UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF POLITICAL SCIENCE, LAW, AND INTERNATIONAL STUDIES

Master's degree in

Human Rights and Multi-level Governance

Unveiling the Role of Artificial Intelligence in Colonial Dynamics

Supervisor: Prof. PIETRO DE PERINI

Candidate: BENEDETTA PESCETTO

Matriculation No: 2090106

A.Y. 2023/2024

UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF POLITICAL SCIENCE, LAW, AND INTERNATIONAL STUDIES

Master's degree in

Human Rights and Multi-level Governance

Unveiling the Role of Artificial Intelligence in Colonial Dynamics

Supervisor: Prof. PIETRO DE PERINI

Candidate: BENEDETTA PESCETTO

Matriculation No: 2090106

A.Y. 2023/2024

ABSTRACT

In the 21st century, we are facing a technological revolution driven by Artificial Intelligence (AI). This technology, fascinating as it is, is permeating our lives and represents one of the most pressing challenges facing nations, societies, communities and individuals.

This study investigates the ethical issues posed by AI and its role as a tool of colonialist domination by big powers and Big Tech. Although these innovative technologies are becoming increasingly integrated into our daily lives, it is crucial for us to recognise that AI is a product of historical data reflecting inequalities and disparities. Using current AI systems as the basis for future decisions risks perpetuating these divergences.

The study initially analyses the concept of Artificial Intelligence and its various forms, reviewing the different legal frameworks proposed by different regions of the world. It then explores the risks associated with the perception of AI as neutral intelligence and the problems associated with algorithmic discrimination. Despite ongoing efforts to promote the inclusiveness of AI in various fields, the debate on AI and advanced technologies has so far been predominantly influenced and dominated by Western culture and wealth that reproduce the forms of oppression of historical colonialism. A particular focus is devoted to the impact of AI on vulnerable individuals and marginalised communities, including indigenous peoples.

The main objective of this study is to highlight how the uncontrolled expansion of the AI empire widens social inequalities, helps perpetuate discrimination against those who have always suffered it during classical colonialism, and negatively impacts marginalised identities and communities, such as indigenous peoples.

The findings underline and highlight the need to develop a decolonised lens for the analysis of data and technologies. The decolonisation of data turns out to be a creative and important endeavour that must start from a deconstruction of the consolidation of a global order based on colonialism, patriarchy, racism and capitalism.

Keywords: *Artificial Intelligence, AI Regulamentation, Algorithmic Discrimination, Data Colonialism, AI Empire, Decolonisation*

TABLE OF CONTENTS

ACRONYMS

AGI - Artificial General Intelligence

AHEG - Ad Hoc Expert Group (UNESCO)

AI - Artificial Intelligence

AI HLEG - High-Level Expert Group on Artificial Intelligence

AI RMF - Artificial Intelligence Risk Management Framework

AIA - Artificial Intelligence Act

AIDA - Artificial Intelligence and Data Act

AIDP - Next Generation Artificial Intelligence Development Plan

ANI - Artificial Narrow Intelligence

CAHAI - Ad Hoc Committee on Artificial Intelligence

CAI - Committee on Artificial Intelligence

CARE principles - Collective Benefit, Authority to Control, Responsibility, and Ethics

CCRA - Corrections and Conditional Release Act

ChatGPT - Chat Generative Pre-trained Transformer

CSC - Correctional Service of Canada

DL - Deep Learning

ECHR – European Court of Human Rights

EO - Executive Order

EU AI Act - European Artificial Intelligence Act

EC - European Commission

FRIA - Fundamental Rights Impact Assessment

G7 - Group of Seven

GAFAM - Google, Amazon, Facebook, Apple and Microsoft

GAI - Generative Artificial Intillegence

GDPR - General Data Protection Regulation

GPAI - General Purpose AI Models

IBM - International Business Machines

ID-SOV - Indigenous Data Sovereignty

IP protection - Intellectual Property Protection

LaMDA - Language Model for Dialogue Applications

MEP - Member of the European Parliament

ML - Machine Learning

NAIIA - National AI Initiative Act

NIST - National Institute of Standards and Technology

OCAP® principles - Ownership, Control, Access and Possession of data in Canada

OECD - Organization for Economic Co-operation and Development

OSCE - Organization for Security and Co-operation in Europe

OSTP - Office of Science and Technology Policy

FAIR Principles - Findable, Accessible, Interoperable, and Reusable

SCC – Supreme Court of Canada

UDHR - Universal Declaration of Human Rights

UNDRIP - United Nations Declaration on the Rights of Indigenous Peoples

UNESCO United Nations Educational, Scientific and Cultural Organization.

US AISI - The United States AI Safety Institute

WEF - World Economic Forum

INTRODUCTION

It is difficult today to spend an entire day without hearing about “artificial intelligence” (AI). We live in an era marked by a growing interest in these technologies, with scientific research demonstrating a keen focus on AI not only within technological and engineering fields but also across anthropology, philosophy, psychology, and law. Artificial Intelligence has become a central topic for reflection to understand its potential impact on humanity and society.

Artificial intelligence is now an integral part of our daily lives; however, most people are not fully aware of it and of the extent to which AI shapes our experiences and interactions. This technological revolution presents one of the most pressing challenges that nations and societies around the world must face in the 21st century.

At the heart of this challenge lies a fundamental issue: the lack of a universally accepted definition of what exactly constitutes Artificial Intelligence. The definitions of AI vary widely across disciplines and contexts, reflecting its diverse perspectives and applications in fields ranging from technology and engineering to philosophy, ethics and law.

It is exactly from these considerations that the questions that animate and enlighten my work arose: To what extent is AI the bearer of neutrality? And to what extent, on the other hand, is it the daughter of cultural prejudices and thus the bearer of Western, white, wealthy thinking? Can AI be discriminatory and be used as an instrument of colonial domination?

The very diversity of definitions underscores the complexity and multifaceted nature of AI, which makes it essential for stakeholders to engage in ongoing dialogue to establish common frameworks and understandings.

As AI evolves rapidly, its implications for society become increasingly profound. The discourse surrounding AI extends beyond technological advancements to include broader ethical, governance, and social impact questions. Navigating these complexities effectively is crucial to ensure that AI developments benefit humanity in a responsible and ethical manner.

In 2021, the European Commission made a significant leap forward by introducing the “first rules on AI”, marking the EU's inaugural regulatory framework for artificial intelligence. In 2023, the European Parliament furthered these efforts by adopting its negotiating position on the AI law, underscoring Europe’s commitment to establishing robust standards for AI governance, and in 2024, the European Parliament formally approved the EU AI Act.

At the same time, the United States made progress with its initiatives, drafting an AI rights bill and issuing an executive order focused on ensuring the safe, secure, and trustworthy use of AI.

China has emerged as a major player in the AI arena, implementing a comprehensive set of legislative measures and national strategies to prioritise the development of AI and communication technologies and to provide regulation that protects citizens' rights.

Beyond Europe, the United States and China, other countries also engaged in legislative efforts related to AI. Brazil advanced with its own AI bill, while Canada developed an Artificial Intelligence and Data Act to evaluate the ethical implications of AI systems.

However, in many nations and regions of the world, there is still a lack of adequate regulation on AI and the use of innovative technologies. This regulatory gap raises critical questions that extend far beyond technological development. How can we ensure that AI systems are safe, respect human rights, are transparent, traceable, non-discriminatory, responsible, and environmentally friendly?

Artificial intelligence poses significant ethical challenges and has a major impact on human rights. Although these new technologies bring extraordinary innovations that will become increasingly integral to our daily lives, it is crucial to recognise that AI is also the product of historical data that reflect existing inequalities and disparities. Using current AI systems as a basis for decisions that will determine our future risks exacerbating existing disparities and perpetuating colonial narratives.

Despite many efforts underway to promote inclusivity in AI in all areas, it should be noted that the debate on AI and advanced technologies has, to date, been predominantly shaped by the “West, whiteness, and wealth”. Considering AI as a

neutral technology carries inevitable risks, as it cannot be inherently so and can create algorithmic discrimination. This phenomenon can manifest itself in the uncontrolled expansion of AI and have a particularly harsh impact on marginalised identities and communities, such as indigenous communities. If in the past it was believed that indigenous societies were destined to disappear due to economic, political, and colonial pressures, today we can observe that this has not occurred. On the contrary, these communities are actively building original identity paths, contributing to making the narrative of modernisation and progress more complex.

These communities in particular, may suffer disproportionate harm due to their historical exclusion from technological decision-making processes and lack of representation in data. Therefore, there is an increasing need to develop a decolonised lens for analysing data and technologies. A decolonised approach requires awareness and active critique of the biases and inequalities embedded in the historical data used to train AI systems. It is imperative to recognise that AI is not developed in a cultural vacuum, but reflects the power structures and biases present in society. Only through such awareness can we work to mitigate the risks of algorithmic discrimination and promote a more equitable and inclusive use of AI.

The first chapter offers an in-depth analysis of the historical evolution of the concept of Artificial Intelligence, exploring its various definitions over time and the many forms it has taken. It starts from the first pioneering theories and visions of AI and then examines the technological and methodological advances that have helped shape the discipline as we know it today.

Subsequently, the legal frameworks pertaining to AI regulation will be thoroughly examined, encompassing the EU AI Act, the draft of the US AI rights bill, China's strategic plan, and the regulatory approaches of Brazil and Canada. This section will highlight the various stages of development in which different countries are in regarding their approaches to AI regulation. It will also explore the different opinions and strategies that these countries have adopted in addressing the challenges and opportunities presented by AI. Countries are at various stages of evolving their approach to AI regulation and have different opinions on how best to do so, and a cohesive and unified global approach to AI regulation is still lacking, despite the

attempts, leading to a fragmented regulatory landscape with significant variations across jurisdictions.

Finally, I will examine the reasons why artificial intelligence cannot be considered as a neutral intelligence. Despite the great opportunities and capabilities that it brings, the risks associated with it are particularly important and must be addressed. In particular, when we refer to individuals or groups that are poorly represented, on the margins of society, or those groups that do not recognise themselves in the dominant thought as indigenous communities. Given that artificial intelligence systems and their algorithms reflect the prejudices, culture, and ideologies of their creators, it is clear that indigenous peoples will face, once again, a new form of domination and discrimination. In this era of AI empire, we are faced with a new form of extractivism; no longer indiscriminate extraction of natural resources and labour force, but human life through data.

Through the use of data that do not take into account cultural diversity and individual particularities, indigenous communities are deleted, and their ontological differences are inaccurately and stereotyped, which only serves to justify oppression and denial of justice. Through *Ewert v. Canada*, it will be shown that colonial thinking is still rooted in the Canadian criminal justice system and that this is exacerbated and not mitigated by the use of automated systems to assess the risk of an inmate.

CHAPTER I

Understanding Artificial Intelligence: The Complexities of its Conceptual Evolution

Introduction

"A curious aspect of the theory of evolution is that everybody thinks he understands it."[1] This phenomenon is also evident in the field of artificial intelligence (AI), where the risk is that people believe they have fully grasped its complexity too soon.

The term "Artificial Intelligence" represents a complex and nuanced concept. In this chapter, our aim is to embark on a journey to untangle the intricate web of AI, tracing its conceptual evolution from its nascent stages to its current multifaceted existence. The goal is to provide a comprehensive understanding of AI, including its historical development, various definitions, and fundamental concepts.

The primary objective of this chapter is to provide a foundational understanding of AI by examining its historical roots and exploring the challenges it has faced from its inception to the present day. We will delve into how AI was conceived, its early applications, and how it has evolved over time, overcoming numerous obstacles, and achieving significant milestones. This historical exploration will lay the groundwork for a more in-depth discussion on the nature and implications of AI.

Subsequently, the multiple definitions of artificial intelligence will be addressed, acknowledging the difficulty of this task given the lack of a universal definition. We will present various relevant definitions proposed by different institutions, highlighting the differences and points of convergence. This section will shed light on how varied the understanding of AI is and how it is perceived through different theoretical and practical lenses.

The notion of artificial intelligence will be examined from a variety of perspectives, emphasising that it "are" many things. Indeed, it becomes evident that AI

[1] A quotation from J. MONOD in E. YUDKOWSKY, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, In Global Catastrophic Risks, 2008, p. 1

not only has faced challenges in its historical and conceptual evolution, but also embodies intrinsic complexities, presenting itself with multiple facets and encompassing various concepts. Finally, the chapter will present the essential concepts of AI, such as machine learning, big data, and will showcase the debate that AI has sparked within the scientific communities, distinguishing between Strong AI and Weak AI, which are fundamental to understanding the current capabilities and limitations of AI, as well as its future potential.

1.1 Tracing the Origins: A Brief History of AI

1.1.1 The Genesis of Artificial Intelligence

The term "Artificial Intelligence" (AI) has taken on a symbolic connotation of modernity, progress, evolution and efficiency. Artificial Intelligence represents one of the most fascinating results of the intersection between the human mind and the computational power of machines. It is undoubtedly one of the most significant technological innovations of our time, its path rooted in a mixture of human ambitions, scientific insights and technological advances, and it is capable of influencing and being influenced by several disciplines, including philosophy, economics, mathematics, neuroscience, psychology, cybernetics, law, cognitive science and linguistics[2].

It is well known in the scientific literature that there is no widely accepted definition of Artificial Intelligence[3] and, as a result, the term "AI" has been used with many different meanings and senses[4].

To fully understand the concept of Artificial Intelligence, it is essential to examine its historical development and the many definitions it has assumed over time. The idea of creating machines capable of thinking dates back to ancient times, but it was in the 20th century that AI began to take shape.

---

[2] M. SOMALVICO, *L'Intelligenza artificiale*, 1987, Rusconi, Milano

[3] Cf. N. J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge, 2009, p. XIII-XIV; P. WANG, *On defining artificial intelligence, in Journal of General Artificial Intelligence*, 10, 2, 2019, p. 1-37; R.J. BRACHMAN, *(AA)AI Presidential Address: (AA)AI More than the Sum of Its Parts*, in AI Magazine, Vol. 27, no. 4, 2006, p. 19–34.

[4] P. WANG, *On defining artificial intelligence, in Journal of General Artificial Intelligence*, 10, 2, 2019, pp. 1-37

To pinpoint the moment of its birth, it is set to coincide with 1956, the year when the Dartmouth Summer Research Project on Artificial Intelligence took place. During this event, a group of US researchers, led by John McCarthy, programmatically founded the new discipline and discussed the expected future development prospects.

This date of birth, commonly identified by the scientific community as the starting point of AI, cannot but also be linked to earlier technological developments that charted the path to follow from a technological point of view.

In 1943, Warren McCulloch (neurophysiologist) and Walter Pitts (mathematician) published an article entitled "A Logical Calculus of the ideas immanent in nervous activity"[5] in the journal "Bulletin of Mathematical Biophysics", in which they showed how an elementary system of artificial neurons could be capable of performing essential logical functions.

Two Harvard undergraduates, Marvin Minsky and Dean Edmonds, then developed the first neural network computer in 1950. Known as SNARC, this device employed 3000 vacuum tubes and exploited the automatic piloting system of a surplus B-24 bomber to simulate a network of 40 neurons[6].

However, one of the crucial moments is the work of Alan Turing, a pioneer of information theory and computer science. In 1950, Turing published his famous article *"Computing Machinery and Intelligence"*[7], in which he proposed the concept of the "Turing test", a criterion to assess whether a machine could be considered "intelligent" and "thinking". In the article, Turing states that addressing this issue requires a clear definition of the concepts of "machine" and "thought". Regarding the first term, Turing offers a description of an ideal machine capable of performing any kind of calculation, a concept that still forms the theoretical basis of all our computers today: the famous Turing machine. Regarding the definition of "thought", Turing proposes an operational approach using the "Turing test" or "imitation game". According to this test, if one cannot distinguish between a machine and a human being, within a certain period of time, one can consider the machine to be intelligent. The purpose of artificial intelligence, then, is to act like a human being, to the point of making itself

[5] W. MCCULLOCH, W. PITTS, *A logical calculus of the ideas immanent in nervous activity,* in Bulletin of Mathematical Biophisiscs*,* Vol 5, pp 115-133, 1943

[6] S. RUSSEL, P. NORVIG, *Artificial intelligence, A Modern Approach (4ª edition.)*, 2020, p. 35

[7] A. TURING, *Computing machinery and intelligence*, in *Mind*, Vol. 59, no. 236, 1950

indistinguishable from the latter. This test helped shape the debate on the cognitive capabilities of machines and stimulated AI research, and is still used today to discriminate between human and artificial intelligence.

Although there were important innovations before that date, the choice of 1956 as the starting point for AI, is mainly attributable to the aforementioned seminar at Dartmouth College in Hanover, New Hampshire. During this event, two Carnegie Tech researchers, Allen Newell and Herbert Simon, stole the show by proposing a reasoning programme, the Logic Theorist (LT), about which Simon stated:

"*We have invented a computer program capable of thinking non-numerically and thereby solved the venerable mind-body problem."*[8]

Despite the various innovations proposed by the various researchers, the most enduring outcome of the workshop was the term "Artificial Intelligence", which scientist John McCarthy coined and according to which:

"*The Artificial Intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs."*[9]

1.1.2 From Birth to Now: AI's Triumphs and Trials

The years following the Dartmouth workshop were the years of great expectations and various currents of thought developed to define the scope and purpose of this field of study. In reality, the Dartmouth event ended without revolutionary results, resolving itself into an occasion for meeting and initiating scientific collaborations[10], rather than continuous research work. Nevertheless, this climate of confidence persisted for the whole of the following decade, characterised by significant progress in the field, which led to the assumption, as we shall see, erroneously, that most of the goals enunciated in 1956 were achievable in a relatively short time.

It is important to emphasise that the aim of Artificial intelligence is not the robotic replacement of human intelligence, a goal considered inadmissible by science

[8] S. RUSSEL, P. NORVIG, *Artificial Intelligence, A Modern Approach* (2^{nd} edition), 2003, pp. 16-18

[9] J. MCCARTHY; *What is Artificial Intelligence*, 2007, available at: http://www-formal.stanford.edu/jmc/whatisai.pdf

[10] S. RUSSEL, P. NORVIG, 2020, op. cit., p. 19

itself, but rather, the focus is on the emulation of those particular skills and capacities peculiar to mankind that a machine can reproduce. Indeed, from the outset, numerous researchers have pursued the development of autonomous agents aiming to emulate human intelligence and behaviour as closely as possible, each with their own interpretations and approaches. Others have defined the purpose of AI to be to think like a human being[11] (defining it as "make computer think"[12] or "machines with minds."[13]). In academic circles, a statement by Herbert Simon is widely quoted as having shown considerable optimism about the prospects of artificial intelligence during that historical period:

"It is not my aim to surprise or shock you – but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until – in a visible future – the range of problems they can handle will be coextensive with the range to which the human mind has been applied."[14]

The hopeful forecasts of the preceding decade collided with reality in the early 1960s, as the anticipated swift advances in the realm of artificial intelligence were notably delayed in materialising. This delay originated predominantly from the challenge of effectively applying the initial technologies developed in the field to real-world challenges[15]. Specialists often refer to this period as the "*first winter of artificial intelligence*" to emphasise the stagnation of progress and investment that characterised it.

Interest in artificial intelligence only grew again with the advent of so-called expert systems. The basis of expert systems was the idea of obviating the problem by developing tools capable of operating in very narrow fields of reality, applying knowledge and decision rules encoded through programming. Expert systems are computer programmes that attempt to reproduce the performance of human experts in solving problems.[16] In less than a decade, the turnover related to artificial intelligence

[11] Among the pioneers and most renowned figures of this movement are Allen Newell and Herbert Simon, American researchers who, in 1957, created a program named the General Problem Solver. A. NEWELL, J. C. SHAW, H. A. SIMON, *Report on a general problem-solving program*, 1959
[12] J. HAUGELAND, *Artificial intelligence: the very idea*, Cambridge (US), London, 1985, p. 2.
[13] *Ibidem*
[14] S. RUSSEL, P. NORVIG, 2020, op. cit., p. 17
[15] *Ibidem*, p 21-22
[16] In Enciclopedia Treccani, appendix V, 1994

has grown from a few million to billions of dollars, giving rise to dozens of specialised companies[17]. This explosion naturally had significant repercussions on research funding, which experienced a renaissance in terms of investment. The 1980s not only saw a resurgence in funding for the field, but also marked a moment of renewed interest in and development of neural networks. Despite major projects, once again, by the end of this decade, expectations of artificial intelligence had exceeded the real possibilities of progress, leading to the *"second winter of artificial intelligence*"[18].

The upswing that took place in the 1990s, with huge investments returning to the sector, enabled the development of increasingly complex algorithms and the growth of the computational capacity of computers. In the context of that period, 11 May 1997 marked a symbolic event in the development of artificial intelligence, the first to capture the attention of the general public: the historic victory of *Deep Blue*. This supercomputer, developed by IBM, defeated the world chess champion of the time, Garry Kasparov.[19]

Then, the abstract concept of rationality emerged, understood as the performance of a model agent, not necessarily equivalent to human intelligence, which sets the goal for artificial intelligence to develop systems that adhere as closely as possible to this ideal canon of rationality[20]. The development of a rational agent became the common goal of many scholars, generating a sharing of results and strategies. The remarkable developments of that period and access to a vast amount of data catalysed a rapid advancement in the field of artificial intelligence, culminating in the emergence of *machine learning.* This period anticipated what happened in the early 2000s, with the spread of the World Wide Web leading to an exponential growth of available data, known as "*Big Data*"[21]. The opportunities offered by big data for the development of artificial intelligence exceeded expectations, leading to significant advances in image recognition and natural language processing.

---

[17] S. RUSSEL, P. NORVIG, 2020, op. cit., pp. 22-23

[18] M. LIM, *History of AI Winters,* 2018 available at https://www.actuaries.digital/2018/09/05/history-of-ai-winters/ (May 11, 2024)

[19] B. WEBER, *Swift and slashing, computer topples Kasprov,* in New York Times, May 12, 1997, available at https://www.nytimes.com/1997/05/12/nyregion/swift-and-slashing-computer-topples-kasparov.html (May 11, 2024)

[20] M. SOMALVICO, 1987, op. cit., p. 4

[21] C. AGATA, *Intelligenza Artificiale, Big Data e nuovi diritti*, in Rivista Italiana di informatica e diritto, vol. 4, no. 1, 2022, pp. 94-97. These are huge and complex data sets, which require advanced tools to be managed, analysed and understood.

The arrival of Big Data and the unprecedented power of computers also enabled the development of *Deep Learning*, a subset of *Machine Learning* techniques characterised by intricate neural networks composed of multiple layers of artificial neurons. Machine Learning itself refers to the field of artificial intelligence where algorithms learn from data and make predictions or decisions without being explicitly programmed. *Deep Learning* emerged as one of the dominant technologies in the field of artificial intelligence in the 2000s and led to great advances in the fields of image and language recognition, speech recognition, machine translation and was accompanied by the victory of a computer system over a human champion[22].

In recent years, AI has undergone considerable changes both at the methodological and content levels. What remains of the characterisation of early AI is the plurality of approaches: Alongside the traditional logical approach of knowledge representation, the subsymbolic approach, which aims to equip AI systems with intelligent capabilities even without a detailed knowledge representation, has gained increasing importance. A great deal of attention has also been paid in recent years to probabilistic and fuzzy methods, which are used to enable efficient reasoning based on uncertain evidence[23]. These models can recognise, represent, manipulate, interpret and utilise data and information that are ambiguous and lack certainty. Probability is suitable for well-defined systems where uncertainty arises from randomness. Fuzzy logic, on the other hand, is ideal for complex systems where uncertainty comes from imprecision or ambiguity. Together, these two methodologies provide powerful tools for addressing various forms of uncertainty.[24]

The prevailing approach today, therefore, holds that AI should aim at developing systems capable not only of reasoning but also of rational behaviour, being able to act adaptively and efficiently in the context in which they operate. It is, in fact, expected to use the expression rational agent[25].

---

[22] In 2016, AlphaGo, a software developed by Google, defeated South Korean champion Lee Sedol in Go, one of the most complicated games in the world.
[23] M. SOMALVICO, 1987, op. cit., p. 4
[24] Data Headhunters, *Fuzzy Logic vs Probability: Handling Uncertainty in Data,* January 5, 2024 https://dataheadhunters.com/academy/fuzzy-logic-vs-probability-handling-uncertainty-in-data/ (July 4, 2024)
[25] S. RUSSEL, P. NORVIG, 2020, op. cit., pp. 3-4

1.2 Exploring the Multifaceted Definition of AI

1.2.1 Difficulties in defining AI

Defining AI is a challenging task; indeed, there exists no universally accepted definition of the concept[26]. The definitions of artificial intelligence have been shaped by the perspectives of various scholars and the nature of technological progress. Initially, the definition reflected the optimism of AI's early days and the belief in the possibility of replicating human intelligence through symbolic processing. However, AI has not remained confined to a single static definition. Over the years, conceptions of what constitutes intelligence and how it can be replicated in machines have undergone significant evolution and, as E. Yudkowsky aptly warned, *"By far the greatest danger of Artificial Intelligence is that people conclude too early that they understand it."*[27]

As previously mentioned, the field of AI research owes much to the contributions of McCarthy, Minsky, Simon, and Newell. Their involvement in the Dartmouth workshop was pivotal, but equally important were their efforts to establish three research centres, which played a significant role in shaping the trajectory of mainstream AI for decades[28]. Their own opinion on AI was as follows:

"By 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action: that in any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity."[29]

Intelligence usually means "the ability to solve hard problems."[30]

"AI is concerned with methods of achieving goals in situations in which the information available has a certain complex character. The methods that have to be used are related

[26] S. RUSSEL, P. NORVIG, 2020, op. cit.
[27] E. YUDKOWSKY, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, In *Global Catastrophic Risks,* 2008, p. 1
[28] P. WANG, 2019, op. cit.
[29] A. NEWELL, H. A. SIMON*, Computer science as empirical enquiry: Symbols and search*, Vol. 19, No. 3, 1976, pp. 113-126
[30] M. MINSKY, *The society of Mind,* The personalist Forun, Vol 3, no. 1, 1987

to the problem presented by the situation and are similar whether the problem solver is human, a Martian, or a computer program."[31]

In its broadest sense, AI is often associated with algorithms. However, this approach is not particularly useful for our analysis. Algorithms have existed prior to the advent of AI and have been extensively employed in various fields beyond AI[32].
In its strictest sense, AI refers to the replication by computers of the intelligence innate to humans.

A prevalent understanding of AI is that it empowers machines to replicate various intricate human capabilities. However, this description lacks specificity, essentially rephrasing the term "artificial intelligence". Without clarification on what constitutes these "complex human abilities", the true essence of AI remains elusive. Similarly, defining AI as the performance of intricate tasks by computers in multifaceted environments also falls short. Alternate definitions strive to elaborate on these abilities and tasks considering AI as a technology that operates effectively and anticipates the demands of its surroundings[33], or the capacity to perceive, take initiatives, to pursue goals and adapt based on feedback. Although they help to better understand the concept of what AI is, they still have limitations.

Given the complexities involved, it's no wonder that defining AI with precision poses significant challenges. After all, it attempts to replicate or simulate something that we still do not fully comprehend ourselves: human intelligence. Indeed, intelligence is integral to our human experience, enabling us to navigate the complexities of the world around us. The etymology of the word *"intelligence"* (the prefix *inter*, meaning "between", and the Latin verb *legere*, which initially meant "to choose, select" and evolved from there into the word "to read")[34] offers intriguing insights into its essence.

The idea of gathering, collecting, and assembling information aligns with the process of cognition and understanding. It reflects the innate human ability to sift through data, make selections, and ultimately arrive at comprehension. This capacity

---

[31] J. MCCARTHY, *Mathematical Logic in Artificial Intelligence,* Daedalus, Vol. 117, No. 1, 1988, pp. 297-311
[32] H. SHEIKH et al, *Mission AI, The New System Technology,* Research for Policy, 2023, pp. 15-19
[33] N. J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, 2009, p. 13
[34] S. DE SPIEGALEIRE et al, *Artificial Intelligence And the future of defense: strategic implications for small-and medium-sized force providers. What is artificial Intelligence*, 2017, pp. 25-42

sets us apart, allowing us to perceive, interpret, and act upon our surroundings in unique ways. In essence, intelligence is not merely about raw knowledge but also about the skillful synthesis and application of information to gain insight and make informed decisions. Absolutely, our understanding of intelligence, both human and artificial, remains an ongoing pursuit with its fair share of complexities, thus, the challenges encountered in defining AI do not stem from any carelessness or oversight, but rather from our prolonged inability to precisely delineate the type of intelligence we intend to artificially imitate.

With the variety of divergent opinions on what AI is, the lack of agreement on a standard assessment (e.g., criteria, benchmarks, milestones) makes it extremely difficult for the industry to maintain healthy growth. Due to the reliance on intuitive yet nebulous notions of intelligence, mainstream AI has evolved within a realm that lacks not only a unified theoretical framework but also consensus regarding overarching research objectives.

AI is frequently associated with cutting-edge technology. As we delve further, it becomes evident that AI has experienced significant momentum in recent years. Notably, advancements in the field of "*machine learning*" (ML) have played a pivotal role, leading to the emergence of "*deep learning*" (DL). Unlike conventional methodologies where computer systems operate based on rigid rules, ML and DL algorithms possess the ability to identify patterns within data, often referred to as "self-learning algorithms"[35]. These techniques, rooted in disparate theoretical underpinnings and applicable to diverse problem sets, have given rise to various AI sub-domains, including knowledge representation, reasoning, planning, machine learning, vision, natural language processing, and robotics. Many researchers align more closely with these specialised sub-fields than with AI itself, viewing AI as an optional label that can be applied or discarded based on its fluctuating public perception, which has experienced considerable volatility over time[36].

The definition of AI has expanded to include both systems that specifically mimic human intelligence (weak AI) and those that aim for a more universal form of intelligence (strong AI or AGI, Artificial General Intelligence). The evolving nature of this scientific discipline implies a constant redefinition of the concept of AI over time.

[35] H. SHEIKH et al, 2023, op. cit.
[36] P. WANG, 2019, op. cit.

Current applications considered AI generally fall into the category of "narrow" or "weak" AI. Indeed, the AI we know today focuses on specific skills, whereas the goal of AGI is to achieve a more complete understanding and simulation of human cognitive capabilities, which, with such a generic definition, will be difficult to achieve.

As previously illustrated, Artificial Intelligence has had a rich and articulated path, characterised by scientific discoveries, technological advances and philosophical debates. Its definitions have been shaped by the visions and insights of its pioneers, as well as the challenges and opportunities offered by research and innovation. AI continues to evolve rapidly, pushing the boundaries of human knowledge and opening up new perspectives on the future of intelligence and technology.

1.2.2 Institutional Perspectives on Artificial Intelligence Definition

In recent years, numerous international and regional institutions have worked to develop a comprehensive definition of artificial intelligence. Among these contributors, the institutions of the European Union have played a significant role, collaborating to shape a unified understanding of AI.

This effort was initiated with a Communication from the European Commission dated April 25, 2018, entitled "Artificial Intelligence for Europe"[37]. In this document, has been proposed a first institutional definition of what Artificial intelligence is:
"*Systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specifc goals."*[38]

This definition seems to reflect a clear reference to human intelligence through the expression "intelligent behaviour", which has been a subject of discussion, while concepts like "some degree of autonomy" remain vague and not well-defined. Subsequently, the *High-Level Expert Group on Artificial Intelligence* (AI HLEG) of the European Commission (EC) was created. This is a body composed of 52 high-level experts established in June 2018 with the aim of providing recommendations and

[37] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the regions, *Artificial Intelligence for Europe,* COM/2018/237 final*,* Brussels, April 25th 2018
[38] *Ibidem,* p. 2

guidance on the ethical and responsible use of artificial intelligence in the European Union. Based on the definition proposed by the European Commission, AI HLEG has developed a new definition:

"Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)."[39]

This definition provides a broad and comprehensive view, highlighting that AI systems can consist of both software and potentially hardware designed by humans to achieve complex goals. It also encompasses various AI approaches and techniques, presenting a holistic perspective that underscores the sophistication and versatility of AI.

Additionally, in 2021, the European Commission developed a Proposal for a Regulation on AI, wherein Article 3 introduces a kind of "official definition" that would have legal binding within the European context:

"'*Artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I*[40] *and can, for a given set of*

[39] High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*. European Commission, 2019, p. 6

[40] Art. 3 par. 1, *Proposal for a Regulation of the European Parliament and of the Council layng down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts,* Brussels, 2021, p. 39. The techniques for developing artificial intelligence systems listed in Annex I of the Proposal, divided into three categories, are: "(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimisation methods".

human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with."[41]
The formula appears to succinctly capture the essential features of those developed by the Expert Group.

On March 13, 2024, the European Parliament endorsed the AI Regulation text, wherein the definition of an AI system as stipulated in Article 3(1) of the prior text underwent slight modifications to better align it with the efforts of international organisations:
"'AI system' means a machine-based system designed to operate with varying levels of autonomy, 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."[42]

In the preceding two definitions, it is evident that they capture the fundamental characteristics of the one elaborated by the Expert Group. However, a notable distinction arises regarding the expressed tendency to confine the scope of AI exclusively to software, with a deliberate avoidance of references to hardware systems.

Furthermore, Recital 10 of article 3 of the EU AI Act, explicitly clarifies that the definition is not intended to encompass traditional software systems or simpler programming approaches "*based solely on rules defined by natural persons to automate operations."*[43] Moreover, the Commission has been tasked with developing guidelines for the implementation of the AI system definition.

Despite significant advancements in this field by the European Union, it is crucial to recognise that it is not the sole international body engaged in the institutional definition of AI. Various organisations and entities have contributed to crafting defining frameworks, primarily characterised by an attempt to delimit the field of AI by listing its main applications, while leaving in the background the theory of the rational agent, which is of great relevance for the definitions seen previously. One of the key actors in the global debate on Artificial Intelligence is UNESCO. The organisation promotes a

[41] Art. 3 par. 1, *Proposal for a Regulation of the European Parliament and of the Council layng down harmonized rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts,* Brussels, 2021, p. 39
[42] Art. 3 par. 1, *EU Artificial Intelligence Act,* 2024, *Definitions*
[43] Art. 3 Recital 10, *EU Artificial Intelligence Act*, 2024

global approach to understanding and addressing its complexities, from education to scientific dissemination, through ethics, and commits to guiding its development and application responsibly. In the 40th plenary session of November 2019, UNESCO appointed an *Ad Hoc Expert Group* (AHEG) to draft a *Recommendation on the ethics of Artificial Intelligence*[44]. This document represents the first regulatory instrument that establishes the ethical principles of AI in accordance with human rights and fundamental freedoms and was finally approved and adopted on November 24, 2021.

The document states that:

"AI systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to: (i) machine learning including deep learning and reinforcement learning; (ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization.

AI systems can be used in cyber-physical systems, including the Internet-of-Things[45]*, robotic systems, social robotics and human-computer interfaces which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work."*[46]

The Organization for Economic Co-operation and Development (OECD) has also recognised the importance of Artificial Intelligence through a Recommendation adopted by the Council of Ministers of its member states on May 22, 2019, subsequently amended on May 3, 2024. This document aims to establish a set of principles and measures aimed at ensuring the responsible and reliable development of AI, while upholding fundamental values centred on humanity and equity. Among its core contents are guidelines to promote transparency, accountability, security, and

[44] UNESCO, United Nations Educational, Scientific and Cultural Organization, *Recommendation on the Ethics of Artificial Intelligence*, November 24 2021

[45] Internet of Things (IoT) refers to that technological development whereby, through the Internet, every object acquires its own identity in the digital world. Thus, IoT is based on the idea of "intelligent" or "smart" objects interconnected with each other in order to exchange the information they possess, collect and/or process.

[46] UNESCO, *November 24 2021*, op. cit., p. 10

privacy in the implementation and use of AI. Additionally, the Recommendation proposes a clear definition of artificial intelligence, which has served as inspiration for the European Union's regulatory framework on AI:

"An AI system is a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."[47]

Another significant player in this arena, in terms of national legal systems, is the Canadian legal system. In 2019, Canada introduced a legislative milestone: the introduction of the first official definition of artificial intelligence embedded in a legal text with legal force. This enactment represents a significant step in providing a clear and binding legal framework to regulate the utilisation and advancement of AI within the nation. The formulation of this definition, arising from a thorough consultative process involving experts, practitioners, and civil society representatives, aims to furnish a robust foundation for guiding public policies, fostering responsible innovation, and safeguarding citizens' rights. This definition diverges from those previously encountered, as it hinges on concepts in which the objective of AI lies in emulating the behaviour and cognitive capacities of human beings. The proposed definition from Canada is encapsulated within the Directive on automated decision-making and delineates artificial intelligence as:

"Information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems."[48]

---

[47] OECD, *Recommendation on Artificial Intelligence*, adopted May 22 2019, amended May 3 2024, OECD member countries approved a revised version of the organisation's definition of an IA system available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (May 10, 2024)

[48] Government of Canada, *Directive on automated decision-making*, April 1 2019, Annex 1, https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592 (May 10, 2024)

1.3 Essential concepts of AI

1.3.1 Machine Learning

Learning, like intelligence, encompasses such a wide range of processes that it is difficult to define precisely. The dictionary defines it with phrases such as "acquiring knowledge, understanding, or skills through study, instruction, or experience" and "modifying a behavioral tendency through experience."[49]

When it comes to machines, we could say, in a very broad sense, that a machine learns each time it adjusts its structure, program, or data (based on its inputs or in response to external information) in order to enhance its anticipated future performance.

The origin of the modern understanding of the term machine learning is usually associated with the psychologist from Cornell University, Frank Rosenblatt, who in 1958 invented, together with his research group, a machine called the "perceptron." This machine had the ability to recognise letters of the alphabet, achieved through three key components: an input layer, an output layer, and an algorithm that allowed it to learn by minimising errors. Once the output was obtained, it was compared with an ideal output value to understand how much more the machine needed to work to improve and get as close as possible to the desired result.

It was Arthur Lee Samuel who first introduced the concept of machine learning in 1959[50]. However, the most iconic definition of Machine Learning was given by Tom Micheal Mitchell, another American scientist:

"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."[51]

This definition encapsulates the essence of Machine Learning, which is to enable computers to enhance their performance through experiential learning. In other words, machine learning aims to empower computers to refine their programming abilities based on the outcomes of tasks or actions they undertake.

---

[49] N. J. NILSSON, *Introduction to Machine Learngin,* Robotics Laboratory, Department of Computer Science, Stanford University, 1998, p. 1

[50] A. L. SAMUEL, *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal of Research and Development, 1959, pp. 206-226. In those studies, he reported the design of a digital computer to behave in a way that, had it been done by humans or animals, would have been described as a learning process.

[51] T. M. MITCHELL, *Machine Learning,* McGraw-Hill Science/Engineering/Math, 1997, p. 2

ML is a subset of artificial intelligence and can now be considered its beating heart. Its task is to instruct computers to learn from data and improve with experience, rather than being specifically programmed to do so. In the field of machine learning, algorithms are trained to identify patterns and correlations within vast datasets, thereby generating optimal decisions and predictions based on such analyses. Machine learning applications progressively improve with use and become increasingly accurate as they gain access to more data.[52]

Machine learning constitutes a vast territory within which we encounter several distinct types and approaches, reflecting the complexity and diversity of the challenges faced by artificial intelligence. Among these main typologies, we can identify the following:

- Supervised Learning: This category forms one of the fundamental foundations of intelligent automation. Supervised learning models are trained using labelled data sets. Learning occurs through the use of numerous examples, which are used to train an algorithm to map input variables to desired outputs. On the basis of these examples, machine learning models become capable of identifying patterns that link inputs to outputs. These models can then apply the rules refined during training to transform new inputs into classifications or predictions. A classic example of supervised learning is the use of various features to determine whether an email should be classified as spam or not.[53]
- Unsupervised Learning: Unlike supervised learning, unsupervised learning models use unlabelled data. Whereas supervised learning consists of mapping the relationships between input and output, unsupervised learning focuses on identifying the intrinsic structures of the data. One of the most common applications of this approach is clustering. Here, the model receives unlabelled input data and determines similarities and differences between various data points, grouping them into clusters based on similar characteristics. These clusters help to categorise the input data, providing a clear picture of the relationships and structures present within the dataset.[54]

---

[52] G. SANGUINETTI, *Machine Learning: accuratezza, interpretabilità e incertezza*. Ithaca: Viaggio nella Scienza XVI, 2020, pp 71-78

[53] D. LESLIE et al, *Artificial Intelligence, Human Rights, Democracy and the Rule of Law, A Primer*, Council of Europe and The Alan Turing Institute, June 2021, p. 9

[54] *Ibidem.*

- Semi-supervised Learning: This method combines a small set of labelled data with a large set of unlabelled data. The algorithms use the labelled data to create an initial model, which is then refined with the unlabelled data. It is particularly useful when labelling data is expensive or difficult, allowing large amounts of data to be exploited with less annotation effort.[55]
- Reinforced Learning: Reinforcement learning models are inspired by the behaviour of biological organisms and learn based on their interactions with a virtual or real environment, rather than on existing data. Reinforcement learning "agents" learn to make decisions based on a series of actions and feedback received from the environment. This type of learning is based on an iterative process in which the agent performs an action, observes the result and receives a reward or punishment. Through this continuous cycle, the agent improves its strategies to maximise cumulative rewards in the long term. Reinforcement learning is particularly useful in contexts where sequential decision-making is required, such as in games, robotics and resource management.[56]
- Deep Learning: Deep learning represents a sophisticated subset of machine learning (ML) methodologies wherein multilayer neural networks are trained on extensive datasets. Derived from conventional neural networks, deep learning significantly outpaces its predecessors, emerging as the predominant computational paradigm within the ML domain[57]. The term "deep" in deep learning refers to the utilisation of deep artificial neural networks[58], which draw inspiration from the intricate workings of neural networks observed in the human brain, along with the handling of vast amounts of intricate and diverse data. This approach involves intricate interactions within the network's multiple layers, progressively extracting more abstract and refined outputs at higher levels of representation. Thanks to interconnected layers of neurons, these

[55] *What Is Machine Learning? Definition, Types, and Examples*, 2024, https://www.coursera.org/articles/what-is-machine-learning (May 16, 2024)
[56] D. LESLIE et al, 2021, op. cit., p. 9
[57] L. ALZUBAIDI, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, in Journal of Big Data, 2021, pp. 1-4
[58] A. SALMAN, *Reti neurali artificiali: dal MLP alle più recenti architetture di Convolutional Neural Networks*, 2017, p 1. An artificial neural network, or simply neural network, is a computational model inspired by the nervous system of living organisms. It is able to learn and acquire knowledge by modifying its structure according to incoming data and internal connections. Information is stored in the connection weights of the network.

networks are particularly effective in analysing complex data such as images, sounds and text.

- Generative AI[59]: This model refers to a subset of machine learning AI technologies that have recently developed the ability to rapidly create new content, including audio, code, images, text, simulations and video. Generative AI (GAI) software transforms user-provided "prompts", which are given in natural language, into a variety of outputs. These outputs can include generating text from other text (Text-to-Text), creating images based on textual descriptions (Text-to-Image), or producing images from other images (Image-to-Image). By utilizing sophisticated algorithms, this technology interprets the given input and produces relevant and contextually accurate results.[60]

Machine Learning represents a pervasive technology that permeates many aspects of our daily lives. This rapidly evolving technology is expanding its horizons at an astonishing pace, revealing only a fraction of its potential. Advances in machine learning promise to further transform numerous sectors, offering innovations that could radically change the way we live and interact with the digital world. The discovery and implementation of new functionalities not only underscore the versatility and power of this technology but also suggest that we are only at the dawn of a technological revolution with largely unexplored potential.

1.3.2 Big Data

Big Data represents one of the most significant innovations in the field of information technology and data management. Providing an analytically precise definition of "*Big Data*" proves to be a complex task. This is because the expression itself refers to an indefinite "magnitude", and the concept of "data" is inherently subjective. Indeed, the perception of what constitutes "data" varies depending on the

[59] Generative AI took the world by storm in the months following the release of ChatGPT, a chatbot based on OpenAI's GPT-3.5 neural network model, in 2022. GPT stands for generative pretrained transformer, words that primarily describe the model's underlying neural network architecture.

[60] P. LICATA, *Generative AI: che cos'è e quali sono le applicazioni di business dei sistemi come ChatGPT*, in Digital4, june 19, 2023, https://www.digital4.biz/marketing/generative-ai-che-cosa-e-quali-sono-le-applicazioni-di-business/ (May 16, 2024)

perspective of the observer and the specific interpretation of its meaning. These characteristics make the nature of "Big Data" challenging to delineate in a singular and precise manner.

Many have attempted to give a definition, but the most widely accepted is due to IBM, which characterises Big Data in terms of four variables (the four Vs): Volume, Variety, Velocity, Veracity[61].

Even when we limit ourselves to the term itself, the most evident parameter is certainly volume. This aspect reflects the immense amount of data that is generated, collected, and stored. When the quantity of data exceeds a certain critical threshold, it becomes practically impossible to analyse it using conventional techniques that require human intervention. Analysts identify Big Data as exceeding the 50 Terabyte threshold or data volume growth of more than 50 per cent per year[62]. In these circumstances, the need arises to adopt machine learning solutions to extract meanings and patterns from this sea of information. Variety refers to the diversity of data types that are generated and collected from a multitude of sources. This variety can include structured data, such as relational databases, and unstructured data, such as text, images, and videos. Similarly to the volume of data, when the complexity and diversity of information exceed a certain threshold[63], managing and analysing this multiplicity requires the use of advanced and specialised tools and techniques. These tools are necessary to extract value from all these heterogeneous data sources, making it once again impossible to rely on traditional techniques. Velocity poses the same problem as the previous parameters. This variable indicates the speed at which data is generated, collected, and analysed. The ability to capture, process, and analyse this data in near real-time is essential for obtaining useful insights and making timely decisions. If a network produces information at a speed too high for human processing, alternative techniques must be adopted. The initial parameters are fundamental and intrinsically linked to the concept of "Big Data". This concept becomes relevant when the computational capacity required to extract meaning from such data exceeds the capabilities of traditional methodologies,

---

[61] G. LONGO, *Big Data e Intelligenza artificiale: che futuro ci aspetta?,* S&F Scienzaefilosofia.It, no. 20, 2018, pp. 15ss

[62] *Cosa sono i Big Data e come vengono utilizzati*?, in BNova, february 24, 2022 https://www.bnova.it/data-science/cosa-sono-i-big-data/ (july 4, 2024)

[63] The value of this threshold depends, from time to time, on the performance of the hardware available at a given time.

making it indispensable to adopt approaches based on automatic techniques capable of emulating some human abilities. In this context, the use of advanced techniques becomes not only useful but necessary to manage the complexity of the data. When discussing the fourth V, veracity, it is important to note that it is not an exclusive parameter of Big Data. Veracity concerns the quality and accuracy of the data. With the enormous volume and speed at which data is generated, questions about its reliability frequently arise. Ensuring that data is accurate, complete, and reliable is essential to draw valid conclusions and make informed decisions. However, as the volume and complexity of data increase, it becomes increasingly difficult to guarantee this accuracy, inevitably affecting the precision of the results obtained. Therefore, veracity, rather than being a distinctive feature, should be seen as a limiting factor that influences the overall handling of Big Data.

Additionally, in some contexts, other "Vs" such as "Value" and "Variability" have been added to reflect further aspects of Big Data. The former represents the potential value that can be derived from data. Big data offers the opportunity to analyse and exploit data to gain valuable insights, identify trends, patterns and correlations, improve decision-making, identify new business opportunities and offer a personalised customer experience[64]. While variability represents the changeability of the meaning of the data according to the context to which they refer and which therefore must be analysed taking into account the different possible interpretations. However, the four main Vs remain fundamental to understanding the complexity and challenges of Big Data.[65].

Big Data represents a revolution in the way organisations collect, store and analyse data. Advanced Big Data technologies are transforming entire sectors, offering new opportunities but also posing new challenges that require innovative solutions.

In today's information age, Big Data has emerged as a fundamental element that is radically transforming every aspect of our existence. Their impact on contemporary times is evident, ranging from accelerating scientific innovation to improving business processes. With their ability to identify complex patterns, predict future developments

[64] M. CASTIGLI, *5 V dei Big data, cosa sono, quale ruolo rivestono*, in BigData4Innovation, 2023 https://www.bigdata4innovation.it/big-data/5-v-dei-big-data-cosa-sono-quale-ruolo-rivestono/ (July 4, 2024)
[65] G. LONGO, 2018, op. cit.

and inform strategic decisions, they occupy a central role in technological and social evolution. This opens up new possibilities for artificial intelligence, medicine, the environment and many other fields. Ethically managing and interpreting this vast source of information is essential to drive progress and ensure the well-being of society in an increasingly interconnected and data-driven future.

1.3.3 Strong AI vs Weak AI

Artificial Intelligence has sparked intense debate within the scientific and philosophical communities on issues such as the nature of intelligence, the possibility that machines might ever become intelligent or develop a mind, and the ethical considerations of creating intelligent machines. Although these questions have been the subject of philosophical discussion for centuries, recent advancements in AI have renewed focus on these topics.

Philosophers and scientists distinguish between two main hypotheses: Weak AI or Narrow AI (ANI - Artificial Narrow Intelligence) and Strong AI or General AI (AGI - Artificial General Intelligence). These two approaches reflect, in a way, two different directions of research and development in the field of simulating the human mind through artificial intelligence.

The distinction between Strong AI and Weak AI is traditionally attributed to John Searle, who in 1980, with the publication of his article "*Minds, Brains and Programs"*[66] introduced the definitions of "Strong AI" and "Weak AI." These definitions are useful for distinguishing between two types of AI efforts. Searle defines Weak AI as an auxiliary tool for the human mind, whereas Strong AI, in contrast, is not merely a simulator of the mind but a genuine mind with accompanying cognitive states.

Strong AI is associated with the assertion that machines are (or will be) capable of thinking, that is, possessing an intelligence indistinguishable from the human mind[67]. Achieving Strong AI is the ultimate goal of many artificial intelligence researchers.

---

[66] J. R. SEARLE; *Minds, Brains and Programs, in The Behavioral and Brain Sciences*, Vol. 3, 1980, Cambridge University Press

[67] CISV, Associazione Italiana per l'Intelligenza Artificiale, Università degli studi di Bari, ONG 2.0, *L'intelligenza Artificiale per lo Sviluppo Sostenibile*, 2021, pp. 23-24

Searle's article attempts to demonstrate that Strong AI (using computers) is impossible. Nevertheless, practitioners of Weak AI (or "cautious" AI) use programs as tools to study the mind, formulating and testing hypotheses about it. Weak AI is also associated with efforts to build programs that assist, rather than replicate, human mental activities, and posits that machines are –or will be– able to behave as if they are intelligent, capable of solving all the problems that human intelligence can solve. Weak AI has already achieved –and continues to achieve– considerable success, while the quest for Strong AI will undoubtedly continue for a long time[68].

On the one hand, then, artificial intelligence systems are identified that are indeed capable of exceeding the capacity of the human mind, e.g. in terms of speed and precision, but which nevertheless remain anchored to the field for which they have been designed. On the other hand, artificial intelligence systems that are capable not only of simulating human behaviour but also of developing their own are identified, regardless of the context in which they are embedded.

To better clarify the distinction between weak AI and strong AI, one could consider the autonomy of these systems. "Artificial General Intelligence" is often described as "human-level AI." Although it remains hypothetical, its main characteristic is a general problem-solving ability that allows it to learn new tasks across various domains. This represents a system that endows a computer with autonomy in thought processes, enabling it to act without requiring human supervision. On the other hand, when people refer to existing technology as "AI", they often classify it as "Narrow AI" (ANI). It is termed narrow precisely because it performs tasks for which human assistance is always essential; the machine, lacking autonomy, would not be able to accomplish any activity without human supervision and can only execute tasks within a very specific and well-defined domain.[69].

The distinction between weak AI and strong AI is fundamental for understanding the development directions and challenges posed by artificial intelligence. While weak AI continues to improve and significantly impact our daily lives, research on strong AI raises critical questions that require careful consideration.

[68] N. J. NILSSON, 2009, op. cit., p. 388
[69] AI: Narrow AI vs. General AI, 2018 https://www.gavinjensen.com/blog/2018/ai-narrow-vs-general (May 16, 2024)

Understanding and managing these technologies are essential to ensure that AI progress contributes positively to society.

Conclusion

In this chapter, the historical evolution of artificial intelligence has been explored, analysing the challenges faced and the successes achieved along the way. From Turing and his pioneering vision to the modern applications of machine learning and big data, we have observed how AI has traversed various stages of development, often accompanied by intense debates and evolving definitions.

One of the main difficulties that emerged is the very definition of AI. Defining AI poses a complex challenge, especially because it attempts to replicate or simulate something—human intelligence—that we ourselves do not yet fully comprehend. This ambiguity is reflected in the debate between strong and weak AI, exploring to what extent machines can truly emulate human intelligence and operate autonomously.

Despite this intrinsic complexity, the landscape of AI definitions is diverse, with various proposals formulated by international and national institutions, each characterised by a distinctive approach in attempting to delineate the boundaries and potentials of this field. Some of these definitions focus on specific and delimited aspects of AI, while others adopt a broader and more inclusive perspective.

Simultaneously, the introduction of concepts such as machine learning has profoundly transformed AI, enabling machines to learn and improve from experiences without being explicitly programmed for every scenario. Alongside big data, which provide vast amounts of information for analysis, these developments have significantly expanded AI's capabilities in decision-making and solving complex problems.

Looking to the future, it is evident that AI will continue to evolve rapidly, with an increasingly profound impact on society. However, beyond the technological potential, it is crucial to consider the ethical, social, and political implications associated with its development. The next chapter will address and examine the regulations proposed by various global actors, driven by the need for a regulatory framework to address potential risks and ensure responsible development and use of artificial

intelligence. This approach aims to promote AI use that is beneficial for all of humanity, maintaining a balance between technological innovation and socio-ethical considerations.

CHAPTER II

Navigating the Legal Terrain: Frameworks for Regulating Artificial Intelligence

Introduction

In our increasingly digital world, where we interact daily with intelligent and advanced systems, there is a pressing need to develop a robust and comprehensive regulatory framework. This framework must achieve two primary objectives: first, to ensure the protection and safety of individuals who engage with these technologies, safeguarding their rights and mitigating potential risks; and second, to stimulate and foster progress and innovation in the field of artificial intelligence, creating a regulatory environment that does not stifle technological development but guides it in a responsible and sustainable manner.

Innovations and advancements in this sector can rapidly disseminate and have a global impact, making the AI market a borderless reality. However, when examining the regulatory context, the landscape changes significantly.

The global AI market is divided into three main areas of regulatory influence: the European approach, which places strong emphasis on data protection and individual rights; the US approach, characterised by a more liberal and market-oriented stance; and the Chinese approach, which adopts a centralised, state-controlled strategy. This regulatory fragmentation presents a significant challenge for creating a harmonised and coherent global regulatory framework, yet it also offers diverse perspectives and governance models that can enrich the international debate on AI.

This chapter aims to examine in detail the various regulatory proposals aimed at governing artificial intelligence, focusing primarily on the regulatory paths adopted by the three main global players: the European Union, the United States, and China. The regulatory approaches of each jurisdiction will be thoroughly analysed, highlighting their specificities and main objectives.

Additionally, the chapter will briefly explore the legislative initiatives of other global actors who, in different but complementary ways, seek to establish guidelines to

regulate the use of artificial intelligence. The focus will be on measures adopted to assess the ethical implications of AI and to promote the responsible development and use of this emerging technology on a global scale.

2. 1 The European Context

2.1.1 Paving the Way for AI Regulation

The European Union, ever since it realised the importance and impact that these new technologies have on human beings and their lives, has always tried to create a regulatory apparatus that takes into account above all the human and ethical implications emerging in the digital age.[70] In light of the social and legal issues that have arisen in this area, several initiatives have been launched to create clear and comprehensive regulations. The European Parliament Resolution of 16 February 2017, which contains recommendations to the Commission for civil law regulations on robotics, highlights these issues. The resolution, entitled "*Civil Law Rules on Robotics"*[71], addresses topics that, starting from literature and science fiction, have become relevant topics of discussion for contemporary society, now perceived as pressing[72]. This Resolution, starting from general reflections on robotics and new technologies in general, introduces some reflections put forward by the European Parliament on issues that concern various fields such as ethics, law, economics, safety, labour, the environment and important considerations regarding the issue of responsibility.[73]

Since the European Parliament Resolution, there have been further developments in the field of AI at the European level, aimed at establishing rules and guiding principles for new technologies. For example, with the Communication of 25

[70] P. MORO, *Intelligenza artificiale e tecnodiritto. Fondamenti etici ed innovazione legislativa,* in P. MORO (eds), *Etica, Diritto e Tecnologia. Percorsi dell'informatica giuridica contemporanea,* Franco Angeli, 2021, pp. 7-24

[71] *Civil Law Rules on Robotics, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*

[72] P. L. DI VIAGGIANO, *Etica, Robotica e Lavoro: Profili D'Informatica Giuridica*, in Revista Opinião Jurídica, Vol. 16, no. 22, 2018, pp. 247-266

[73] *Ibidem*

April 2018 entitled "*Artificial Intelligence for Europe*"[74], the European Commission has encouraged and promoted the necessary regulation to address emerging AI-related issues. Furthermore, in June 2018, the European Alliance for Artificial Intelligence was established, a platform for thousands of stakeholders to discuss the technological and societal implications of AI online. Finally, in April 2019, the European Commission endorsed the basic requirements set out in ethical guidelines developed by an expert group to ensure trustworthy AI. According to the guidelines, trustworthy AI should be: lawful, ethical and robust[75].

In addition, on 19 February 2020, the European Commission issued the White Paper on Artificial Intelligence "*A European approach to excellence and trust*"[76]. This paper aims to promote the adoption of trustworthy AI while addressing the risks associated with this technology, such as lack of transparency in decision-making processes, gender or other discrimination, and intrusions into individual privacy. Given the centrality that AI has and will continue to have in the years to come for humanity, the reliability of this technology is not only a prerequisite, but also a prerequisite for its deployment. It is essential that AI develops with respect for European values and fundamental rights, such as human dignity and privacy.

Subsequent EU regulatory interventions also follow this direction. The conclusions of 21 October 2020 "*The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change*" [77], aim to secure the EU's fundamental rights and values in the age of digitisation, to promote EU digital sovereignty and to actively contribute to the global debate on the use of artificial intelligence in order to shape the international framework.

With the Communication entitled "*2030 Digital Compass: The European way for the Digital Decade"*[78] of 9 March 2021, the European Commission presented a

---

[74] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and Committee of the Regions, Artificial Intelligence for Europe,* COM(2018) 237 final, April 25, 2018

[75] High-Level expert group on Artificial Intelligence, European Commission, *Ethics Guidelines for Trustworthy AI*, April 8, 2019

[76] *WHITE PAPER, On Artificial Intelligence - A European approach to excellence and trust,* February 19, 2020, COM(2020) 65 final

[77] Presidency Conclusions, *The Rights in the context of Artificial Intelligence and Digital Change*, October 21, 2020

[78] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and Committee of the Regions, 2030 Digital Compass: the European way for the Digital Decade, COM(2021) 118 final,* March 9, 2021

vision and outlook for Europe's digital transformation by 2030 in order to ensure that AI is developed in a way that respects people's rights and makes Europe ready to operate in the digital age.

Ultimately, the proposal for a regulation of the European Parliament and of the Council establishing harmonised rules on AI was presented on 21 April 2021.

2.1.2 The European AI Act: A Milestone in Governance

The first proposal for the AI Act was presented by the European Commission in April 2021 as a political commitment of President von der Leyen. In her political programme for 2019-2024, entitled *"An Union that strives for more"*[79]. President von der Leyen declared that the Commission would put forward a regulation for a harmonised European approach on the humane and ethical use of artificial intelligence. In December 2022, the European Council adopted its common position. Subsequently, in June 2023, after the adoption of some amendments by the European Parliament, the legislative act entered the phase of "trilogues", the final negotiations. These negotiations involved three legislative bodies: the European Commission, the Council and the European Parliament, with the intention of developing a final version of the AI Act by the end of 2023 or the beginning of 2024. On 13 March 2024, the European Parliament formally approved the EU AI Act with a large majority of 523 votes in favour and 46 against[80], and on 1 August 2024, the law came into force.

The three legislative bodies - Commission, Council and Parliament - had to develop an agreed text through the EU decision-making process. Due to the numerous differences between the Parliament, which prioritises the democratic legitimacy of European law, and the Council, consisting of the representatives of the EU governments, several scepticisms emerged concerning a perfect balance between the protection of rights and the protection of the economy and social interests of the member states. The areas around which the tension between the European institutions

[79] U. VON DER LEYEN, *A Union that strives for more, My agenda for Europe, Political Guidelines for the next European Commission 2019-2024*, European Commission, 2019.
[80] M. FAZLIOGLU, *Contentious areas in the EU AI Act trialogues*, IAPP, 2023 https://iapp.org/news/a/contentious-areas-in-the-eu-ai-act-trilogues/ (May 21, 2024)

was greatest were several, including: the definition of AI, the list of prohibited AI applications (such as, for example, the use of this technology for biometric surveillance in publicly accessible spaces), high-risk AI obligations, core models and governance. Furthermore, the AI law is a proposal for a regulation. For this reason, when it comes into force, it will be directly applicable and immediately enforceable in the Member States[81]. Therefore, the definitions it contains become particularly critical, as they will not be subject to differences between national implementing regulations.

One of the crucial points of discussion was the clarification of the definition of AI. Initially proposed by the Commission in the appendix, the Council later downgraded its role, while the Parliament agreed to incorporate it directly into the body of the text, in Article 3. Furthermore, as we have already seen above, the Parliament saw fit to align the definition of AI with that proposed by the OECD in 2019. This change was motivated by the need to avoid ambiguities and legal uncertainties that could undermine fundamental rights.[82] Therefore, all computational systems used in the identified high-risk sectors, regardless of whether they are considered AI or not, have been included in the concept of "AI systems".[83]

Regarding the definition of high-risk AI, the Parliament favoured an extension of the criteria to include a broader range of systems, while the Council preferred a narrower definition. The Parliament, pressed by the urgency to pass the European law on AI within the year, felt that a broader definition could facilitate a greater consensus among Member States. Such an approach is also considered more appropriate to deal with future advanced technologies. On the contrary, the Council had to respond to civil society concerns about potential human rights violations. The current definitions have two critical issues: too broad a definition could be vague and general, leaving room for interpretation; on the other hand, more precise definitions could undermine the effectiveness of the law and exclude future developments in AI[84].

To balance the transformative potential of artificial intelligence on society and the economy, with its significant benefits, and the risks it poses to fundamental rights,

---

[81] Art 288, *Treaty on the Functioning of the European Union TFUE*
[82] T. MADIEGA, *EU Legislation in Progress, Artificial Intelligence Act*, EPRS - European Parliamentary Research Service, 2024
[83] European Parliament, *Legislative Train, A Europe fit for the digital age*, 2024
[84] *European Commission of 2021 on proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (artificial intelligence act) and amending certain union legislative acts, COM/2021/206 final, April 21, 2021*

security and the smooth functioning of the single market, the AI Act was developed following a risk-based approach. Thus, artificial intelligence systems are classified according to the level of risk they pose to individuals and society. This classification system identifies four categories of risk: unacceptable, high, limited and minimal.

The AI systems that present an unacceptable risk are those that contradict fundamental EU values and principles, such as respect for human dignity, democracy and the rule of law. In the final moments of the negotiations, MEP Dragoș Tudorache revealed in an interview that there were still some problems with the AI Act's prohibitions article. The Council was calling for exemptions for national security and law enforcement, while the European Parliament was advocating a stricter approach, proposing an outright ban on facial recognition technology in public spaces[85]. In the final version of Article 5, the following are prohibited:

- "AI systems that use harmful manipulative 'subliminal techniques';
- AI systems that exploit the vulnerabilities of a natural person or a specific group of persons;
- AI systems used to assess or classify natural persons or groups of persons; Systems used to conduct risk assessments of natural persons;
- the use of AI systems that create or extend facial recognition databases; the use of biometric categorisation systems that individually classify natural persons on the basis of their biometric data in order to infer or deduce their race, political opinions, trade union membership, religion or philosophical beliefs, sexual life or sexual orientation;
- 'real-time' remote biometric identification systems in publicly accessible locations for law enforcement purposes, except in a limited number of circumstances such as:

i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons;

[85] J. FLAMING-JONES, EU Policy. *EU AI Act nearing agreement despite three key roadblocks – co-rapporteur*, Euronews.next https://www.euronews.com/next/2023/10/23/eu-ai-act-nearing-agreement-despite-three-key-roadblocks-co-rapporteur (May 21, 2024)

ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;

iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years."[86]

The most controversial issue has been that of high-risk AI systems. High-risk AI systems are those that can have a "systemic", that is, significant impact on fundamental rights or the security of individuals. These systems are subject to strict obligations and requirements before they can be placed on the market or used.

The definition and limitation of high-risk AI systems was the subject of heated debate among the co-legislators, leading to many changes compared to the Commission's initial proposal. The main problem was to find a balance: on the one hand, there was a desire not to expand the list of high-risk uses too much in order to avoid additional burdens on companies; on the other hand, the European Parliament insisted that controls should not be reduced, ensuring that certain applications could not be used without due guarantees and precautions.[87] Article 6 Para. 2[88] of the AI Act makes a reference to Annex III of the text, which lists a number of use cases deemed "high risk" by the European legislator. These top-level categories of high-risk AI systems subject to the most stringent obligations under the Act are: Critical infrastructure; Biometric identification of natural persons; Educational and vocational

[86] Art. 5 (1) (h), *EU Artificial Intelligence Act*, 2024, Prohibited AI Practices

[87] R. PANETTA, *AI Act, requisiti e obblighi per i sistemi ad alto rischio: tutto quello che c'è da sapere,* in Agenda Digitale, 2024 https://www.agendadigitale.eu/industry-4-0/ai-act-requisiti-e-obblighi-per-i-sistemi-di-ia-ad-alto-rischio-tutto-quello-che-ce-da-sapere/ (May 21, 2024)

[88] Art. 6, *EU Artificial Intelligence Act*, 2024: "*1. Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered high-risk where both of the following conditions are fulfilled: (a) the AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonisation legislation listed in Annex II; (b) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonisation legislation listed in Annex II. 2. In addition to the high-risk AI systems referred to in paragraph 1, AI systems referred to in Annex III shall also be considered high-risk*."

training; Employment/workforce management; Essential private and public services; Law enforcement; Border control; Administration of justice and democratic processes.[89]

The big point scored by Parliament in the negotiations was the introduction of the Fundamental Rights Impact Assessment (FRIA). The compromise reached envisages it as mandatory for public entities and private entities offering public services. Before their public use, the likely consequences for those particular categories of people at risk and the solutions to be adopted in terms of human control and internal organisation will have to be identified.[90] Therefore, high-risk AI providers are required to implement a risk management system during the life cycle of the AI. They must primarily identify risks to health, safety and fundamental rights. For risks that cannot be eliminated, mitigation solutions must be provided. It is essential to inform and, if necessary, train the users of the system. Particular attention must be paid to children and vulnerable people in risk management.

The AI Regulation also imposes obligations on low-risk AI systems, i.e. those systems that can influence users' rights or choices, but to a lesser extent than high-risk systems. This category includes most AI systems. The regulation specifies that the production and use of such systems, which pose a limited risk to the rights and freedoms of individuals, will be subject to simple transparency obligations.

Specifically, article 52[91] stipulates that systems that interact with individuals, systems for emotion recognition and biometric categorisation that are not prohibited, and systems that generate or manipulate "deep fake" content, must clearly inform the

[89] Annex III, *EU Artificial Intelligence Act*, 2024
[90] R. PANETTA, 2024, op. cit.
[91] Art 52, *EU Artificial Intelligence Act*, 2024, *Transparency obligations for certain AI systems: "Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use. This obligation shall not apply to AI systems authorised by law to detect, prevent, investigate and prosecute criminal offences, unless those systems are available for the public to report a criminal offence. Users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences. Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated. However, the first subparagraph shall not apply where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties. Paragraphs 1, 2 and 3 shall not affect the requirements and obligations set out in Title III of this Regulation."*

user that they are interacting with an AI system or that a particular piece of content has been created through AI. This transparency obligation is crucial to enable users to use the technology in an informed and knowledgeable manner.[92]

Finally, AI systems that present little or no risk are those that have no direct impact on fundamental rights or the security of individuals, and that offer ample room for choice and control to users. These systems are free from any regulatory requirements in order to encourage innovation and experimentation.

Another major change introduced by Parliament during the negotiations was the imposition of a specific regime for foundation models, which are classified as high-risk systems. These models, defined in the Artificial Intelligence Act as "General Purpose AI Models" (GPAI), are computer models that, partly due to training on a vast amount of data, can be used for a variety of tasks, either individually or as components of an AI system.[93] Their ability to serve multiple purposes and their centrality in the expanding market of AI-based systems and applications have given foundation models a crucial role in the artificial intelligence debate. The growing popularity of these models, such as Open AI's GPT and Google's LaMDA, together with the associated risks and the perceived significant change in AI development, has stimulated intense public and political debate.[94]

During the legislative process, the Council and the Parliament took different approaches to the regulation of foundation models. While the Council opted for an initially lighter regulatory framework, with the possibility of introducing stricter rules following an analysis by the European Commission, the Parliament supported the imposition of stringent rules from the outset. This divergence has contributed to some

[92] I. DE FEO, A. AFFERNI, *AI Act: il Regolamento sull'Intelligenza Artificiale adottato dal Parlamento UE,* 14 March 2024 https://www.dirittobancario.it/art/ai-act-il-regolamento-sullintelligenza-artificiale-adottato-dal-parlamento-ue/ (May 28, 2024)

[93] Art 3, 63, *EU Artificial Intelligence Act*, 2024, "*GPAI model means an AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications. This does not cover AI models that are used before release on the market for research, development and prototyping activities"*

[94] M. BORGOBELLO, *AI Act: ecco come regolerà l'intelligenza artificiale generativa*, in Agenda Digitale, February 5, 2024, https://www.agendadigitale.eu/mercati-digitali/ai-act-ecco-come-regolera-i-foundation-model/ (May 29, 2024)

uncertainty in the debate, complicating the measurement of the distance between the two regulatory approaches.[95]

The adopted AI Act distinguishes between two categories of models: generic GPAI models and "systemic" GPAI models[96]. The latter, because of the "systemic risks" they may entail at the European level, are subject to stricter regulation than the generic ones. Generic GPAI models are subject to transparency requirements[97], which include the availability of detailed technical documentation to make their operation, including data training processes, understandable to the European AI Office and to third parties interested in integrating these models into their systems. Such regulation is considered reasonable and should not be a significant obstacle to the development of these models. Furthermore, anyone wishing to market a foundation model in the EU must appoint a representative in the territory.[98] There is also an obligation to adopt a policy that respects copyright law.[99]

Systemic GPAI models, in addition to the obligations of generic models, must comply with additional requirements, making regulation more pervasive. These obligations include: (a) the evaluation of the model according to standardised protocols, including the conduct and documentation of 'adversarial testing' in order to identify and mitigate systemic risk; (b) the assessment and mitigation of possible EU-wide systemic risks arising from the development, marketing or use of AI models with systemic risk (c) the tracking, documentation and timely reporting of serious incidents and corrective measures to the European AI Bureau and relevant national authorities; and (d) ensuring an adequate level of cybersecurity protection of the model and its physical infrastructure[100].

Undoubtedly, the Commission has managed to build itself an absolutely prominent role in the management of foundation models within the AI Act, with

---

[95] I. GENNA, *The regulation of foundation models in the EU AI Act,* in International Bar Association, April 12, 2024 https://www.ibanet.org/the-regulation-of-foundation-models-in-the-eu-ai-act (May 29, 2024)
[96] Art 51, *EU Artificial Intelligence Act*, 2024, *Classification of General-Purpose AI Models as General-Purpose AI Models with Systemic Risk*
[97] Art 53, *EU Artificial Intelligence Act*, 2024, *Obligations for Providers of General-Purpose AI Models*
[98] Art 54, *EU Artificial Intelligence Act*, 2024, *Authorised Representatives of Providers of General-Purpose AI Models*
[99] Article 53(1)(c), *EU Artificial Intelligence Act*, 2024, *Obligations for Providers of General-Purpose AI Models*
[100] Art 55, *EU Artificial Intelligence Act*, 2024, *Obligations for Providers of General-Purpose AI Models with Systemic Risk*

significant competences not only in regulatory and enforcement matters, but also in the sector's industrial policy. In particular, the Commission will have exclusive competence and wide discretion in the enforcement phase of the system, including the identification of systemic GPAI models, a phase feared by some European governments because it entails the automatic application of the obligations under the regulation. The Commission will also have a key role in overseeing the Codes of Conduct to ensure compliance[101].

The last issue negotiated during the trialogues was the clarification of governance, concerning the enforcement of the AI law and coordination between the various national and EU authorities. The new regulation establishes, in Title VI, a governance framework with the aim of coordinating and supporting national enforcement.

In particular, an Office for AI[102] will be established within the Commission, with a strong link to the scientific community to support its work. The AI Office will oversee the most advanced AI models, help promote standards and testing practices, with common rules in all Member States, and will be endowed with a range of administrative, advisory, interpretative and enforcement powers, as well as responsibility for coordinating cross-cutting activities. The European AI Office will be an independent body of the Union and will have legal personality. It will be the centre of expertise for AI across the EU and will play a significant role in the implementation of the AI Act[103].

Since the structure of the Artificial Intelligence Act is similar to that of the General Data Protection Regulation (GDPR)[104], it will assign various competences to

[101] M. BORGOBELLO, 2024, op. cit.
[102] Art 56, *EU Artificial Intelligence Act*, 2024, *Establishment of the European Artificial Intelligence Board*
[103] *European Parliament of 2023 on Artificial Intelligence Act on amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). Amendment 122, Proposal for a regulation, Recital 76*
[104] The regulation was drafted and adopted by the European union and put into effect on May 25, 2018. It is a regulation governing the way in which companies and other organisations process personal data. The legislation aims to give every individual control over the use of his or her data, protecting the 'fundamental rights and freedoms of natural persons'. With this in mind, the regulation establishes precise and strict requirements for data processing, transparency, documentation and user consent for organisations processing personal data in the European Union.

national law enforcement actors in an Artificial Intelligence Committee, similar to the European Data Protection Board.[105]

At the national level, according to the text of the AI Act, each Member State will have to establish national authorities with the competences assigned by the Regulation. These authorities will be responsible for enforcing sanctions for violations of the AI Act. The national authorities will have to operate independently, impartially and without bias, and will have to be provided with the necessary technical, financial, human and infrastructural resources to effectively fulfil their tasks.

The EU's AI Regulation is set to become the world's first comprehensive horizontal legislative instrument for artificial intelligence. The Regulation is seen as a sea change for the regulation of artificial intelligence, just as the EU's General Data Protection Regulation was for the regulation of data protection a few years ago.[106] The impact of the new document will not stop at EU borders. In fact, some EU politicians believe it is a key goal of the AIA to establish a global standard, so much so that some talk of a race to regulate AI. This framework implies that it is not only useful to regulate AI systems, but that being among the first major governments to do so will have a broad global impact to the benefit of the EU, often referred to as the "Brussels effect". However, even if some components of the AIA will have major effects on global markets, Europe alone will not be able to establish a comprehensive new international standard for AI.[107]

2.1.3 Shaping the Future: The Council of Europe's AI Convention

Understanding the need for democratic oversight in driving AI innovation, the Committee of Ministers of the Council of Europe, in September 2019, established the terms of reference for the Ad Hoc Committee on Artificial Intelligence (CAHAI)[108]. The role of this committee was to explore the feasibility and core elements of a legal

[105] M. FAZLIOGLU, 2023, op. cit.
[106] *Ibidem*
[107] A. ENGLER, *The EU AI Act will have global impact, but a limited Brussels Effect*, in Brookings, 2022 https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/ (May 21, 2024)
[108] https://www.coe.int/en/web/artificial-intelligence/cahai

framework that governs the design, development, and deployment of AI systems, ensuring that they are in line with the principles of human rights, democracy and the rule of law of the Council of Europe.

As a key step, CAHAI's feasibility study[109], approved in December 2020, analysed possible international legal responses to fill legislative gaps. This study also concluded that existing legal structures are inadequate to safeguard these values in the context of AI and to create a reliable environment for AI and data-driven technologies. Therefore, they considered the creation of a new legal framework to be necessary[110].

After CAHAI completed its mandate from 2019-2021, it was succeded by the Committee on Artificial Intelligence (CAI)[111]. Since 2021 the CAI has been working on drafting a Convention that guides the development, deployment, and use of AI systems based on the Council's human rights, rule of law, democracy and innovation-friendly standards. The AI Convention takes a comprehensive approach to ensure that the use of artificial intelligence does not undermine the fundamental principles and rights set out in the ECHR.

In February 2023, the CAI decided to release a revised version of the *"Zero Draft" Framework Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law*[112]. By July 2023, a consolidated working draft of this framework convention had been published[113]. At its eighth meeting in December 2023, the CAI decided to make public the *draft Framework Convention*[114] incorporating the results of the second reading. Meanwhile, the Convention remained open to new ideas and articles, which could only be proposed by member states, as civil society was excluded by United States representatives at the beginning of 2023.[115]

---

[109] Council of Europe, Ad Hoc Committee on Artificial Intelligence, *Feasibility Study*, Strasbourg, 17 December 2020

[110] D. LESLIE et al, 2021, op. cit., p. 6

[111] https://www.coe.int/en/web/artificial-intelligence/cai

[112] *Council of Europe, Committee on Artificial Intelligence (CAI), Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law,* Strasbourg, 6 January 2023

[113] COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI), *Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, July 2023

[114] COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI), *Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law,* December 2023

[115] L. BERTUZZI, *US obtains exclusion of NGOs from drafting AI treaty,* EURACTIV, 2023, https://www.euractiv.com/section/digital/news/us-obtains-exclusion-of-ngos-from-drafting-ai-treaty/ (June 4, 2024)

The CAI finalised the text of the *Draft Framework Convention on AI, Human Rights, Democracy, and the Rule of Law*[116] during their March 2024 meeting. This document is built around several key principles governing AI system activities throughout their lifecycle, including: respect for human dignity and autonomy[117], transparency and oversight[118], accountability and responsibility[119], equality and non-discrimination[120], privacy and personal data protection[121], reliability[122], and the promotion of safe innovation[123].

On 17 May, at its 133rd session in Strasbourg, the Committee of Ministers approved and adopted the *Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*. This Convention will be opened for signature at the Conference of Ministers of Justice to be held in Vilnius, Lithuania, on 5 September 2024.

The EU AI Act aims to regulate products using artificial intelligence in the EU internal market, while the AI Convention intends to protect the fundamental rights of persons involved and affected in AI systems. While the EU will directly implement the AI Act through its member states, the AI Convention will be based on principles and establish legally binding individual human rights, applicable to both EU member states and non-member states that choose to ratify and sign the Convention in the future.[124]

Council of Europe Secretary General Marija Pejčinović remarked: "The Framework Convention on Artificial Intelligence is a first-of-its-kind, global treaty that will ensure that Artificial Intelligence upholds people's rights. It is a response to the need for an international legal standard supported by states in different continents which share the same values to harness the benefits of Artificial intelligence, while mitigating the risks."[125]

---

[116] *Draft Framework Convention on artificial intelligence, human rights, democracy and the rule of law*, March 2024 https://rm.coe.int/-1493-10-1b-committee-on-artificial-intelligence-cai-b-draft-framework/1680aee411 (June 4, 2024)
[117] *Ibidem,* Art. 7
[118] *Ibidem,* Art. 8
[119] *Ibidem,* Art. 9
[120] *Ibidem,* Art. 10
[121] *Ibidem,* Art. 11
[122] *Ibidem,* Art. 12
[123] *Ibidem,* Art. 13
[124] Committee of Minsters, *Council of Europe adopts first international treaty on artificial intelligence*, 17 May 2024, https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence (June 4, 2024)
[125] *Ibidem*

2.2 The United States Approach

Currently, there is no comprehensive federal legislation that regulates the development of artificial intelligence or specifically prohibits or limits its use. However, there are some federal guidelines and protections in place.

The United States has embarked on a course of regulation of artificial intelligence characterised by considerable "levity". This approach has been motivated by the need to preserve economic and industrial development by avoiding restrictive interventions that could slow down or limit innovation with excessive procedures and controls[126]. One of the main motivations of US policymakers is to ensure that the United States plays a leading global role in the development and implementation of artificial intelligence. To this end, Washington has been reluctant to adopt or even propose a radical EU-style regulatory regime governing AI applications and oversight, for fear of slowing down innovation.

The Trump administration in February 2019 issued Executive Order 13859 "*Maintaining American Leadership in Artificial Intelligence*" [127] which introduced the US Artificial Intelligence Strategy, the four main objectives of which can be traced back to the promotion of research and development, the creation of prerequisites for greater public confidence in AI applications, the training of a skilled workforce capable of benefiting from the use of AI, and the protection of the US technology sector from takeover attempts and possible attacks by competitors and foreign countries[128].

Following the directives of the executive order, in August 2019, the National Institute of Standards and Technology (NIST) presented its Plan for Federal Engagement in Developing Technical Standards. The plan identifies areas of technical standards for AI and non-technical standards for AI that inform policy decisions, such as "social and ethical considerations," "governance," and "privacy"[129]. Significantly, in the context of developing standards for the social and ethical considerations of AI, it is recognised that it is important to distinguish between technical and non-technical

[126] E. STRADELLA, *Le fonti nel diritto comparato: analisi di scenari extraeuropei (Stati Uniti e Cina)*, in "DPCE Online", vol. 51, 2022, p. 233-234, https://www.dpceonline.it/index.php/dpceonline/article/view/1569 (July 3, 2024)
[127] Executive Order No. 13859, Federal register, vol. 84, N. 31, 11 February 2019.
[128] *Ibidem,* sect 1
[129] *Ibidem,* p. 12

standards. This is because not all AI-related issues involving social and ethical aspects can be fully addressed through the development of technical standards.

In recent years, several laws have been proposed to regulate artificial intelligence. Among the most impactful proposals is the "*Algorithmic Accountability Act"*[130] of April 2019. This act requires specific commercial entities to conduct assessments of high-risk systems that use personal information or make automated decisions using artificial intelligence or machine learning. It is considered the first federal legislative attempt to regulate AI across all industries.

More recently, two significant bills have been introduced: the "*Algorithmic Justice and Online Platform Transparency Act*"[131] and the "*Artificial Intelligence Training Act*"[132]. The former introduces new disclosure requirements for online platforms regarding the algorithmic processes used to customise content or services for users. Its goal is to enhance transparency and accountability in algorithm usage, particularly in influencing user decisions. The latter aims to ensure that the workforce is well-informed about the capabilities and risks associated with artificial intelligence.

In addition, another important example is the "*National AI Initiative Act*" of 2020 (NAIIA)[133], which was last updated in 2023. This legislation aims to support research and development in AI and established the National Artificial Intelligence Initiative Office. This office is tasked with overseeing and implementing the U.S. national strategy on AI, complementing efforts to regulate and promote transparency in the use of AI technologies across various sectors.

[130] Congress.gov, H.R.2231, 116th Congress, Algorithmic Accountability Act of 2019.
[131] H.R. 3611, *"Algorithmic Justice and Online Platform Transparency Act",* 117th Cong., 28 May 2021
[132] Congress.gov. S.2551, 117th Congress, *AI Training Act*, Introduced 07/29/2021.
[133] H.R.6216, National Artificial Intelligence Initiative Act of 2020, 116th Congress.

2.2.1 Soft-Law Instruments: The U.S. AI Bill of Rights and The Risk Management Framework

In October 2022, the Office of Science and Technology Policy (OSTP) of the White House published the "*Blueprint for an AI Bill of Rights*"[134]. This document resulted from collaboration among the OSTP, academics, human rights groups, the general public, and major companies such as Microsoft and Google[135]. Its purpose is to guide the design, development, and implementation of artificial intelligence and other automated systems, aimed at protecting the rights of American citizens.

The blueprint applies to automated systems that can have a significant impact on the rights, opportunities, or access to essential resources or services of individuals and communities. It consists of five core principles and related practices that will guide the design, use, and development of AI algorithms that operate using biometric data such as facial recognition, fingerprints, retinal screening, and DNA. The goal is to reduce discrimination, protect citizens, and define concrete steps that companies and governments must take to develop algorithms that promote economic and social progress without compromising civil rights and democratic values.

The principles listed are as follows:

- Safe and Effective Systems[136]: Emphasizes that artificial intelligence systems should be developed after careful consultation with experts and auditors to identify risks and potential impacts. Each AI system should undergo testing during development, risk identification, mitigation, and continuous monitoring to demonstrate its safety and efficiency. If tested systems reveal dangerous outcomes for citizens or fail to meet government-imposed standards, they should not be used.

---

[134] The White House Office of Science and Technology Policy (OSTP), *Blueprint for an AI Bill of Rights: Making Automated Systems Works for the American People,* October 2022, The White House, Washington, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf (July 3, 2024)

[135] E. GLOVER, *AI Bill of Rights: What You Should Know,* in Builtin, 2024 https://builtin.com/artificial-intelligence/ai-bill-of-rights (July 3, 2024)

[136] The White House Office of Science and Technology Policy (OSTP), *Blueprint for an AI Bill of Rights: Making Automated Systems Works for the American People,* October 2022, The White House, Washington, p. 5, https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf (July 3, 2024)

- Algorithmic Discrimination Protections[137]: Refers to discrimination, defined as treatment that unjustifiably disadvantages individuals based on ethnicity, gender, sexual orientation, age, disability, or any other protected category under the law. Developers of AI systems should take continuous and proactive measures to protect individuals and communities from algorithmic discrimination. These measures should include fairness assessments, the use of representative data from society as a whole, and public evaluation of the impact algorithms have on society.
- Data Privacy[138]: Data collected and used by algorithms must always include the consent of citizens. Digital surveillance should not be used in schools, workplaces, homes, or other contexts where its use could limit individual rights and freedoms.
- Notice and Explanation[139]: Developers of AI systems must provide clear documentation that includes a description of the general functioning of these systems and the circumstances in which they are used.
- Human Alternatives, Consideration, and Fallback[140]: Individuals subjected to the use of an AI system should, where appropriate, be able to refuse automated systems and have access to a person who can evaluate and address issues encountered. The option to receive a human assessment and remedy should be accessible, fair, effective, accompanied by adequate training of operators, and should not impose an unreasonable burden on the public. Particularly in sensitive areas such as criminal justice, employment, education, and health, automated systems should ensure the possibility of obtaining a human assessment in cases of adverse or high-risk decisions.

The technical appendix provides a detailed explanation of each principle's relevance through concise summaries and concrete examples of the issues each principle aims to address. It also illustrates practices and technical standards, such as the use of independent assessments, which businesses, public administrations, and other organisations can adopt to implement these principles. These practices and standards

[137] *Ibidem*
[138] *Ibidem,* p. 6
[139] *Ibidem*
[140] *Ibidem,* p. 7

serve as guidelines for the overall design of technology. Additionally, the technical appendix describes how the principles can be applied in practice, with particular attention to their interaction with existing laws and policies, which are not modified, replaced, or reinterpreted in light of the Blueprint for an AI Bill of Rights.

In the glossary of terms included in the Appendix of this document, a detailed definition of "automated system" is provided. According to this definition, an automated system is described as:

"*Any system, software, or process that uses computation as whole or part of a system to determine outcomes, make or aid decisions, inform policy implementation, collect data or observations, or otherwise interact with individuals and/or communities. Automated systems include, but are not limited to, systems derived from machine learning, statistics, or other data processing or artificial intelligence techniques, and exclude passive computing infrastructure. "Passive computing infrastructure" is any intermediary technology that does not influence or determine the outcome of decision, make or aid in decisions, inform policy implementation, or collect data or observations, including web hosting, domain registration, networking, caching, data storage, or cybersecurity. Throughout this framework, automated systems that are considered in scope are only those that have the potential to meaningfully impact individuals' or communities' rights, opportunities, or access."*[141]

It is crucial to emphasise that the document applies exclusively to automated systems that have the potential to significantly influence the rights, opportunities, or access of the American public to essential resources or services, generally excluding many industrial and operational applications of artificial intelligence.

Additionally, it should be emphasised that the Blueprint for an AI Bill of Rights issued by the White House does not carry legal binding authority. Rather, it serves as a set of recommendations aimed at advancing the protection of citizens' rights within the field of artificial intelligence. While these recommendations are not enforceable by law, they mark a significant step and a foundational initiative toward fostering ethical practices and ensuring accountability in AI development and deployment.

Another key document was adopted in January 2023. After several drafts and rounds of public consultation, the US National Institute of Standards and Technology

[141] *Ibidem,* p. 10

(NIST)[142] released its "*Artificial Intelligence Risk Management Framework*"[143] (AI RMF). This framework represents an approach to identify, assess and manage risks associated with AI, marking a significant step in the development of robust AI governance and risk mitigation strategies.

The document is divided into two main sections. The first section discusses the material scope of the document and the characteristics of reliable AI. In addition, the target audience and the concept of risk are examined.

The RMF focuses primarily on different types of risk that can be integrated into risk management. The framework begins with the identification of a risk, defining it as "the composite measure of an event's probability of occurring and the magnitude or degree of the consequences of the corresponding event"[144].

The identified risks must then be evaluated, a process that can be carried out by following specific criteria established by the US NIST in the RMF, referred to as the "characteristics of a trustworthy IA". These criteria include: Validation, Reliability, Accuracy, Robustness, Safety, Security, Resilience, Transparency, Accountability, Explainability, Interpretability, Privacy-enhancement, Fairness.

The second section, on the other hand, describes four categories of actions that framework users can take to concretely manage AI-related risks in line with the RMF. These actions include: governing, mapping, measuring and managing risks. Each action is divided into sub-categories, to which actors and activities are in turn assigned. The first function - govern - involves the implementation of a general policy for AI risk management, facilitating the other three functions[145]. The second function - map - focuses on identifying and defining risks in a specific context[146]. The third function – measure - concerns the assessment and analysis of the identified risks[147]. Finally, the management function aims to prioritise, monitor and resolve the risks present[148].

---

[142] The first draft of the "AI Risk Management Framework" document dates back to March 2022 and a second draft was released by NIST in August 2022.
[143] National Institute of Standards and Technology (NIST), U.S department of commerce, *Artificial Intelligence Risk Management Framework (AI RMF 1.0),* January 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf (July 3, 2024)
[144] *Ibidem,* p. 4
[145] *Ibidem,* pp. 21-24
[146] *Ibidem,* pp. 24-28
[147] *Ibidem,* pp. 28-31
[148] *Ibidem,* pp. 31-33

The AI Risk Management Framework is designed to support companies in the development and deployment of artificial intelligence systems by helping them assess and manage the risks associated with these technologies. The framework provides guidelines and recommendations that are voluntary in nature, which means they are not legally binding and should not be interpreted as mandatory regulations.

2.2.2 Executive Order on AI: Strengthening Governance

30 October 2023 marks a significant turning point for the evolution of artificial intelligence-based systems and products. On this date, President Biden signed Executive Order 14110, which introduces a series of measures to ensure the safe and regulated development of artificial intelligence, while protecting citizens from potential abuse. These initiatives aim to create a regulatory framework that balances technological innovation with the protection of people's rights and safety.

The Executive Order on the "*Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*" [149] defines a policy framework to manage the risks associated with artificial intelligence. It guides agency action for the regulation of AI systems and tools in the health sector and promotes AI innovation in all sectors, including health and human services. It sets new benchmarks for AI security, protecting the privacy of Americans, advancing equity and civil rights, and promoting competition and innovation[150].

The Executive Order (EO) establishes eight guiding principles and priorities to promote and regulate the use of AI:

(i) Artificial Intelligence must be safe and secure[151]: Cybersecurity, the use of biotechnology, critical infrastructure protection and post-implementation monitoring are considered essential in this theme.

---

[149] THE WHITE HOUSE, October 30, 2023, JOSEPH R. BIDEN JR., *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ (July 4, 2023)

[150] Stanford University, Human-Centered Artificial Intelligence HAI, *Artificial Intelligence Index Report 2024,* chapter 7: Policy and Governance, p. 9

[151] *Ibidem, section 2, (a)*

(ii) Promoting responsible innovation, competition, and collaboration will allow the United States to lead in AI and unlock the potential of technology to solve some of the most difficult challenges in society[152]: This effort requires investment in research, development, education, training and capabilities related to AI, as well as addressing new intellectual property issues.

(iii) The responsible development and use of AI require a commitment to supporting American workers[153]: Support American workers by creating new jobs and sectors, and incorporate collective bargaining to ensure that workers take advantage of these opportunities. Provide training and vocational education to support a diverse workforce, and ensure that the implementation of AI does not "undermine rights, worsen job quality, encourage undue surveillance of workers, reduce market competition, introduce new health and safety risks, or cause harmful workforce disruptions."

(iv) "Artificial Intelligence policies must be consistent with my Administration's dedication to advancing equity and civil rights"[154]: It is crucial that those who develop and use AI are obliged to adhere to standards that prevent unlawful discrimination and abuse, including in the judicial system and the federal government. Only in this way can Americans have confidence in AI as a tool to promote civil rights, civil liberties, fairness, and justice for all.

(v) The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected[155]: protect consumers' interests by enforcing existing consumer protection laws and taking safeguards against fraud, unintentional bias, discrimination, privacy violations and other potential harms resulting from artificial intelligence.

(vi) Americans' privacy and civil liberties must be protected[156]: Protect privacy and civil liberties by ensuring that the collection, use and storage of data are lawful and secure and reduce risks to privacy and confidentiality.

---

[152] *Ibidem, section 2 (b)*
[153] *Ibidem, section 2 (c)*
[154] *Ibidem, section 2 (d)*
[155] *Ibidem, section 2 (e)*
[156] *Ibidem, section 2 (f)*

(vii) It is important to manage the risks from the Federal Government's own use of AI and increase its internal capacity to regulate, govern, and support responsible use of AI[157]: The federal government is committed to providing comprehensive training to its employees to understand the benefits, risks and limitations of AI in their functions, as well as to modernising the IT infrastructure, overcoming bureaucratic hurdles and ensuring the safe and respectful adoption of AI.

(viii) The Federal Government should lead the way to global societal, economic, and technological progress[158]: This action includes engagement with global allies and partners to develop a framework to mitigate AI risks, unlock the potential of AI for good and join forces on shared challenges.

The extensive Order contains directives for almost all 15 executive departments, urging them to use their regulatory powers to monitor and mitigate the risks associated with artificial intelligence. In addition, the Order calls for the development of practical applications for AI technology and the safe implementation of such technologies.

The US government aims to develop a specific architecture for artificial intelligence that has clear regulatory and institutional references to guide the controlled and safe development of this technology. Starting with the fundamental principles, Sections 4 to 11 of the Order are structured to reflect each of the eight guiding principles. Each section outlines concrete policy objectives, specific tasks and detailed guidelines for federal agencies to implement over the next year. Among the various initiatives is the creation of "The United States AI Safety Institute" (US AISI), which, using the NIST AI Risk Management Framework as a reference, will create guidelines, tools and practices for risk assessment and mitigation in the use of AI[159].

The executive order also stipulates that, before releasing a new artificial intelligence system to the public, its creators must provide the federal government with the results of their security tests[160].

---

[157] *Ibidem, section 2 (g)*
[158] *Ibidem, section 2 (h)*
[159] The White House, *Fact Sheet: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence,* November 2023 https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harris-announces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/ (July 5, 2024)
[160] *Ibidem, section 4.2*

Starting from these foundational elements, there is a commitment to draft an initial version of specific AI legislation that integrates and harmonises, on one hand, the operational risk management content developed in the NIST AI Risk Management Framework, and on the other hand, incorporates the principles outlined in the Blueprint for an AI Bill of Rights.

2.3 China's AI Governance: Strategic Plans and Policies

China is recognised as a key player in the field of artificial intelligence, showing dedication not only to technological advancement and expanding market infrastructure but also to regulating the sector. Over the years, China has implemented a comprehensive range of legislative acts and state plans that prioritise the development of AI technologies.

The regulatory framework began to take shape in 2012 during the 18th National Congress of the Chinese Communist Party. It was during this event that the importance of rapidly integrating artificial intelligence into the economy, society, and national defense was underscored. The stated objective was to advance scientific development and technological innovation with the aim of cultivating a new generation of AI, promoting an intelligent economy, and establishing a smart society[161].

In 2015, the "*Made in China 2025*"[162], plan was launched, a decade-long state-led initiative aimed at transforming China into a global leader in high-tech manufacturing by 2025. This ambitious program sought to enhance industrial innovation, strengthen national production capabilities, and reduce dependence on foreign technology, positioning China as a global hub for production and technological innovation[163].

Until 2016, artificial intelligence was included among many other technologies mentioned in Chinese policy documents. While recognised as a valuable resource for

[161] Report of Hu Jintao to the 18th National Congress of the Communist Party of China, 8 November 2012, http://cpc.people.com.cn/n/2012/1118/c64094-19612151.html# (July 5, 2024)
[162] The State Council People of China, *Made in China 2025* plan issued, 2015 https://english.www.gov.cn/policies/latest_releases/2015/05/19/content_281475110703534.htm
[163] J. MCBRIDE, A. CHATZKY, *Is 'Made in China 2025' a Threat to Global Trade?, 2019* https://www.cfr.org/backgrounder/made-china-2025-threat-global-trade (July 5, 2024)

achieving political and social goals, AI had not yet gained clear distinction either technologically or, consequently, in regulations compared to other emerging technologies[164].

In 2017, China marked a significant turning point with the adoption of the "*Next Generation Artificial Intelligence Development Plan*" (AIDP)[165] by the State Council. This document represented a substantial declaration of intent, positioning China as an aspiring global leader in AI theories, technologies, and applications. The AIDP is considered the most comprehensive national strategy ever formulated on AI, encompassing ambitious initiatives and objectives for research and development, industrialisation, talent development, education and skill acquisition, as well as the establishment of standards, regulations, ethical norms, and security measures.

This plan set clear objectives through 2030, structured in three successive phases: first, to align China's AI industry with global competitors by 2020; second, to achieve global leadership in certain AI sectors by 2025; third, to become the world's leading centre for AI innovation by 2030. Additionally, the document outlines the Chinese government's intention to attract top international talent in AI, enhance domestic workforce training in AI, and take a leading role in shaping global laws, regulations, and ethical standards to promote responsible AI development. It emphasises the need to "Strengthen research on legal, ethical, and social issues related to AI, and establish laws, regulations and ethical frameworks to ensure the healthy development of AI".[166]

Similarly, in 2018, Xi Jinping underscored the importance of enhancing research capabilities, assessing, and preventing potential risks associated with the development of artificial intelligence. He emphasised the need to safeguard public interests and national security, ensuring that AI is reliable and under control. Xi Jinping highlighted the urgency of integrating interdisciplinary expertise and advancing studies on laws, ethics, and social issues related to AI. Furthermore, he proposed the creation and

[164] E. STRADELLA, 2022, op. cit., p. 222
[165] State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, *A Next Generation Artificial Intelligence Development Plan*, July 2017 https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf (July 5, 2024)
[166] State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, *A Next Generation Artificial Intelligence Development Plan*, July 2017, p. 25 https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf (July 5, 2024)

refinement of regulations, institutional systems, and ethical frameworks to promote the healthy development of artificial intelligence.[167]

At the World Conference on Artificial Intelligence held on July 9, 2021, the National Industrial Information Security Development Center officially released the first "*White Paper on Trustworhty artificial Intelligence*"[168]. This document presents a comprehensive framework for AI, thoroughly detailing the elements that should characterise trustworthy artificial intelligence. The reliability features of AI are summarised into five main aspects: "transparency, security, fairness, accountability and privacy".[169]

The paper argues that the concept of AI reliability now extends beyond mere definitions of AI technologies, products, and services, encompassing systematic methodologies that embrace all necessary stages to create trustworthy AI. Additionally, the White Paper examines various technologies that support AI reliability, such as system stability, enhanced explainability, privacy protection, and fairness.[170]

In September 2021, China's National Professional Committee for the Governance of the New Generation of Artificial Intelligence published the "*Code of Ethics for the New Generation of Artificial Intelligence*"[171] to integrate ethical principles into the AI life cycle and provide guidance for individuals, legal entities and other related institutions engaged in AI-related activities[172]. Article 1, in fact, states that: "*This specification aims to integrate ethics and morality into the entire life cycle of artificial intelligence, promote fairness, justice, harmony and security, and avoid problems such as bias, discrimination, privacy and information leakage."*[173] Furthermore, in Article 3, it sets out six fundamental ethical requirements that must guide all AI-related activities: i) Enhancing human welfare; ii) Promoting fairness and

---

[167] *Xi Jinping chaired the ninth collective study session of the Political Bureau of the CPC Central Committee and delivered a speech*, 2018 https://www.gov.cn/xinwen/2018-10/31/content_5336251.htm (July 5, 2024)

[168] China Academy of Information Technology, JD Explore Academy, *White Paper on Trustworhty artificial Intelligence, july 2021,* http://www.caict.ac.cn/english/research/whitepapers/202110/P020211014399666967457.pdf (July 7, 2024)

[169] *Ibidem,* p. 8

[170] *Ibidem,* pp 11-14

[171] *Code of Ethics for the New Generation of Artificial Intelligence*, *新一代人工智能伦理规范》发布,* 26 september 2021 https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html (July 7, 2024)

[172] *Ibidem,* art 2

[173] *Ibidem,* art 1

justice; iii) Protecting privacy and security; iv) Ensuring controllability and credibility; v)Enhancing accountability; and vi) Enhancing ethical literacy[174].

More recently, in March 2022, the General Office of the Central Committee of the Communist Party of China together with the General Office of the State Council issued the "*Opinions on Strengthening the Ethical Governance of Science and Technology*"[175]. This document aims to improve the ethical system in science and technology, strengthen governance, prevent and effectively manage ethical risks, and promote benefits to society. Five essential requirements have been outlined to improve ethical governance in science and technology: (i) integrate ethical principles into all phases of science and technology activities; (ii) accelerate the creation of a legal system for ethical governance in these fields; (iii) adapt governance practices and ethical standards to quickly and flexibly address the challenges posed by technological innovation; (iv) develop and improve a science and technology ethics system in line with country-specific conditions; and (v) adopt an open development approach, enhancing international exchanges and actively promoting global ethical governance in science and technology.

Beijing's approach to regulating artificial intelligence is becoming more structured and complex, reflecting an attempt to balance support for technological innovation with the need to control its associated risks. In July 2023, the Cyberspace Administration of China - the country's main Internet control and censorship system - published a set of guidelines to regulate the generative AI sector, the so-called "*Interim Measures for the Administration of Generative Artificial Intelligence Services*"[176], which came into force on 15 August.

The Measures aim to promote innovative and high-quality use of generative AI, while ensuring the protection of the intellectual property rights involved. They require providers to support the core values of socialism according to the Chinese model, guaranteeing intellectual property protection (IP protection), transparency, accuracy and

[174] *Ibidem,* art 3

[175] General Office of the CPC Central Committee and the General Office of the Council, *Opinions on strengthening the ethical governance of science and technology, 关于加强科技伦理治理的意见,* 2022, https://www.gov.cn/zhengce/2022-03/20/content_5680105.htm (July 7, 2024)

[176] *Interim Measures for the Administration of Generative Artificial Intelligence Services, 生成式人工智能服务管理暂行办法,* 13 July 2023, https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm (July 7, 2024)

reliability, without discrimination[177]. As stated in Article 1, the Measures apply to developers and providers of generative artificial intelligence systems for text, images, audio, video and other content intended for persons living in the People's Republic of China[178].

The legislation outlines principles for the provision and use of generative AI, including the following:

- Take appropriate measures to prevent discrimination on the basis of ethnicity, religious beliefs, state, region, gender, age, employment and health status in the process of algorithm processing, selection of training data, model generation and optimisation, and in service provision[179];
- Respect intellectual property rights, corporate ethical standards, keep business secrets and not use algorithms, data, platforms and other tools to create a monopoly and engage in unfair competition[180];
- Respect the rights and interests of individuals, avoiding harm to their physical and mental health and not violating their data protection rights [181];
- Depending on the type of service, implement effective measures to increase the transparency of generative artificial intelligence services and improve the accuracy and reliability of the content produced[182].

Furthermore, the Measures require that providers of generative artificial intelligence services conduct training data processing activities, such as initial training and optimisation training. These activities involve using data and base models from legitimate sources[183], respecting the intellectual property rights of others[184], obtaining consent from individuals for the use of their personal information, or complying with other administrative provisions as per the law[185]. They must also adopt effective measures to enhance the quality of training data, emphasising authenticity, accuracy,

[177] *Ibidem*, art 4 (1)
[178] *Ibidem*, art 2
[179] *Ibidem*, art 4 (2)
[180] *Ibidem*, art 4 (3)
[181] *Ibidem*, art 4 (4)
[182] *Ibidem*, art 4 (5)
[183] *Ibidem*, art 7 (1)
[184] *Ibidem*, art 7 (2)
[185] *Ibidem*, art 7 (3)

objectivity, and diversity of data, while complying with all applicable administrative regulations[186].

Unlike the EU AI Act, which addresses artificial intelligence broadly, Chinese law specifically regulates generative artificial intelligence, leaving other forms unregulated. These regulatory initiatives demonstrate a balance between China's desire to support AI innovation and the need to control and guide its impact on society and the economy. Thus, China emerges as a significant player in the global landscape of AI regulation, with an approach that could influence global trends in this field.

2.4 Other Global Players in AI Regulation

2.4.1 Brazilian Draft on AI Regulation

In Latin America, Brazil clearly stands out for its dynamic in addressing the regulation of artificial intelligence. Between 2019 and 2021, the country submitted three legislative proposals concerning AI to Congress. Although none of these were passed as law, Bill 2338/2023 "*Brazil's Proposed AI Regulation*"[187], introduced in May 2023, aims to become a comprehensive piece of legislation that will define the regulatory framework for Brazil's ethical and responsible use of artificial intelligence.[188] The bill is the result of a comprehensive effort to create a new bill that replaces three bills that have been pending in Congress for the past four years (5051/2019[189], 21/2020[190] e 872/2021[191]).

The legislative proposal seeks to address the potential risks and negative impacts of AI while promoting its benefits. The creation of a commission in March 2022 marked

[186] *Ibidem*, art 7 (5). Regulations to be complied with are the "Information Security Law of the People's Republic of China", the "Data Security Law of the People's Republic of China", the "Personal Information Protection Law of the People's Republic of China" and other relevant regulatory requirements of the relevant competent authorities.
[187] Senado Federal, Senador Rodrigo Pacheco, Projecto de Lei n° 2338, de 2023
[188] A. BAIG, *Brazil's New AI Law: What You Should Know,* in Securiti AI, 20 march 2024. https://securiti.ai/brazil-ai-regulation-and-law/ (July 13, 2024)
[189] Senado Federal, Senador Styvenson Valentim, Projeto de Lei n° 5051, de 2019
[190] Senado Federal, Eduardo Bismarck, Projeto de Lei nº 21 de 2020
[191] Senado Federal, Senador Veneziano Vital do Rêgo, Projeto de Lei nº 872 de 2021

the beginning of this effort, which lasted nearly 240 days and involved meetings, seminars, and public hearings.[192]

The Brazilian AI regulation proposal aims to safeguard fundamental rights and ensure the adoption of safe and reliable systems, fostering human well-being, supporting the democratic regime, and promoting scientific and technological development.

Articles 2 and 3 lay down foundations and guiding principles for the development and use of AI, including respect for human rights, democratic values, equality, non-discrimination, plurality, and respect for labour rights[193]. They also provide guiding principles, such as the importance of accountability[194] as well as measures to prevent, mitigate, and address systemic risks that may arise from intentional or unintentional use and effects of AI-based systems.[195]

Although the proposed Brazilian regulation on artificial intelligence has not yet been turned into law, there is currently no legal definition of artificial intelligence in Brazil. However, in the proposal, an artificial intelligence system is described as "*a computational system, with varying degrees of autonomy, designed to infer how to achieve a given set of objectives, using approaches based on machine learning and/or logic and knowledge representation, through input data from machines or humans, with the aim of producing predictions, recommendations, or decisions that may influence the virtual or real environment."*[196]

Moreover, the proposal classifies AI systems according to different risk levels: Article 13 requires providers to conduct a preliminary assessment to classify the risk level as "Excessive" or "High", and these risk assessments must be carried out before the AI system is introduced to the market or deployed in service. [197] Systems considered to be of "Excessive"[198] risk will be prohibited, including those that exploit vulnerabilities of specific groups or use subliminal techniques. The article also bans the

---

[192] *Access Alert | Brazil's New AI Bill: A Comprehensive Framework for Ethical and Responsible Use of AI Systems,* in Access Partenrship, 5 May 2023, https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/ (July 13, 2024)
[193] Projecto de Lei n° 2338, de 2023, Art 2. https://www25.senado.leg.br/web/atividade/materias/-/materia/157233
[194] *Ibidem,* Art 3, sec. IX
[195] *Ibidem,* sec. XI
[196] *Ibidem,* Art 4
[197] *Ibidem,* Art 13
[198] *Ibidem,* Art 14-15-16

use of such systems by public entities to evaluate, classify, or rank individuals based on their social behaviour or personality traits for access to goods and services and public policies in an illegitimate or disproportionate manner. Article 17 identifies high-risk sectors and applications, which encompass AI systems used for the following purposes: critical infrastructure security, education, hiring, human resources management, and health.

Providers and operators of AI systems must also establish governance structures and internal processes capable of ensuring the safety of the systems and the protection of the rights of affected individuals. These must include at least: transparency, adequate data management measures to mitigate and prevent potential discriminatory biases, the legitimisation of data processing in accordance with data protection regulations, the adoption of appropriate parameters for the separation and organisation of data for training, testing, and validating the system's outcomes, and the implementation of suitable cybersecurity measures from the design phase through to the operation of the system.[199]

These measures apply throughout the life cycle of AI systems – particularly those that pose a high risk – and require documentation, testing, and bias prevention measures. Article 24 stipulates that the impact assessment must take into account various factors related to the AI system, such as foreseeable and known risks, associated benefits, the likelihood and severity of negative outcomes, operational logic, conducted tests and evaluations, mitigation measures, training and awareness, transparency measures for the public, and others. Additionally, the assessment must be accompanied by regular quality control tests and a justification of the system's residual risk.[200]

Despite the fact that the final version of the bill has yet to be approved and may undergo further modifications during the legislative process, the Brazilian proposal stands as a significant step towards the regulation of artificial intelligence in the country. This proposal represents a comprehensive effort to ensure that the adoption and implementation of AI systems are guided by ethical and responsible principles. The emphasis on risk assessment, transparency, and ethical governance underscores the commitment to fostering a safe and equitable AI ecosystem in Brazil.

[199] *Ibidem,* Art 19
[200] *Ibidem,* Art 24

2.4.2 Canadian Artificial Intelligence and Data Act

On June 16, 2022, the Government of Canada introduced the "*Artificial Intelligence and Data Act*" (AIDA)[201] as part of Bill C-27[202], known as the "*Digital Charter Implementation Act*." This bill also includes the "*Consumer Privacy Protection Act*"[203], which aims to modernise privacy laws in the Canadian private sector, and the "*Personal Information and Data Protection Tribunal Act*"[204], which would establish an appeals tribunal for decisions made by the Office of the Privacy Commissioner.

AIDA is a risk-based legislation designed to protect fundamental rights, human health and safety, and democracy. Its aim is to balance the risks associated with artificial intelligence with the opportunities to promote responsible and reliable innovation in the field of AI.

The stated purposes of the AIDA are: "(i) to regulate the cross-border trade of artificial intelligence systems by establishing common requirements applicable across Canada for the design, development, and use of such systems; and (ii) to prohibit certain conduct related to artificial intelligence systems that may cause serious harm to individuals or damage to their interests (particularly, biased outputs)" [205].

The risk-based approach, including its definitions and key concepts, has been developed to align with and adapt to emerging international standards in the AI sector. This includes the EU AI Act, the principles of the Organisation for Economic Co-operation and Development (OECD)[206], and the Risk Management Framework (RMF)

---

[201] HOUSE OF COMMONS OF CANADA, BILL C-27, *The Artificial Intelligence and Data Act (AIDA),* First reading, june 16, 2022, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

[202] *Bill C-27*, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts*, November 22, 2021, https://www.parl.ca/LegisInfo/en/bill/44-1/C-27

[203] Government of Canada, *Consumer Privacy Protection Act,* https://ised-isde.canada.ca/site/innovation-better-canada/en/consumer-privacy-protection-act

[204] HOUSE OF COMMONS OF CANADA, BILL C-27, *Personal Information and Data Protection Tribunal Act*, First reading, june 16, 2022, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

[205] HOUSE OF COMMONS OF CANADA, BILL C-27, *The Artificial Intelligence and Data Act (AIDA),* 16 June 2022, sec. 4, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

[206] OECD*, AI Principles Overview*, adopted May 22 2019, amended May 3 2024, OECD member countries approved a revised version of the organisation's definition of an AI system available at https://oecd.ai/en/ai-principles

from the National Institute of Standards and Technology (NIST) in the United States, integrating seamlessly with existing legal frameworks in Canada[207].

In October 2023, François-Philippe Champagne, the Minister of Innovation, Science, and Industry, introduced several amendments to the AIDA. These amendments include the definition of "high-impact artificial intelligence systems" and the assessment of potential harms caused by such systems in specific contexts. [208]

Regarding risk categorisation, an artificial intelligence system has been specified to be classified as a "high-impact system" if its use falls into one of the seven categories identified in the proposed amendments. Specifically, an artificial intelligence system will be considered "high-impact" [209] when employed in the following situations:

- To determine employment-related issues;
- To decide whether to provide services to an individual, or the type or cost of services to be provided to an individual, or to establish the priority of services to be provided to individuals;
- To process biometric information in matters related to: (i) the identification of an individual, except where such information is processed with the individual's consent to authenticate their identity; or (ii) the assessment of an individual's behaviour or mental state;
- In matters related to content moderation and prioritisation, specifically, AI systems used for: (i) moderating content on an online communication platform, including a search engine or social media service; or (ii) prioritising the presentation of such content;
- In healthcare or emergency services;
- By a court or administrative body in decisions concerning an individual who is part of a proceeding;
- To assist a police officer, as defined in the Criminal Code, in the exercise and execution of their powers, duties, and law enforcement functions.[210]

---

[207] *The Artificial Intelligence and Data Act (AIDA) – Companion document,* https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document#fn20

[208] A. LACASSE, *Canadian Parliament's Bill C-27 hearing delves deeper into AIDA*, IAPP https://iapp.org/news/a/canadian-parliaments-bill-c-27-hearing-delves-deeper-into-aida

[209] *Ibidem,* sec 5. A concept that was originally introduced in AIDA. The definition has been modified by the proposed amendments, p. 38

[210] *Ibidem.*

Crucially, AIDA emphasises the importance of addressing potential AI risks prior to the deployment or use of systems. Although the AI risk management principles and their associated requirements can be implemented throughout various stages of the AI lifecycle, they are particularly vital for those deploying AI systems[211]: Human oversight and monitoring, Transparency, Fairness and equity, Safety, Accountability, Validity and robustness.[212]

Additionally, the AIDA requires that those responsible for a high-impact AI system take measures to identify, assess, and mitigate the risks of harm or distorted outputs that may result from the use of the system. [213] "Harm" is defined in the AIDA as: *"(a) physical or psychological harm to an individual; (b) harm to an individual's property; (c) economic loss to an individual*" [214]. "Biased output", on the other hand, is defined as "*content generated, decision, recommendation, or prediction made by an artificial intelligence system that discriminates negatively, either directly or indirectly and without justification, against an individual based on one or more prohibited grounds of discrimination established in the Canadian Human Rights Act (1985), or on a combination of such prohibited grounds*."[215]

The AIDA, therefore, aims to ensure that Canadian citizens are protected from potential harms of AI, and that AI risks are properly managed at every stage of the AI lifecycle, from design to operation. The goal is to establish concrete standards for the AI sector that reflect fundamental risk management principles and support reliable and responsible innovation in the field of AI. Starting in 2024, Bill C-27, which includes the AIDA, is under review by the House of Commons committee, and its approval will depend on the progress of the legislative process.

---

[211] *What You Should Know: Canada's Artificial Intelligence and Data Act*, in Lumenova, April 2024, https://www.lumenova.ai/blog/canada-ai-and-data-act-what-you-should-know/ (July 14, 2024)

[212] *The Artificial Intelligence and Data Act (AIDA) – Companion document,* https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document#fn20

[213] H. CHAMBERS, *Canada: AI and Data Act - Key takeaways*, in DataGuidance, 2023, https://www.dataguidance.com/opinion/canada-ai-and-data-act-key-takeaways (July 14, 2024)

[214] *The Artificial Intelligence and Data Act (AIDA),* 16 June 2022, sec. 5, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

[215] *Ibidem*

2.4.3 Toward a Global Governance of AI

In addition to the national regulatory measures that each country is implementing to regulate artificial intelligence, it will be equally crucial to establish sound and coordinated international governance. This is because AI, by its very nature, has a global reach that transcends territorial boundaries and local jurisdictions, with ramifications that extend throughout the world. Although, at present, the development of advanced AI systems is highly concentrated in a limited number of countries, it is clear that access to many AI models cannot be easily contained or restricted within national borders. Consequently, international cooperation becomes indispensable to address the challenges and opportunities associated with this emerging and highly influential technology.

Thus, it becomes imperative to develop a set of shared international rules to guide the development and responsible use of artificial intelligence. Such rules will not only need to set global standards, but also ensure that governments and private actors, operating within their respective sovereign jurisdictions, are held accountable for how they use these technologies. This emerging global regulatory framework will also need to be fully consistent with already established international law, including existing norms and conventions, such as the UN Charter, International Humanitarian law, the Universal Declaration of Human Rights, and other important multilateral treaties, which provide an essential legal and moral basis for the future of AI worldwide[216].

In October 2023, Chinese President Xi Jinping announced a global AI governance initiative[217], emphasising the need to establish international rules and standards to guide the safe and responsible use of this emerging technology.

Subsequently, the G7 published international guiding principles as part of the "Hiroshima Process", an initiative aimed at regulating organisations developing advanced artificial intelligence systems[218]. These principles aim to promote the responsible development of AI, addressing issues such as transparency, security and

[216] E. KLEIN, S. PATRICK, *Envisioning a Global Regime Complex to Govern Artificial Intelligence*, 2024, p. 5

[217] Chinese Ministry of Foreign Affairs, *Global AI Governance Initiative,* October 20, 2023, https://www.fmprc.gov.cn/eng/xw/zyxw/202405/t20240530_11332389.html (August 28, 2024)

[218] Japanese Ministry of Foreign Affairs, *G7 Leaders' Statement on the Hiroshima AI Process*, October 30, 2023, https://www.mofa.go.jp/ecm/ec/page5e_000076.html (August 28, 2024)

accountability. In parallel, the G7 introduced a related code of conduct[219], which provides concrete guidelines for companies and institutions engaged in the research and implementation of AI technologies, ensuring that internationally agreed ethical and regulatory standards are met.

Straight after, the UK convened the first global summit on AI security[220], which concluded with the publication of the Bletchley Declaration[221]. This document calls on countries to actively collaborate through existing international forums and other initiatives, with the aim of jointly addressing challenges related to the development and safe use of AI.

The United Nations Educational, Scientific and Cultural Organisation (UNESCO) Recommendation on the Ethics of AI[222], adopted in 2021 by 193 states, the principles of the Organisation for Economic Cooperation and Development (OECD) [223], and the multidisciplinary research reports of the Global Partnership on AI[224] are fundamental to this set of initiatives, as they are built upon them.

As noted above, multiple international attempts have been made to establish global governance for AI, but the challenge of reaching a common and effective agreement remains significant. In order to achieve an international agreement, it is crucial to consider that the major powers and the major tech companies are competing for the political, economic, and social benefits of this technology.

---

[219] Japanese Ministry of Foreign Affairs, *G7 Leaders' Statement on the Hiroshima AI Process*, October 30, 2023, https://www.mofa.go.jp/ecm/ec/page5e_000076.html (August 28, 2024)

[220] Foreign, Commonwealth, and Development Office, Department for Science, Innovation, and Technology; and the AI Safety Institute, November 2023, https://www.gov.uk/government/topical-events/ai-safety-summit-2023 (August 28, 2024)

[221] *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*, UK Government, November 1, 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

[222] UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2022, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

[223] OECD*, AI Principles Overview*, adopted May 22 2019, amended May 3 2024, OECD member countries approved a revised version of the organisation's definition of an AI system available at https://oecd.ai/en/ai-principles

[224] *Global Partnership on Artificial Intelligence GPAI,* https://gpai.ai/projects/

Conclusion

The use of AI must be carefully regulated to avoid creating new issues, vulnerabilities, and problems related to ethics, transparency, jurisdictional control, misinformation, social cohesion, and the integrity of democratic institutions. It is evident that the potential risks of AI have been recognised both in Europe and the United States, as well as in Asia, with China playing a significant role. However, countries are at different stages in developing their approaches to AI regulation and have varying opinions on how best to achieve it.

On the one hand, Europe adopts a human-centred perspective, based on classifying AI systems according to risk and establishing a range of horizontal obligations, with a focus on protecting individual rights and creating a harmonised legal framework. The introduction of the AI Act and initiatives from the Council of Europe represent crucial steps towards regulation that balances innovation with fundamental rights.

On the other hand, the United States adopts a more business-friendly approach, preferring a more fragmented and wide-ranging regulatory environment, with the aim of encouraging development and maintaining global leadership in AI. The U.S. tends to avoid stringent regulations that could limit technological development, while still introducing guidelines to protect citizens from the risks associated with AI.

China, with its ambitious and centralized strategy, aims to become the global leader in AI, integrating rigorous ethical and regulatory principles to ensure controlled and safe development of AI technologies.

Beyond these three powers, other regions in the world are moving towards AI regulation. As seen, Brazil and Canada are proposing laws that assess the risks associated with specific uses of AI and protect citizens' rights. However, many countries and regions still lack adequate regulation on AI and innovative technologies.

These distinct approaches reflect the varying priorities and values of each context, providing an overall view of the challenges and opportunities in creating a global regulatory framework for artificial intelligence. Despite the significant developments achieved so far, AI regulation requires a delicate balance between promoting technological innovation and protecting human rights. However, there

remains uncertainty as to whether AI systems can fully respect such rights, ensuring transparency, accountability, and the absence of discrimination.

As we have seen, there have been international efforts to regulate artificial intelligence and find common ground. However, despite these significant developments, any attempt to govern AI on a global scale will face powerful incentives that could hinder such regulations. In fact, major powers and leading companies are engaged in a competitive race to obtain the geopolitical and economic benefits associated with this emerging technology.

It is crucial to understand that artificial intelligence is not neutral. In fact, AI learns from the data on which it has been trained, and these data often reflect existing inequalities and injustices in society. Therefore, if we use current AI systems to make important decisions in the future, there is a risk that these inequalities may be perpetuated or even amplified, rather than addressed and corrected.

The next chapter will examine how these new technologies can give rise to algorithmic discrimination, perpetuating existing biases and having a significant impact on social justice and equity. It will explore how seemingly impartial algorithms can, in fact, amplify existing social inequalities, contributing to the creation of new forms of modern domination and control.

CHAPTER III

Is AI Truly Neutral? How AI Can Perpetuate Discrimination and Contribute to New Forms of Colonialism

Introduction

Artificial intelligence is often represented as neutral technology capable of making objective decisions based on data. However, this perspective has ignored the fact that AI can perpetuate systems of discrimination and oppression not only by continuing, but also by intensifying them.

The purpose of this chapter is to demonstrate that the dynamics of oppression that characterised historical colonialism are reproposed today in apparently new forms, but with the same underlying logic, and that the myth of AI neutrality can lead to a false sense of security and a misperception that can have serious consequences. The arrival of modern technologies, especially artificial intelligence and data analysis, brought about an evolution in the face of colonialism, a face now apparent not through the possession of territory, but through the governance of information and digital assets.

The extensive utilisation of data by artificial intelligence poses the risk of reinforcing stereotypes and biases, given that algorithmic models are developed based on data sets that embody historical prejudices, such as racism, sexism and economic exclusion. Consequently, this leads to outcomes and decisions that, instead of being unbiased, turn out to be systematically discriminatory toward individuals and groups that are already marginalised.

The age of automation, with its computer-based logic, is profoundly shaping human nature and culture. This logic, applied through algorithmic control, accentuates the inequalities between the North and the South. This trend creates the basis for a “new AI empire”, in which decision-making power is concentrated in the hands of a few global players, while the most vulnerable are excluded or exploited.

This chapter will explore the dangers associated with AI being regarded as an impartial intelligence. Furthermore, we will reflect on the danger that AI may evolve, or has already evolved, into a tool that contains within it a new kind of technological

colonialism. The nations and multinationals that dominate the development and distribution of AI could impose their values, standards, and power systems on technologically less advanced communities and countries, perpetuating global inequalities and limiting the sovereignty of entire populations.

Finally, it will be examined how this phenomenon of discrimination and technological colonialism manifests itself in a particularly pervasive way on communities that have historically already suffered the worst forms of oppression: the indigenous peoples. These communities, already victims of centuries of colonisation and exploitation, are now facing a new threat that, although disguised as technological progress, reproduces the same logic of exclusion and domination as in the past.

Through the analysis of the *Ewert v. Canada* court case, it will be shown how various studies and the Canadian Supreme Court's ruling highlight how automated artificial intelligence systems, far from resolving inequalities, can even reinforce them. These systems, built on data steeped in historical and cultural biases, end up perpetuating structural discrimination. Instead of acting as instruments of equity and inclusion, modern technologies continue to treat indigenous peoples as marginalised subjects, reinforcing their exclusion and depriving them of their right to self-determination.

3.1 Artificial Intelligence between the Myth of Neutrality and the Risk of Discrimination

"*Everyone is entitled to all the rights and freedoms set forth in this Declaration, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status*."[225]

In the Universal Declaration of Human Rights, the principle of non-discrimination is recognised as one of the fundamental principles for the full exercise of human rights and, according to the principle of equality, should be guaranteed as "*All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of*

[225] Art 2, *Universal Declaration of Human Rights* UDHR, 10 December 1948.

brotherhood"[226]. The principle of equality, in fact, is the prerequisite for the effective enjoyment of any individual right, representing, the first fundamental curb to possible arbitrary behaviour of authority.

The prohibition of discrimination is part of the essential core of general international law, forming part of *Ius Cogens*, that is, that group of mandatory rules that bind all states without exception. This principle is enshrined in numerous international legal instruments, starting with Article 1 of the United Nations Charter[227], passing through Article 2 common to the two 1966 International Covenants on Civil and Political Rights and on Economic, Social and Cultural Rights respectively[228], up to the Convention on the Rights of the Child (Article 2)[229].

The principles of non-discrimination and equality, firmly rooted in a wide range of international treaties and conventions, have played a fundamental role in promoting inclusion and countering the structural and systemic discrimination that has characterised many societies throughout history. In recent years, discrimination, which is a phenomenon closely linked to human behaviour, has undergone new and partly unexpected transformations. The entry of technology into human, public and private space and, with it, artificial intelligence techniques has brought with it important consequences on, among others, the phenomenology of discrimination.

Artificial intelligence, with its ability to process data on a large scale and make autonomous decisions, offers enormous potential in various fields, from economics to justice and healthcare. However, its use is not without risks. Indeed, AI can reproduce, or even amplify, existing inequalities if not carefully designed and used. Distorted

---

[226] Art 1, *Universal Declaration of Human Rights,* 10 December 1948.

[227] Art 1, *United nations Charter,* 1945, "*The Purposes of the United Nations are: (...) promoting and encouraging respect for human rights and for fundamental freedoms for all without distinction as to race, sex, language, or religion"*

[228] Art 2 (1), *International Covenant on Civil and Political Rights* (ICCPR) and the *International Covenant on Economic, Social and Cultural Rights* (ICESCR), 1966, "*Each State Party to the present Covenant undertakes to respect and to ensure to all individuals within its territory and subject to its jurisdiction the rights recognized in the present Covenant, without distinction of any kind, such as race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status."*

[229] Art 2, *Convention on the Rights of the Child*, 1989, *"States Parties shall respect and ensure the rights set forth in the present Convention to each child within their jurisdiction without discrimination of any kind, irrespective of the child's or his or her parent's or legal guardian's race, colour, sex, language, religion, political or other opinion, national, ethnic or social origin, property, disability, birth or other status."*

databases, non-transparent algorithms and automated decisions can lead to invisible but extremely harmful forms of discrimination.

The risk of discrimination is inherent in any decision-making or assessment process involving the individual taken by those exercising power over them, whether formal or informal.

In the present era of rapid technological progress, automated systems play an increasingly important role in decision-making processes, with artificial intelligence often at the centre of these dynamics. On the one hand, the application of algorithms offers considerable advantages in terms of simplification, making decision-making processes faster and more agile by automating weightings and evaluations that would otherwise require human intervention. However, this apparent simplification also adds more complexities. Indeed, smart technologies add a further layer of difficulty to the decision-making process, as algorithmic evaluation requires a particularly sophisticated handling by the human operator.

The amount of data that an algorithm is capable of analysing far exceeds the processing capabilities of a human being and this vastness of information makes it extremely complex for the human operator to perform a critical processing of the information provided by the machine. The use of advanced technologies, including artificial intelligence systems, does not automatically reduce or increase the risk of discriminatory outcomes in decisions. What really influences the actual effects from this point of view are the specific characteristics of the algorithm and the way in which it is used.[230]

3.1.1 The Illusion of the Objectivity of the Machine

Algorithms are often thought of as neutral tools based on objective calculations. This perception stems from the belief that the use of digital technologies reduces human

[230] L. RINALDI, *L'intelligenza artificiale come nuova frontiera dei diritti fondamentali*, Tesi di Dottorato, Università degli studi di Trento, 2023, pp 140-142

error, leading to more accurate, quick, and more reliable results.[231] This belief is justified by the role that technological innovation has played in history, especially since the Industrial Revolution. With the introduction of automated machinery, production was accelerated and standardised, drastically reducing defects by eliminating the margin for human error[232]. Similarly, the most advanced technologies, including artificial intelligence systems, are often associated with an aura of objectivity, efficiency and neutrality.

However, when it comes to artificial intelligence, the situation becomes much more complex. AI systems, employed to support or substitute decision-making and evaluation processes, are based on the analysis of data. However, the data itself is essentially human and "earthy", as it is produced by humans. This makes it potentially incomplete, biased and not always capable of accurately representing external reality and carries with it the risk of reflecting pre-existing social and cultural prejudices and expectations. In this respect, the data sets used should be broad and representative, to avoid some biases that lead to inaccurate or discriminatory results.

"Data are assumed to accurately reflect the social world, but there are significant gaps, with little or no signal coming from particular communities. While massive datasets may feel very abstract, they are intricately linked to physical place and human culture"[233]. With this study, Kate Crawford, one of the leading scholars of the interactions between discrimination and artificial intelligence, emphasised that data seem neutral and abstract, but in truth they are not because they only show a part of reality, the part that humans can see or that is shown to them.

After all, it cannot be expected that data should capture the entire complexity of society, as all data are, by their very nature, statistical and probabilistic. They should be interpreted in this way, and not taken as absolute truths. The problem arises when data are treated as if they were unquestionable truths, diverting attention from their probabilistic nature. This is especially concerning due to the fact that, even though the data are man-made and therefore subjective, they are then used as if they were objective

---

[231] E. M. CAMPBELL et al, *Overdependence on technology: An unintended adverse consequence of computerized provider order entry*, in AMIA Annual Symposium Proceedings, 2007, pp. 94-98, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2710605/ (August 27, 2024)

[232] J. DE VRIES, *The industrial revolution and the industrious revolution*, in The Journal of Economic History, Vol. 54, No. 2, 1994, p. 249-270

[233] K. CRAWFORD, *The Hidden Biases in Big Data*, in Harvard Business Review, 2013. https://hbr.org/2013/04/the-hidden-biases-in-big-data (August 27, 2024)

to make decisions and unbiased distinctions. This divergence from reality can make it difficult to understand how artificial intelligence techniques work, the potential biases inherent within them, and the accountability of those who make determinations predicated on the data[234].

If we consider that data do not exist on their own, but are created and influenced by humans, in terms of the causes of discrimination, we can see a similarity between traditional discrimination, caused directly by human action, and discrimination resulting from the use of artificial intelligence. The main difference is that, in the second case, discrimination is mediated by data and machines, not coming directly from humans. However, the common element in both cases is the human factor.

Moreover, behind all human behaviour, whether it directly causes discrimination or interacts with data, there is a cultural element that affects both human conduct and the data itself. This link with culture, which is not external but internal to the data, makes the causes of AI-related discrimination similar to those of traditional discrimination.

The perception of objectivity and infallibility that has historically accompanied technological innovations should, therefore, be revised in the context of artificial intelligence. Although these tools may offer high reliability, it is crucial to recognise that there is still a margin of error and that automated decisions are not immune to inaccuracy.

3.1.2 AI-Derived Discrimination

Anti-discrimination law[235] uses the term "discrimination" to refer to laws, decisions, criteria and practices that place certain individuals at a disadvantage because of certain "protected characteristics". These characteristics are listed in a non-fixed and evolving catalogue that, in the contemporary West, includes "sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other

[234] C. NARDOCCI, *Intelligenza Artificiale e Discriminazioni,* 2021, pp. 15-16

[235] The notion of anti-discrimination law adopted here refers to that area of law that focuses on defining the characteristics of discrimination, identifying its forms and types, and the means to prevent and counteract it in its external manifestations.

opinion, membership of a national minority, property, birth, disability, age or sexual orientation"[236]. If we take the general legal concept of discrimination as a reference, algorithmic discrimination can be defined as that prescription, decision or practice that entails disadvantages for certain persons on account of their characteristics protected by law, where such decisions or practices are adopted or implemented through the use of algorithms, including those based on artificial intelligence.[237]

There can be a number of reasons that a computer system and its algorithms may lead to discriminatory decision-making processes. First, an algorithm is a defined sequence of steps or instructions designed to perform calculations or solve problems, particularly by a computer. In computing, algorithms are translated into software, allowing the system to process input data according to specific rules and generate the corresponding output[238]. The lack of neutrality of an algorithm is often described by the term "bias", which is now mainly associated with a negative meaning. However, originally, the term denoted any deviation from a standard of operation[239]. When speaking of algorithmic discrimination, bias refers to any reason that leads the algorithm to produce a result that differs from that expected by its proper functioning.

Describing and analysing the steps in the programming of artificial intelligence is an important process, because through this process, one can identify the moments when a variable that may lead to the biased functioning of the machine, what one may call bias time, is introduced.

Five mechanisms have been identified in the programming of artificial intelligence and its subsequent operation, from which disproportionately adverse outcomes and discriminatory impacts against already marginalised communities can occur. [240]

With regard to the first mechanism, the risk of discrimination is considered to lie in the identification and relationships that are established between the "target variable", the characteristic that the system searches for, and the "class label", the category that is

[236] Art 21, Charter of Fundamental Rights of the European Union (2000/C 364/01), December 1, 2009
[237] G. GOMEZ, *Intelligenza artificiale, profilazione e nuove forme di discriminazione*, in Teoria e Storia del Diritto Privato, 2022, pp. 11-19.
[238] S. SILVA, M. KENNEY, *Algorithms, platforms, and ethnic bias: an integrative essay*, in Phylon (1960-) Vol. 55, No. 1 & 2, SUMMER/WINTER 2018, p. 11
[239] In in Online Etymology Dictionary, Bias (n.) "oblique or diagonal line," from French biais "a slant, a slope, an oblique", https://www.etymonline.com/word/bias (August 27, 2024)
[240] S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact,* in California Law Review, 2016, pp. 67 ss.

associated with it. The problem of associations between the "target variable" and the "class label" realised by the algorithm occurs especially when characteristics do not fit well into a binary scheme with only two options. This leads to the high risk that automated systems do not adequately consider individual circumstances, resulting in unfair treatment of vulnerable groups. The concern is, therefore, that decisions on how to identify characteristics and categories may have greater negative effects on protected classes. Ultimately, the risk of discrimination may be related to the choice of characteristics or categories, or both[241].

The second mechanism concerns the collection and selection of data, known as "data training". Artificial intelligence techniques are based on examples or models created using data from which machines learn. Sometimes, machines can operate autonomously and make inferences that deviate from the initial model. For this reason, it is crucial that data are collected and selected with care. If the data is biased or influenced by biases, the resulting model is likely to be discriminatory, following the principle known as "garbage in, garbage out", whereby incongruous, inaccurate or out-of-date data can only produce unreliable decision-making results, creating real traps, often mostly invisible. Ultimately, data quality is crucial: less accurate data increases the risk that the machine will make unreasonable and, therefore, discriminatory distinctions.[242]

"Feature selection" represents the third mechanism, i.e. the selection of individual characteristics that the model uses. If this selection is biased or wrong, excluding important aspects or exaggerating some to the detriment of others, the model may end up making unfair distinctions. Discrepancies in the selection of characteristics thus result in further discrimination against individuals and groups.

The fourth mechanism concerns the use of *proxies*, i.e. the elements that the algorithm uses to make differentiations. Proxy discrimination occurs when human decision-makers indirectly discriminate against a legally protected class by using seemingly neutral data, rather than explicitly exploiting prohibited traits. For example, an employer might exclude applicants from predominantly black neighborhoods by using zip codes as a proxy, rather than directly refusing to hire black applicants. This practice constitutes disparate treatment. Although the employer's method appears

[241] *Ibidem,* p. 680
[242] C. NARDOCCI, 2021, op. cit., pp. 20-21

neutral, it effectively achieves the same discriminatory outcome. The decision-maker understands the correlation between the neutral data and the protected class but avoids directly using sensitive feature data. Since the discrimination is not based on explicitly prohibited traits, it can often be legally defended as being driven by neutral considerations, despite undermining the principle of equality[243].

The fifth and final mode of discrimination in artificial intelligence techniques is the most obvious and easiest to identify. It occurs when the programmer acts intentionally, with a partial selection of data, to harm one group over others. This phenomenon, called "masking", occurs when discrimination results directly from the intentional actions of the programmer in the model development phase. In these cases, the machine learning model is deliberately designed to include biases.

Ultimately, there are two further aspects that can adversely affect the functioning of artificial intelligence techniques and cause discrimination: the updating of data - to adapt the machine to progress - and their pollution - the deliberate feeding of incorrect data[244].

AI systems can be discriminatory not because the system is "bad" per se, but because it inherits bad behaviour that it then repeats. Like previous technologies, artificial intelligence will inevitably embody the values of those who created it. We risk developing AI systems that replicate a limited and privileged perspective of society, reinforcing existing biases and stereotypes. Those who have already faced marginalisation or prejudice for years, now with AI risk facing a further threat since "currently the loudest voices debating the potential dangers of superintelligence are affluent white men"[245].

[243] X. CHEN, *Algorithmic proxy discrimination and its regulations,* in Computer Law & Security Review Vol. 54, September 2024
[244] C. NARDOCCI, 2021, op. cit. p. 22
[245] K. CRAWFORD, *Artificial Intelligence's. White Guy Problem,* in The New York Times, 2016. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html (August 27, 2024)

3.2 Artificial Intelligence: The New Face of Colonialism

3.2.1 From Traditional Colonialism to Data Colonialism

During colonialism, Europeans dispossessed the indigenous peoples of their ancestral lands, occupied their territories, forced them to work as slaves and serfs for the landowners, committed atrocious violence and established a condition of dependency and exploitation through strategic policies of underdevelopment[246].

Similarly to the technical skeleton of classical colonialism, the creation of the technological ecosystem for benefit and plunder is rooted in a new form of colonialism that is distinctive of the 21st century: *data colonialism*[247].

The concept of digital colonialism was first introduced by Herbert Schiller in 1976 in his essay "*Communication and Cultural Domination*"[248]. Schiller analysed how the emergence of a new technological era favoured the dominant countries, widening the gap between the more developed global regions and the poorer ones. He argued that through "important communication equipment and foreign-produced software", developing nations and poorer economies were being subjected to the control and will of the dominant global powers[249].

As Kwet argued, this structural form of dominance is exercised through centralised ownership and control of the "three fundamental pillars of the digital ecosystem: software, hardware and network connectivity"[250].

If railways and shipping lanes once represented the "open veins" of the colonies, today digital infrastructures perform a similar function: the "digital veins" that cross the oceans, cabling a technological ecosystem owned and controlled by a few multinational technology corporations that monitor, collect data and offer services created specifically for digital fiefdoms, using proprietary applications, corporate clouds and centralised Internet services[251].

---

[246] M. KWET, *Digital colonialism: US empire and the new imperialism in the global south.* In Sage Journals, Race & Class, Vol. 60(4), 2019, pp. 3–26.
[247] G. KAKAR, *Cognitive dysphoria: Evaluating the paradigm shift of artificial intelligence technology in digital colonialism.* In Indian Journal of Artificial Intelligence and Law Vol 2, 2021, p. 7- 10.
[248] H. SHILLER, *Communication and Cultural Domination,* 1976
[249] G. KAKAR, 2021, op. cit., p. 9.
[250] M. KWET, 2019, op. cit.
[251] M. KWET, *Digital colonialism: the evolution of American empire,* In ROAR Magazine, 2021, https://roarmag.org/essays/digital-colonialism-the-evolution-of-american-empire/ (September 3, 2024).

As stated by Nick Couldry and Ulises Ali Mejias, Data Colonialism is a "term for the extension of a global process of extraction that started under colonialism and continued through industrial capitalism, culminating in today's new form: instead of natural resources and labor, what is now being appropriated is human life through its conversion into data"[252].

While historical colonialism entailed the annexation of territories, resources and populations, data colonialism adopts a more subtle and pervasive mode of domination: it involves the capture and control of human life itself through the appropriation and exploitation of personal data for profit. If this is true, similarly to how historical colonialism fuelled the rise of industrial capitalism, data colonialism is setting the stage for a capitalism based on the market and the exploitation of data. In this new scenario, human life is annexed to capital through various means, including digital platforms, which transform personal data into valuable economic goods[253].

Therefore, we are at the centre of a new phase of colonialism, structurally linked to the development of capitalism. This process negatively affects the quality of life in two main ways: on the one hand, through constant monitoring of personal data and surveillance by those who collect these data; on the other hand, because it converts human life itself into a direct resource for capitalist production.

Through the so-called "data relations" – new types of human interaction that facilitate the extraction and commodification of personal data – global social life becomes an "open" resource for extraction, seemingly at the disposal of capital. In order for personal data to be freely available for appropriation, it is necessary to regard it as a natural resource, i.e. a resource that simply appears to exist. This cultural process of considering data a natural resource that can be extracted as if it were inert nature is facilitated by the normalisation and trivialisation of big data in everyday life[254].

There are several ways in which data colonialism integrates human life into capital. For instance, digital platforms act as technological instruments that create a new social dimension for capital. In this newly created space, human interactions can be continuously monitored, recorded and quantified as "data". Thanks to these platforms

[252] N. COULDRY, U. A. MEJIAS, *The Costs of Connection: How data is Colonizing Human Life and Appropriating It for Capitalism,* 2019, Preface, p. xix
[253] *Ibidem,* p. xi
[254] N. COULDRY, U. A. MEJIAS, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, in Sage Journals, Television & New Media, Vol. 20(4), 2019, pp. 336–349.

every aspect of life, that is also spheres previously outside of formal economic dynamics can be treated under the umbrella of commodification.

Another way in which human life is integrated into capital is the rapid spread of data-driven logistics across all aspects of production. Leaving its primary focus on managing the flow of products through global supply chains, logistics has expanded its "logic" to all forms of production, including human and non-human components, with the aim of optimising their management through data. This approach involves massive data collection and processing, transforming many work activities that were previously managed differently.

Finally, a third way in which human life is absorbed into capital is through new social relationships from social, in which individuals themselves track their activities to generate data. This can happen voluntarily, but often also occurs because individuals need to fulfil work requirements or to fulfil contractual obligations, such as those related to insurance or social security[255]. Recent studies have shown how these self-collection practices of data can lead to new forms of discrimination and inequality[256].

Analogously to the historical use of legal fiction, which labelled lands inhabited for millennia as "terra nullius" or "no man's land", and thus open to exploitation without any legal impediments and interference, there would appear to be a common-sense view about the naturalness of data appropriation, which, similar to historical colonialism, is based on extensive ideological work[257].

For example, data are frequently described as "the new oil", describing it as a representation of a lost resource for mankind that is once again extractable and usable for any purpose and for the benefit of companies. The World Economic Forum (WEF) stated: "personal data will be the new 'oil'—a valuable resource of the 21st century[258]…becoming a new type of raw material that's on par with capital and labour"[259]. This claim is based on the construction of data as a "raw material" with

[255] N. COULDRY, U. A. MEJIAS, 2019, op. cit.
[256] cf. V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor,* 2018; C. O'NEIL, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* 2016
[257] *Ibidem.*
[258] World Economic Forum (WEF), *Personal Data: The Emergence of a New Asset Class,* 2011, p. 5
[259] *Ibidem,* p. 7

intrinsic value. Thus, data become a fundamental form of resource appropriation and extraction[260].

Data colonialism combines the extractive and predatory practices of traditional colonialism with modern digital quantification techniques.

To appropriate data, data colonialism relies on different forms of extractive rationalities. One social rationality regards much of the work that fuels data mining as worthless, reducing it to "mere sharing". Then, there is also a practical rationality, which gives companies the exclusive power to process data and thus to appropriate it. Finally, a political rationality that portrays society as the natural beneficiary of these extractive processes, in the same way that, during historical colonialism, it was believed that humanity would benefit from the so-called "civilisation" project"[261].

A further parallel to historical colonialism can be drawn with the *Requerimiento* of the Spanish Empire, the absurdity of which was first described by Bartolomé de Las Casas. This proclamation, read by the *conquistadores* in Spanish to populations that did not speak the language, was intended to introduce the natives to the new world order under which they would be colonised and to obtain their acceptance, on pain of extermination, which often happened anyway, regardless of obedience. Today, in the age of data colonialism, the same mechanism is reproduced through the administration of documents likewise: the Terms of Service. These texts, full of obscure clauses, represent a modern form of appropriation claim by big corporations.

Digital colonialism represents a new form of relationship between poorer regions and post-industrial nations induced and developed by importing communication equipment and software from abroad, along with engineers, technicians and information protocols[262].

Multinational technology corporations exert significant control over local development, dominate markets and extract profits from the global South, thanks to their dominance over the digital architecture. This structural control results in new

---

[260] S. MEZZADRA, B. NEILSON, *On the multiple frontiers of extraction: excavating contemporary capitalism*, In Cultural Studies, Vol 31 (2-3), 2017, pp. 185-204 https://www.tandfonline.com/doi/full/10.1080/09502386.2017.1303425?scroll=top&needAccess=true (September 3, 2024)

[261] N. COULDRY, U. A. MEJIAS, 2019, op. cit.

[262] T. L. MCPHAIL, *eColonialism Theory: How Trends are Changing the World,* in The World Financial Review, 2014

forms of imperialism, with giants such as Google, Amazon, Facebook, Apple and Microsoft (GAFAM) assuming the role of new imperialists on the international stage.

In the perspective of digital colonialism, foreign powers and companies are establishing infrastructures in the global South to serve their interests, consolidate economic and cultural domination and impose privatised models of governance. These companies develop digital technologies that ensure their power over the essential functions of the technology ecosystem, thus strengthening their ability to exercise widespread control on a global scale[263].

In Johannesburg, the company Vumacam is building the nationwide CCTV network in collaboration with foreign multinationals including Chinese company Hikvision. Vumacam has erected poles with cameras whose footage is sent to security rooms across the country which in turn use AI tools to track the movements of the population. The result is the rapid creation of a mass, centralised, coordinated and privatised surveillance system. Through the narrative that greater surveillance equals greater security, and considering that there are three times as many private security agencies in South Africa as there are police, this market for control has developed in a pervasive and widespread manner. It is a fact, however, that the clients of these security agencies are predominantly white, while blacks do not even have the opportunity to complain and protest about excessive control. At the political level, there is no mention of the fact that inequalities are at the root of increased crime, and instead of investing in reducing them, this system of privatised control has been favoured. In fact, as an MIT Technology Review study suggests, cameras recreate "the digital equivalent of passbooks, or internal passports, an apartheid-era system that the government used to limit Black people's physical movements in white enclaves"[264]. Consequently, these instruments are used with the same colonial criteria and logic, contributing to the marginalisation of the poorest.

In contrast to the past, data colonialism is no longer limited to a single pole of colonial power, such as the West, but also involves new global players, such as China. This phenomenon complicates the traditional conception of the geography of the global

[263] M. KWET, 2019, op. cit
[264] K. HAO, H. SWART, *South Africa's private surveillance machine is fueling a digital apartheid,*in MIT Technology Review, 2022, https://www.technologyreview.com/2022/04/19/1049996/south-africa-ai-surveillance-digital-apartheid/ (September 20, 2024)

South, an idea that has so far distinguished resistance and identity on the basis of divisions between former colonisers and colonised. On the contrary, the new data colonialism operates on two levels: on a global level, extending its control everywhere, and on a domestic level, also influencing the populations of its home countries. The elites that drive this process benefit from colonisation in both dimensions and the North-South, East-West divisions no longer count in the same way[265].

3.2.2 The Empire of AI and Its Roots

Technological progress has often been criticised by scholars as a harmful ideology that, despite claiming to create a more equitable and interconnected world, actually reinforces and amplifies existing dominant modes of oppression. Generative artificial intelligence, with ChatGPT as one of its most emblematic examples, represents the latest incarnation of this ideology, celebrated for its "extraordinary potential" but closely linked to existing power dynamics.

Herbert Marcuse, who clearly articulated his concerns about the role of modern technology, stated that "specific purposes and interests of domination are not foisted upon technology "subsequently" and from the outside; they enter the very construction of the technical apparatus". He also criticised the idea of direct use of technology, arguing that "technology is always a historical-social project: in it is projected what a society and its ruling interests intend to do with men and things"[266].

In its current form, artificial intelligence is essentially seen as a means to generate profits: the technologies that have driven recent developments in AI, such as machine learning and deep neural networks, have become not only new tools of production, but also elusive entities that fuel speculation[267].

When considering the broader lifecycle of algorithms, materials, data, logistics, and knowledge, as well as the political, economic, cultural, and ideological structures that support them, we can realise how Artificial Intelligence has thoroughly permeated

[265] N. COULDRY, U. A. MEJIAS, 2019, op. cit.
[266] H. MARCUSE, *Negations: Essays in Critical Theory,* 1968, p. 168.
[267] C. ARADAU, M. BUNZ, *Dismantling the apparatus of domination?: Left critiques of AI,* in Radical Philosophy, no. 2.12, 2022, pp. 10–18

our existence, operating as a veritable empire. Indeed, we are not just living in the "age of Artificial Intelligence" but in the era of the AI empire, where this technology is shaped in a complex manner by political, historical, cultural, racial, gender, and class relations.

Regarding the "infrastructures of data and information empire," these are defined by Aouragh and Chakravartty as "both the material stuff of cables and wires that have long been seen as modern public goods as well as the 'soft' and more amorphous networks of cultural exchange shaped by European and American colonial power". These infrastructures encompass the totality of "both technical and cultural systems that create institutionalised structures whereby goods of all sorts circulate" and are "both central as digital nodes for financial transactions and trade, and key in squeezing down dissent or co-opting social movements"[268].

The expansion of imperialism through artificial intelligence goes far beyond the mere technological infrastructure. It is a complex and diffuse set of "actors, arrangements, technologies and logics" that operate in a decentralised manner in both time and space[269]. This system includes new and old mechanisms of domination, such as continuous surveillance, the exploitation of both physical and intangible labour, the collection and recording of sensory data and biological and social processes. These data are in turn translated into models for classification and prediction with the intention of controlling and manipulating human behaviour.

At the base of this AI empire, there is a persistent discourse on racism and colonialism, situated within a conception of technology that has prevailed in the West. However, as already mentioned, this global system is not only dominated by the West, in addition to the "great technological empires" of the United States, those of Europe and China also play a significant role.

It is also evident that the AI empire is closely intertwined with capitalism, with its relentless pursuit of profit and growth, which fuels both the Western neo-liberal model based on data, information and technology, and the Chinese model of a "digital empire", seemingly non-capitalist but nonetheless centred on unstoppable market

[268] M. AOURAGH, P. CHAKRAVARTTY, *Infrastructures of empire: Towards a critical geopolitics of media and information studies*, in Sage Journals, Media, Culture & Society, Vol. 38(4), 2016, pp. 559–575.
[269] E. ISIN, E. RUPPERT, *Data's empire: Postcolonial data politics*. (eds) Data Politics, 2019, pp. 207–227.

expansion. Several scholars have also highlighted the role of AI in perpetuating patriarchal structures, embodied by the toxic masculinity that pervades Silicon Valley culture[270].

These systems of oppression should, therefore, be considered as an interconnected whole in which the various components amplify and reinforce each other within the context of AI. Therefore, it is important to adopt a transnational and intersectional perspective to uncover their interconnectedness and political and historical interdependencies.

Feminist anthropology has long demonstrated how women's bodies have historically represented the "first colony", highlighting the deep link between colonialism and capitalism with patriarchal structures. As a result, it is no surprise that AI systems are replicating dynamics of exclusion, violence and discrimination. The increasing adoption of generative AI tools increases the risk of spreading false accusations based on misleading evidence, the amplification of misogynistic, homophobic or transphobic discourses and positions AI itself as both the cause of the problem and the potential solution to the same problem[271].

The links between algorithmic discrimination and colonial racism are perhaps among the most obvious. Algorithms, designed to automate procedures and formulated on the collection of data from an inherently racist society, tend to reproduce and perpetuate in their results and decisions the same racist outcomes[272].

The racist implication of AI - predictive policing and risk assessment of recidivism that has a detrimental effect on people of colour and marginalised groups - have become emblematic examples of the damage caused by algorithms. These mechanisms of excessive surveillance and discrimination, fuelled by AI, work together to perpetuate inequalities and contribute to the affirmation of racial hierarchies - concepts theorised as the "New Jim Code"[273]. Some scholars refer to the concept of

---

[270] J. TACHEVA, S. RAMASUBRAMANIAN, *AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order,* in Sage Journals, Big Data & Society, 2023 https://journals.sagepub.com/doi/10.1177/20539517231219241#bibr58-20539517231219241 (September 4, 2024)

[271] *Ibidem.*

[272] P. BENANTI, *Anche l'intelligenza artificiale va decolonizzata*, in Africa, 2020, https://www.africarivista.it/paolo-benanti-anche-lintelligenza-artificiale-va-decolonizzata/173072/ (September 4, 2024)

[273] R. BENJAMIN, Race After Technology: Abolitionist Tools for the New Jim Code, 2019

"Racial Capitalism" to highlight the connection between racism, white supremacy, and capitalism in the age of AI dominance. This system acts "not to homogenise but to differentiate—to exaggerate regional, subcultural, and dialectical differences into 'racial' ones"[274].

Thus, despite the formal end of colonialism, its consequences and the informal structures of colonial practices continue to persist. The industry no longer exploits labour through large-scale slavery, which was once justified by the spread of racist beliefs that dehumanised entire populations, but has developed new methods to exploit precarious, low-cost labour, often coming from the Global South, maintaining the implicit idea that these populations do not need, or are not deserving of, decent wages and economic stability[275].

Moreover, the dominance of the English language in the age of artificial intelligence is not an insignificant detail; on the contrary, language is a pillar of a community's identity, influencing and reflecting its culture, history, and worldview. Therefore, imposing English on non-English-speaking societies perpetuates colonial models of thought and behaviour, thus oppressing the creativity and self-determination of these populations. Nowadays, this function seems to be accomplished with even more far-reaching implications by the spread and utilisation of technologies like ChatGPT, described as "the fastest growing consumer application in history" and other forms of generative AI.[276].

3.2.3 How the AI Empire Operates

Our daily lives are increasingly shaped and influenced by advanced technologies that operate in often invisible but deeply pervasive ways. Among the key concepts to understand the impact of these emerging technologies, emerges that of extractivism, already addressed in the context of data colonialism. Traditionally linked to the exploitation of natural resources, the term extractivism takes on a new relevance in the

[274] J. TACHEVA, S. RAMASUBRAMANIAN, *2023, op. cit.*
[275] K. HAO, *Artificial intelligence is creating a new colonial world order,* MIT Technology Review, 2022 https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/ (September 4, 2024)
[276] J. TACHEVA, S. RAMASUBRAMANIAN, *2023, op. cit.*

digital world. Here, the concept refers to a mechanism that is no longer limited to the extraction of physical resources, but extends to the collection and appropriation of personal and collective data.

One of the central aspects of the extractive vision is the datification, defined by Couldry and Mejias as "quantifying human life through digital information, very often for economic value"[277]. Through this process, data is transformed into a commodity to be bought and sold in the big data market of the AI empire.

Another essential feature of the artificial intelligence empire is automation, especially in the field of cognitive work. Today, automation is not limited to making existing processes more efficient; it has the ambition of developing machines that not only match but exceed human capabilities, aiming at the creation of a superintelligence. This advanced technological evolution not only replaces manual work, but also extends to the intellectual one, radically transforming the very conception of work[278].

However, this growing reliance on algorithmic decision-making has revealed serious problems, especially for marginalised communities. Examples such as automated risk assessment show that racial and gender groups may be ignored or misrepresented due to lack of adequate data. For instance, in 2020, it was discovered that Zoom's video chat software, used to detect the face on a virtual background, was built on a facial recognition algorithm programmed with a default setting on whiteness. This problem was brought to light by the case of a black professor, whose face was constantly being "erased" and removed whenever he tried to use a virtual background during a conversation on Zoom, simply because the technology failed to recognise him correctly due to a "prototypical whiteness" that makes racialised subjects invisible[279].

Furthermore, despite the apparent "magic" of artificial intelligence algorithms such as ChatGPT, these systems are not fully automated. They are highly dependent on so-called "ghost work" of human annotators and moderators, often employed in regions of the world with lower labour costs. Former colonised countries, such as the Philippines, Kenya and India, have become ghost job centres for US and UK companies. Data annotation work, fundamental to supporting AI innovation, extends

[277] N. COULDRY, U. A. MEJIAS, 2019, op. cit.
[278] J. TACHEVA, S. RAMASUBRAMANIAN, 2023, op. cit.
[279] C. ARADAU, M. BUNZ, 2022, op. cit., pp. 10–18

and reflects the historical economic relationship between colonisers and colonised, perpetuating the dynamics inherited from their colonial histories[280].

The AI empire is characterised by an intrinsic essentialism that leads to the erasure of cultures, identities, people and communities. This approach reduces the complexity and diversity of social categories to a set of predefined "intrinsic traits", resulting in the representation of heterogeneous individuals—with plural and divergent values, interests, lifestyles, and moral and political commitments—as if they belonged to homogeneous groups.

In addition to the cancellation of diversity, this new empire relies on constant surveillance to determine the role and place of each individual in society. Personal data, such as skin colour, height, facial scans, and voiceprints, collected from surveillance cameras, are fed into integrated platforms. These data are then linked to national identification numbers to create unique and comprehensive personal profiles.

The cutting-edge technologies based on artificial intelligence, which are implemented to manage and control the population, are often presented under the veil of technological progress. In countries of the Global South, these technologies represent a test ground for AI before being exported worldwide. In this context, essentialism and surveillance are intertwined, as AI not only reduces people to stereotyped categories, but also uses advanced surveillance tools to consolidate and amplify these simplifications, perpetuating a uniform and limited vision of human identities.[281]

Roma residents in Europe, religious and ethnic minorities in China, members of marginalised castes in India and indigenous peoples in various regions of the world, among others, are not only subject to constant surveillance but are also labelled as groups "excess" to be controlled and confined.

For example, the Uyghurs, a Muslim minority in China's Xinjiang region, are subjected to close video surveillance by the Beijing government in the re-education camps[282] where they have been "deported". These surveillance systems are programmed

[280] P. BENANTI, 2020, op. cit.
[281] J. TACHEVA, S. RAMASUBRAMANIAN, *2023, op. cit.*
[282] According to the Chinese authorities, the thousands of Uyghurs who are removed from their villages and towns to be interned in the re-education camps are students and not prisoners. The aim of the stay in such facilities is brainwashing. The basic idea on the Chinese side is the transformation of thought: not to completely transform thought, but to eliminate its 'extremist elements'. Another underlying principle is to intern Uyghurs before they can commit a crime, thus arbitrarily deciding which individuals might break the law in the future.

to detect changes in various facial expressions, to recognise the emotional state of people, with the aim of controlling their behaviour and implementing preventive measures to contain people deemed "dangerous".[283]

The indigenous population have been subject to containment practices for centuries, but today artificial intelligence systems are amplifying and aggravating these processes. Although direct physical methods are no longer used, modern technologies continue to impose forms of control that are equally violent and dehumanising[284].

The colonial era's inherited geopolitical power imbalances have had a significant influence on artificial intelligence governance as well. This has been clearly demonstrated in the recent race to establish global ethical guidelines for AI. Developing countries in Africa, Latin America and Central Asia have largely been disregarded from the discussion tables, leading some of these countries to refuse to participate in international data flow agreements. As a result, developed nations continue to disproportionately benefit from the global standards that have been shaped in their favour, while developing countries remain significantly behind[285].

Finally, geopolitical power imbalances inevitably also influence the way in which AI is used to help developing countries. Big Tech companies are currently promoting and financing projects that claim to use datafication for the "social good". Behind this apparent benevolence, however, there is a deeper and more relevant objective: the progressive transformation of the social landscape, or at least of much of it, to consolidate the role of Big Tech as main providers of social solutions and knowledge.

The idea that Big Tech, operating in one part of the world and with a concentration of resources, can decide how to interpret and solve social problems on a global scale represents, in the light of colonial history, a considerable usurpation of power. No input from local communities is sought on this substitution of social knowledge or its implications for data collection and processing and they are not

---

[283] CF. N. BETRO, *Xinjiang: sorveglianza speciale per gli uiguri,* in Il Caffè Geopolitico, 2021, https://ilcaffegeopolitico.net/172007/xinjiang-sorveglianza-speciale-per-gli-uiguri (September 21, 2024), Amnesty International, *Cina, rapporto di Amnesty International: crimini contro l'umanità ai danni dei musulmani dello Xinjiang,* 2021, https://www.amnesty.it/cina-rapporto-di-amnesty-international-crimini-contro-lumanita-ai-danni-dei-musulmani-dello-xinjiang/ (September 21, 2024)

[284] L. VERACINI, *Containment, Elimination, Endogeneity: Settler Colonialism in the Global Present,* Rethinking Marxism, Vol. 31(1), 2019, pp. 118–140.

[285] P. BENANTI, 2020, op. cit.

consulted. As a result, the freedom of populations to define their own concept of social good and their version of social knowledge is systematically ignored[286].

Initiatives such as "AI for good" or "AI for sustainable development" often adopt a paternalistic approach, forcing developing countries to depend on existing AI systems, rather than participating in the design of new systems adapted to their specific context[287].

The AI Empire stands on a worldview where some lives, cultures and ways of being are considered more valuable than others. With its omnipresent presence, apparent efficiency and deceptive objectivity, it perpetuates a sophisticated form of oppression. Marginalised communities around the world continue to bear the brunt of this automated violence.

3.3 The Role of AI in Perpetuating the Dynamics of Oppression on Indigenous Peoples

3.3.1 The Ewert v. Canada Case: Artificial Intelligence and the Persistence of Colonialism in Legal Systems

The case of *Ewert v. Canada*[288] highlights the deep link between algorithmic discrimination and the legacy of colonialism in the Canadian judicial system. Jeffrey Ewert has been incarcerated for more than 30 years in Canadian federal prisons, where he is serving two life sentences for murder and attempted murder. He is a member of the indigenous *Métis* community, to which he belongs both ethnically and culturally.

In 2007, for the first time, he raised an issue regarding the use of actuarial risk assessment tools employed by the Correctional Service of Canada (CSC) that measure violent and sexual recidivism and psychopathy. The CSC is a federal government body in Canada in charge of the operation of all prisons and detention facilities housing

[286] J. VIERA MAGALHAES, N. COULDRY, *Giving by taking away: Big tech, data colonialism and the reconfiguration of social good*. International Journal of Communication, 2021, Vol. 15, pp. 343–362.
[287] P. BENANTI, 2020, op. cit.
[288] *Ewert v. Canada*, 2018 SCC 30, [2018] 2 S.C.R. 165. In literature cf. E. HILL, J. WOLFE, E*wert v. Canada: Shining Light on Corrections and Indigenous People*, in The Supreme Court Law Review: Osgoode's Annual Constitutional Cases Conference, Vol. 94, 15, 2020, pp. 391-413; A. M. HAAG et al, *An introduction to the issues of cross-cultural assessment inspired by Ewert v. Canada*, in Journal of Threat Assessment and Management, Vol. 3, 2, 2016, pp. 65–75

persons sentenced to two years or more in confinement. Its mandate revolves around safe custody, rehabilitation, and facilitation to enable reintegration into society[289].

Mr. Ewert subsequently continued to file a number of similar complaints, accompanied by internal appeals against the decisions relating to these complaints. All these complaints and appeals were essentially centred on the same issue, as explained by Judge Beaudry in the 2007 decision of the Federal Court:

"These risk assessment instruments were designed by and for western people and when they are used in assessing Aboriginal offenders they produce a discriminatory effect that places Aboriginal prisoners in a disadvantaged position in the federal correctional system. Mr. Ewert characterised these assessment tools as racist and a contributing factor to the over representation of Aboriginal peoples in Canadian correctional institutions."[290]

Mr Ewert argued that these instruments, although presented as impartial, were in fact inherently discriminatory against indigenous peoples and violated the Corrections and Conditional Release Act (CCRA) [291], and also Sections 7[292] and 15[293] of the Charter of Rights and Freedoms [294].

The working papers recommended that indigenous people be recognised as a particularly disadvantaged group and that it would be necessary to address differences in results. Additionally, they recommended that certain components of corrective operations and programs be coded and designed to specifically address the requirements of indigenous detainees. When the Parliament adopted the CCRA in 1992, these proposals were implemented. The CCRA represents a significant revision of the Canadian criminal justice system. It was created with the intention of providing a

---

[289] Correctional Service Canada (CSC), officially established on 10 April 1979, https://www.canada.ca/en/correctional-service.html

[290] *Ewert v. Canada (Attorney General)*, 2007 FC 13, at paras. 7-14

[291] *Corrections and Conditional Release Act* (S.C. 1992, c. 20), hereinafter CCRA. An Act concerning corrections, conditional release, and the detention of offenders, and to create the office of the Correctional Investigator.

[292] *Canadian Charter of Rights and Freedoms,* 1982, section 7: *"Everyone has the right to life, liberty and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice."* https://laws.justice.gc.ca/eng/const/page-12.html

[293] *Canadian Charter of Rights and Freedoms,* 1982, section 15. (1) *"Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability."* https://laws.justice.gc.ca/eng/const/page-12.html

[294] *Canadian Charter of Rights and Freedoms, Part I of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (U.K.), 1982, c. 11, hereinafter "Charter".* https://laws.justice.gc.ca/eng/const/page-12.html

precise legal framework for the treatment of individuals serving sentences in federal prisons and their final conditional release programs for reintegration into society. The need to treat all categories of detainees fairly and justly, including indigenous peoples and other ethnic minorities, who have historically received unfair treatment in the criminal justice system, was one of the driving forces behind the creation of this document. In reality, there are special provisions for indigenous people in the corrective system. In particular, Article 4(g) declares:

"The principles that guide the Service in achieving the purpose referred to in section 3 are as follows:

(g) correctional policies, programs and practices respect gender, ethnic, cultural, religious and linguistic differences, sexual orientation and gender identity and expression, and are responsive to the special needs of women, Indigenous persons, visible minorities, persons requiring mental health care and other groups..."[295]

The Supreme Court decision in Ewert highlights the stark contrast between the reality experienced by indigenous peoples in the Canadian penitentiary system and the stated objectives of the Corrections and Conditional Release Act.

Although the CCRA aims to promote fair and rehabilitative treatment for all prisoners, in the case of indigenous peoples, often disadvantaged and victims of systemic discrimination, these objectives are difficult to achieve in practice.

Despite the arguments presented by Mr. Ewert in 2007, all complaints and appeals were rejected by the CSC, which argued that the actuarial instruments used are valid risk predictors.

Subsequently, in 2015, Mr. Ewert began a new proceeding before the Federal Court to challenge the continued use of the same assessment tools by the CSC, reiterating that the use of these tools violated the requirement of the CCRA to consider the "special needs" of indigenous prisoners and also the obligation, provided for in Article 24(1), according to which: "The Service shall take all reasonable steps to ensure that any information about an offender that it uses is as accurate, up to date and complete as possible."[296]

[295] *Corrections and Conditional Release Act,* CCRA (S.C. 1992, c. 20), section 4(g)
[296] *Ibidem,* section 24(1)

Despite the statements of Judge Phelan who had stated that the "actuarial tests are susceptible to cultural bias and therefore are unreliable"[297] and that the continued use of these tests by the CSC, despite legitimate concerns, violated both Article 24(1) of the CCRA and Article 7 of the Charter, the Federal Court of Appeal, in examining the CSC's appeal, concluded that Mr. Ewert had failed to demonstrate with sufficient evidence that "the assessment tools produce or are liable to produce erroneous results and conclusions"[298].

Thus, after being initially heard at first instance[299] and then dismissed on appeal[300], the appeal has now reached the Canadian Supreme Court.

In October 2017, Mr. Ewert again challenged the use of five different tools[301] to assess his psychopathy and the risk of violent and sexual recidivism[302].

The tools used are largely based on static data collections, which include elements such as age at the time of crime, work and education history, marital and family status, criminal history, and substance abuse. These data are unchanging and do not take into account the possible significant changes that a person may have made in his or her lifetime, leading to a higher risk score. Furthermore, as the Supreme Court recognised in *R. v. Ipeelee*[303], the lasting influence of colonial oppression and the social conditions that have resulted from it continue to contribute to the high incidence of incarceration among aborigines. The specific context of indigenous peoples, including trauma and experiences related to crime, is not taken into account and therefore the reliability of data and systems is doubtful and their use may lead to systemic discrimination[304].

One crucial aspect of the issue raised by Ewert is that the technologies used have been mainly trained and applied for assessments of people not belonging to indigenous

---

[297] *Ewert v. Canada (Correctional Service),* 2015 FC 1093, para. 75
[298] *Ewert v. Canada (Correctional Service),* 2016 FCA 203, para 21
[299] Federal Court (Phelan J.), 2015 FC 1093, 343 C.R.R. (2d) 15.
[300] Federal Court of Appeal (Dawson J.A., Nadon and Webb JJ.A. Concurring), 2016 FCA 203, 487 N.R. 107.
[301] The actuarial tools being referenced are the Hare Psychopathy Checklist — Revised (PCL-R), Violence Risk Appraisal Guide (VRAG), Sex Offender Risk Appraisal Guide (SORAG), Static-99, and Violence Risk Scale — Sex Offender (VRS-SO).
[302] *Ewert v. Canada*, 2018 S.C.J. No. 30, [2018] 2 S.C.R. 165, at para. 11
[303] *R. v. Ipeelee*, 2012 SCC 13, [2012] 1 S.C.R. 433, at para. 60
[304] D. MILWARD, *Sweating it Out: Facilitating Corrections and Parole in Canada Through Aboriginal Spiritual Healing,* 2011, p. 47

minorities. Therefore, Ewert argues that these tools are not sufficiently accurate for its case, as the tests themselves are subject to cultural prejudice.

The Supreme Court has recognised as valid the concerns raised by Mr. Ewert since 2000. The majority of judges, represented by Judge Wagner, pointed out that contrary to what the CSC claimed, the test results fall within the concept of "information" as defined in Section 24. For this reason, the CSC had a legal obligation to "take all reasonable steps" to ensure that the information produced by the instruments was accurate when applied to indigenous people in the prison system. Furthermore, it was found that the CSC has never conducted audits on the reliability of risk assessment tools when used for decisions concerning members of ethnic and cultural minorities, despite specific concerns raised in this regard.[305]

The *majority opinion* also pointed out that there are statistical data showing less favourable treatment of indigenous people by the Canadian justice system. These defendants tend to receive more severe penalties and worse risk assessments than others.

The ruling states that the growing use of advanced technologies may amplify these discriminations, making them more difficult to detect than is currently the case. This risk is due to the fact that these technologies may mask discrimination behind apparent scientific and technical objectivity[306].

"In the context of the case at bar, this required, at the very least, that the CSC take seriously the credible concerns that have been repeatedly raised according to which information derived from the impugned tools is of questionable validity with respect to Indigenous inmates because the tools fail to account for cultural differences. By disregarding the possibility that these tools are systematically disadvantaging indigenous offenders and by failing to take any action to ensure that they generate accurate information, the CSC fell short of what it is required to do under s. 24(1) of the CCRA"[307].

[305] *Ewert v. Canada*, 2018 SCC 30, [2018] 2 S.C.R. 165, at para. 45-67
[306] *Ewert v. Canada*, par. 65, "*Thus, the clear danger posed by the CSC's continued use of assessment tools that may overestimate the risk posed by Indigenous inmates is that it could unjustifiably contribute to disparities in correctional outcomes in areas in which Indigenous offenders are already disadvantaged. […]"*
[307] *Ewert v. Canada*, par. 66

The Supreme Court has recognised that these algorithms do not take into account the historical impact of colonialism, which left deep traumas in indigenous peoples and caused social inequalities and excessive criminalisation of indigenous communities. This case may demonstrate how the use of algorithmic tools can continue to perpetuate forms of systemic discrimination and highlights the need to address cultural biases in order to ensure a more equitable and impartial treatment in the criminal justice system.

It is interesting to note that Statistics Canada reported that even when compared to non-aboriginal individuals who have similar socio-demographic characteristics, the risk of victimisation for aboriginal individuals was 58% higher than for non-aboriginal individuals. It also showed that the victimisation rate of aboriginal women is almost three times higher than that of non-aboriginal women[308].

For this reason, the introduction and use of risk assessment tools such as those challenged by Ewert in criminal proceedings must be accompanied by particularly rigorous checks, which The Canadian Supreme Court found to be lacking in this case. Human control of AI is essential at all stages of its implementation to ensure a fair role and monitor its effectiveness in providing "justice", especially for historically disadvantaged communities, particularly indigenous peoples.

3.3.2 Indigenous Data and Colonialism: Toward a Global Response for Self-Determination

"Indigenous Peoples have always been "data warriors". Our ancient traditions recorded and protected information and knowledge through art, carving, song, chants and other practices"[309].

Before the arrival of the imperial powers, the indigenous peoples possessed a rich heritage of knowledge and had gathered an important body of data that was under

[308] S. PERREAULT, *Violent victimization of Aboriginal people in the Canadian provinces, 2009,* Component of Statistics Canada catalogue no. 85-002-X, Juristat, 2011 https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2011001/article/11415-eng.pdf?st=niePbvho (September 7, 2024)

[309] T. KUKUTAI, J. TAYLOR, *Data sovereignty for indigenous peoples: current practice and future needs*, in T. KUKUTAI, J. TAYLOR (eds), *Indigenous Data Sovereignty: Toward an agenda*, 2016, pp. 1-22

their full control. The collection and preservation of information was an integral part of most, if not all, indigenous cultures, manifested through cultural practices passed down from generation to generation. With colonialism, however, the systems of knowledge of indigenous peoples were progressively supplanted, at least in public discourse, by the data and narratives imposed by the imperial powers and their colonisers[310].

As previously stated, artificial intelligence is still built upon pre-existing data, which has traditionally overlooked, distorted, silenced, stereotyped, and marginalised indigenous peoples' knowledge, cultural expressions, resulting in their oppression and deprivation of political rights. Thus, aborigines and other groups affected by settlement colonialism and its narratives continue to be exposed to prejudice and discriminatory policies in constant evolution, further strengthened by the increasing use of artificial intelligence systems.

Some scholars argue that AI could open the way to serious consequences if its management does not focus and prioritise the experiences of those living on the margins of society. The idea of technological progress, promoted under the name of "development", represents an additional form of oppression for those who have already suffered the effects of the "developments" imposed by the colonising states on their ancestral lands[311].

To function properly, artificial intelligence needs large amounts of data – "big data". To effectively predict results, an algorithm must be able to accurately reflect the precise representations of the problems a community faces based on available data.

Indigenous communities, which have already been victims of violence legitimated by colonial narratives, are increasingly concerned about the collection, analysis, and use of data concerning them. This is because, data as a form of narration "plays a powerful role in constituting reality through their underpinning methodologies by virtue of the social, cultural, and racial terrain in which they are conceived, collected, analysed, and interpreted."[312]

---

[310] I. POOL, *Colonialism's and postcolonialism's fellow traveller: the collection, use and misuse of data on indigenous peopl*e, in T. KUKUTAI, J. TAYLOR (eds), Indigenous Data Sovereignty: Toward an agenda, 2016, pp. 57-76

[311] N. SANGMA, *Artificial Intelligence and Indigenous Peoples' Realities*, in Cultural Survival, 2024, https://www.culturalsurvival.org/publications/cultural-survival-quarterly/artificial-intelligence-and-indigenous-peoples-realities (September 8, 2024)

[312] M. WALTER, C. ANDERSEN, *Indigenous Statistic*: *a Quantitative Mesearch Methodology,* 2013, p. 9

Above all, the extractive nature of AI, which is based on indiscriminate data collection, evokes concerns about cultural appropriation. This phenomenon particularly affects indigenous communities whose culture has historically been the subject of theft and misappropriation[313]. Colonial governments in Australasia and North America, for example, have accumulated a great deal of information on indigenous peoples, using these statistics as "evidence" to monitor populations and justify political interventions[314]. Data are continuously extracted from native communities, without their input or permission on how they are collected, used or applied[315].

The management of indigenous data and narratives is often in the hands of non-indigenous actors, which creates significant gaps in the representation of these peoples' experiences. Indigenous people are frequently described through the lens of deficits, disparities and disadvantage. This external control perpetuates distorted narratives, which do not accurately reflect the historical, social and political forces that have shaped and continue to influence their lives. Such narratives, often used to justify injustice towards indigenous peoples, were created to serve the violent designs of colonialism, concealing the injustices suffered and denying the need for adequate remedies[316]. In addition, they contribute to perpetuating the disappearance of native identities.

Data collected and monitored from non-indigenous sources often leads to erosion of visibility and recognition of aboriginal groups and individuals in various ways. First, these populations are erased when governments fail to collect and provide comprehensive data on these communities, often driven by political, economic or social interests. Or, this is where the data are not disaggregated – that is, not broken down into categories such as gender, region or tribe. By using aggregated data, the complexities of indigenous experiences, including concepts such as sovereignty and self-determination, are overlooked. The aggregated data do not reflect the diversity between and within indigenous communities, and often these "are regularly excluded from study or put in

[313] R. CHANDRAN, *FEATURE-Indigenous groups in NZ, US fear colonisation as AI learns their languages*, in Reuters, Media & Telecom, 2023 https://www.reuters.com/article/idUSL8N2UQ0EC/ (September 9, 2024)
[314] T. KUKUTAI, M. WALTER, *Recognition and indigenizing official statistics: Reflections from Aotearoa New Zealand and Australia,* in Statistical Journal of the IAOS 31, 2015, pp. 317–326
[315] T. KUKUTAI, J. TAYLOR, 2016, op. cit.
[316] I. FALEFUAFUA TAPU, T. KAMAILELAULI'I FA'AGAU, *New Age Indigenous Instrument: Artificial Intelligence & Its Potential for (De)colonialized Data*, in Harvard Civil Rights-Civil Liberties Law Review Vol. 57, 2022, pp. 739-741

the catch-all miscellaneous category of 'other'"[317], thus becoming invisible. Invisibility contributes to perpetuating the erroneous idea that native peoples no longer suffer from racism or discrimination. However, indigenous communities continue to suffer from deep-rooted injustices and suffering that many non-indigenous people find it difficult to recognise.

These systems therefore seem to be built with the intention of marginalising, and the data used to erode and damage native communities for political, ideological and social gain[318]. So, if indigenous data is managed by non-indigenous people and the natives do not have sovereignty over their own data, "they will simply be recolonised in this information society"[319].

The concept of "indigenous data" goes beyond the traditional meaning of "data" as it includes information, data and knowledge that have an impact on indigenous peoples. According to some scholars, "the boundaries between data, information and knowledge, as defined in the western context, are much more fluid in the indigenous world"[320]. Indigenous data, therefore, includes much more than bits and bytes[321], extending to elements such as the cultural heritage contained in the languages, traditions, practices, technologies, natural resources and territories of indigenous communities.

In response to poor data management practices by non-indigenous peoples, the Indigenous Data Sovereignty (ID-SOV) movement was born, which deals with every stage from collection of data on indigenous peoples to communication. The ID-SOV affirms the right of indigenous peoples and nations to own, control, access and use data concerning their members, traditions, territories and natural resources[322].

The sovereignty of indigenous data is founded on the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP)[323], an international

---

[317] K. GOODLUC, *The Erasure of Indigenous People in U.S. COVID-19 Data,* in HIGH COUNTRY NEWS, 2020, https://www.hcn.org/articles/indigenous-affairs-the-erasure-of-indigenous-people-in-us-covid-19-data/ (September 8, 2024)
[318] I. FALEFUAFUA TAPU, T. KAMAILELAULI'I FA'AGAU, 2022, op. cit., pp. 742-746
[319] R. CHANDRAN, 2023, op. cit.
[320] S. R. CARROL et al, *Indigenous Data Governance: Strategies from United States Native Nations, in Data Science Journal*, Vol 18 (31), 2019, p. 2
[321] *Ibidem*
[322] R. LOVETT et al, *Good Data Practices for Indigenous Data Sovereignty and Governance,* 2018, pp 26-35
[323] United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), adopted by the General Assembly on 13 September 2007

document adopted in 2007 after a quarter-century of work to define minimum standards for the protection of the rights of indigenous communities. In particular, articles 18 and 19 of the declaration highlight the right of indigenous peoples to participate in decisions that affect them through their modalities and procedures. In this context, the sovereignty of indigenous data provides a real opportunity to apply these articles, allowing communities to exercise control over their own data[324].

Inspired by the First Nations OCAP® principles[325], which in the 1990s set standards for ownership, control, access and possession of data in Canada, the Indigenous Data Sovereignty Networks in New Zealand, Australia and the United States, together with Aboriginal scholars, leaders and allies, have recognised the urgent need to create global principles for the governance of these data[326].

The current trend towards open data and open science does not adequately address the rights and interests of natives. Current principles of open data movement, such as the FAIR model[327] (findable, accessible, interoperable, and reusable), focus mainly on making data more easily shared between different entities, but they overlook the power inequalities and historical contexts that affect indigenous communities.

Thus, the CARE principles for indigenous data governance[328] were formulated by the International Data Sovereignty Interest Group, a network within the Research Data Alliance[329]. The four principles – Collective Benefit, Authority to Control, Responsibility, and Ethics – oppose data extraction issues in an attempt to move toward decolonisation of data.

The CARE principles are based on the idea that the data from indigenous communities should be used in such a way that has the real benefits for those communities, ensures equitable development, increases innovation, allows better

---

[324] *Ibidem*

[325] The First Nations developed a new model that established collective and broad-based control of their data.

[326] S.R. CARROL et al, *The CARE Principles for Indigenous Data Governance*, in Data Science Journal, Vol. 19 (43), 2020, pp. 1–12 https://datascience.codata.org/articles/10.5334/dsj-2020-043 (September 9, 2024)

[327] Principles FAIR were developed in 2014 and published in 2016. https://force11.org/info/the-fair-data-principles/

[328] Research Data Alliance, International Indigenous Data Sovereignty Interest Group, September 2019, *CARE Principles for Indigenous Data Governance*, The Global Indigenous Data Alliance. https://www.gida-global.org/care (September 9, 2024)

[329] Research Data Alliance, International Indigenous Data Sovereignty Intertest Group, 2017, https://www.rd-alliance.org/groups/international-indigenous-data-sovereignty-ig (September 9, 2024)

decision making, and fosters local expertise. Beyond that, such standards also emphasise respect for indigenous traditional knowledge and practices, and ensure that communities are not excluded from the process.

In order to ensure more equitable and inclusive governance, it is essential that indigenous peoples have sovereignty over their data and that their needs and ambitions are taken into account. It is only through control of their data that natives can preserve their traditional knowledge and protect their cultural identity. It is also necessary for them to play an active role in the management of their own resources so as to initiate a process of decolonisation of artificial intelligence systems. This sovereignty represents a crucial step towards the recognition of their self-determination and the construction of a more prosperous and sustainable future, in which indigenous peoples are not only the recipients and victims of decisions but also protagonists of their social, economic and cultural development.

Conclusion

The increasing dependence on technology is rapidly spreading into all aspects of daily life, both public and private. Linked to this phenomenon are many opportunities and innovations, but also significant risks. In particular, there is a risk that technology might exclude and marginalise people and their rights. First among the most at-risk rights are those covering the principles of equality and nondiscrimination. The more established this technology becomes, the very real danger exists that the old inequalities and dominance only become magnified, and new forms of discrimination and colonisation will manifest themselves and further undermine any notions of equitable access to resources, opportunities, and rights of those already experiencing social and economic disadvantage.

As we have seen, artificial intelligence systems are far from neutral or objective. In contrast, they embody the prejudices of their creators, this group of developers is often relatively homogeneous, mainly composed of young engineers and entrepreneurs, predominantly white, men and well-off, who bring with them their unique worldview.[330]

[330] J. TACHEVA, S. RAMASUBRAMANIAN, 2023, op. cit.

This elite-dominated vision becomes the foundation for the architecture of the AI empire.

In the case of *Ewert v Canada*, it became clear how algorithmic discrimination is closely linked to the historical legacy of colonialism persisting in the Canadian judicial system. It cannot be denied that artificial intelligence is very attractive for its potential to predict court decisions, but it also poses the risk of aggravating existing barriers and discriminations indigenous communities face in legal systems. Predictive algorithms often provide results that are "neither justice nor predictive"[331].

Before courts and legal systems start to rely on artificial intelligence, it is essential that those developing these technologies take into account narratives that have historically harmed indigenous communities by depriving them of the right to self-determination. Without data that reflect the reality and specific conditions of indigenous peoples, non-indigenous researchers risk, whether intentionally or not, filling gaps with their own prejudices based on colonial stereotypes. Like other applications of artificial intelligence, risk assessment tools used in criminal proceedings are also based on historical data. Unless these distortions are corrected, AI will continue to perpetuate and legitimise violence and ignore the experiences of indigenous communities.

For this reason, guaranteeing the sovereignty of data for indigenous people and involving the communities concerned directly in the elaboration and management of technologies supported by artificial intelligence becomes of utmost importance. The rights of self-determination of indigenous peoples require respectful representations that reproduce the reality of their data and experiences, rather than being distorted by external prejudices.

It is only with genuine respect for data sovereignty and inclusive collaboration that we can actually build systems representative of diversity and the realities of indigenous communities, making sure that their voices are respected and included within all the spheres of decision-making. Rooted in unique community and ethical visions, the indigenous perspective contributes valuably to tackling the challenges arising in the wake of artificial intelligence systems. The absence of the inclusion of such ways would be a serious lost opportunity for the global community, which could henceforth be deprived of a more holistic and inclusive ethical way.

---

[331] A. D. REILING, *Courts and Artificial Intelligence,* 11 International Journal for Court Administration 8, 2020, p. 4

CONCLUSIONS

It is evident that we are in the midst of a digital technological revolution driven by Artificial Intelligence and it is clear that the spread of this technology is unstoppable. This is why it is necessary to pose questions and to look at artificial intelligence systems not as neutral and impartial systems, but as systems that have a profound impact on the daily lives of individuals and also on global society.

To what extent is AI the bearer of neutrality? And to what extent, on the other hand, is it the daughter of cultural prejudices and thus the bearer of Western, white, wealthy thinking? Can AI be discriminatory and be used as an instrument of colonial domination?

To seek answers to these complex questions, I first attempted to reconstruct the evolution of the concept of Artificial Intelligence, its origins and definitions. It emerged that, in reality, to date there is no universally agreed definition of AI and this already represents a not insignificant epistemological problem. The analysis of the AI governance of the European Union, the United States, and China, the major players in the race for global digital regulation, then highlighted the differences in the orientations that have guided them and that we can respectively identify as “rights-based”, “market-based”, and “state-based” models.

Despite significant national and international developments in the regulation of AI, any attempt to govern artificial intelligence on a global scale faces major obstacles posed by the great Powers and Big Tech, committed to reaping the political, social and technological benefits deriving from AI.

Examining more closely the algorithms on which AI is based, which in turn are based on data collection, it became apparent that there are five moments in the programming or functioning of AI techniques from which disproportionately unfavourable results can arise and have a discriminatory effect on marginalised groups such as indigenous communities. It was always thought that algorithms were neutral tools because they were based on objective data, and that the use of digital technologies reduced human error by producing results that were accurate, fast and adhered to the objective reality of things. In actual fact, however, data is inherently human and earthly because it is created by man himself. Therefore, it may not accurately represent reality

and may reflect pre-existing social and cultural prejudices and expectations. Once it is shown that the cultural element affects both human conduct and the data itself, i.e. that the link with culture is not external but internal to the data, it is evident that the causes of the possible discriminatory effects of AI are similar to those of traditional discrimination.

Through the analysis of the *Ewert v. Canada* case, it was demonstrated that the AI systems used for the psychological assessment and violent and sexual recidivism of indigenous inmates are unable to guarantee fair and non-discriminatory treatment. The Supreme Court ruling states that the increasing use of advanced technologies may amplify these discriminations, making them more difficult to detect than is currently the practice. This risk is due to the fact that such technologies can mask discrimination behind an apparent scientific-technical objectivity. The Supreme Court has recognised that these algorithms do not take into account the historical impact of colonialism, which left profound traumas in indigenous peoples and caused social inequalities and excessive criminalisation of indigenous communities.

Therefore, it can be said that traditional and algorithmic discrimination have the same root in the cultural legacy of classical colonialism, which is now reappearing in a new form called Data Colonialism. Indeed, as stated by Couldry and Mejias, Data Colonialism is a "term for the extension of a global process of extraction that started under colonialism and continued through industrial capitalism, culminating in today's new form: instead of natural resources and labor, what is now being appropriated is human life through its conversion into data".

If historical colonialism entailed the annexation of territories, natural resources and populations, data colonialism entails that human life, through digital platforms, acquires a market value intertwined with the evolution of today's capitalism. Thus, data colonialism combines the extractive and predatory practices of traditional colonialism with modern digital quantification techniques. Social and global life becomes an open resource and personal data are equated with an extractable natural resource subject to exploitation for the benefit of capitalism. Thus, just as the concept of nature and land, which in the space of four centuries, from the Enlightenment to the present day, has gone from being a living organism to a mere inert extractable resource, contributing to a

purely extractivist view of the world, today that same view has shifted to bodies and life itself.

It should not be forgotten that colonialism and capitalism have their roots in a patriarchal system that is culturally constructed on women's bodies and, therefore, on bodies/data constitutively considered inferior. Patriarchy as a system is not an additional system, it is not the product of capitalism, it is not a consequence of colonisation, it is not a form of racism, but as Adriana Guzman, a Bolivian feminist political activist and founder of Community Feminism, states, it is the system that produces all the discrimination experienced by humanity and nature.

It is no coincidence that many of the most in-depth studies on the discrimination acted out by AI and its algorithms have been carried out by women and indigenous women precisely because they bring a different worldview from the dominant extractivist one.

Therefore, in order to try to decolonise data and thus liberate it from colonial prejudices and discrimination, it is certainly important to continue working in the area of AI regulation, but a radical change in the way of thinking and looking at the world is also necessary. The rationality on which colonial thinking was based experienced as absolute and universal must be deconstructed. To do so requires that everyone participate at the decision-making table, that equal space be given to visions of forms of life other than those of dominant groups, that the diversity and richness of human experience be expressed, and that the values of all communities be incorporated. The exclusion of indigenous perspectives and socially marginalised groups risks biasing AI systems towards the experiences and views of dominant groups and helps perpetuate a form of cultural myopia.

Decolonising data is an exercise in imagination and creativity deeply connected to alternative epistemologies. Datafication produces homologation by subjecting us to total algorithmic control, while resistance to data colonialism must revalue the idea of the "other" and the heterogeneity of reality. Despite the complexity of this technological scenario, we must remember that AI is a human endeavour, enriched by this idea of heterogeneity and the contributions of all communities.

The great power that multinational corporations have acquired by collecting huge amounts of data and the influence they have on the policies of governments makes

the introduction of this new approach extremely complex, but individuals and territories in large parts of the world should not be victims and bear the price of empire and the development of AI. There is an ethical need to create alternative pathways that involve deconstructing the reinforcement of a colonialist, patriarchal, and capitalist global order, but it is a fact that many debates are already underway and that movements of resistance to colonialism are joining struggles for dignity and justice.

BIBLIOGRAPHY

A. BAIG, *Brazil's New AI Law: What You Should Know,* in Securiti AI, 20 march 2024. Last visited July 13, 2024. https://securiti.ai/brazil-ai-regulation-and-law/

A. ENGLER, *The EU AI Act will have global impact, but a limited Brussels Effect*, in Brookings, 2022. Last visited May 21, 2024. https://www.brookings.edu/articles/the-eu-ai-act-will-have-global-impact-but-a-limited-brussels-effect/

A. L. SAMUEL, *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal of Research and Development, 1959, pp. 206-226.

A. LACASSE, *Canadian Parliament's Bill C-27 hearing delves deeper into AIDA*, IAPP, 2023. Last visited July 13, 2024. https://iapp.org/news/a/canadian-parliaments-bill-c-27-hearing-delves-deeper-into-aida

A. M. HAAG, A. BOYES, J. CHENG, A. MACNEIL, R. WIROVE, *An introduction to the issues of cross-cultural assessment inspired by Ewert v. Canada*, in Journal of Threat Assessment and Management, Vol. 3, 2, 2016, pp. 65–75

A. NEWELL, H. A. SIMON*, Computer science as empirical enquiry: Symbols and search*, Vol. 19, No. 3, 1976, pp. 113-126

A. NEWELL, J. C. SHAW, H. A. SIMON, Report on a general problem-solving program, 1959

A. SALMAN, *Reti neurali artificiali: dal MLP alle più recenti architetture di Convolutional Neural Networks*, 2017, p 1.

A. TURING, *Computing machinery and intelligence*, in *Mind*, *Mind*, vol. 59, no. 236, 1950

A.D. REILING, *Courts and Artificial Intelligence,* 11 International Journal for Court Administration 8, 2020, p. 4

*Access Alert | Brazil's New AI Bill: A Comprehensive Framework for Ethical and Responsible Use of AI Systems,*in Access Partenrship, 5 May 2023. Last visited July 13, 2024. https://accesspartnership.com/access-alert-brazils-new-ai-bill-a-comprehensive-framework-for-ethical-and-responsible-use-of-ai-systems/

AI: Narrow AI vs. General AI, 2018. Last visited May 16, 2024. https://www.gavinjensen.com/blog/2018/ai-narrow-vs-general

Amnesty International, *Cina, rapporto di Amnesty International: crimini contro l'umanità ai danni dei musulmani dello Xinjiang,* 2021. Last visited September 21, 2024. https://www.amnesty.it/cina-rapporto-di-amnesty-international-crimini-contro-lumanita-ai-danni-dei-musulmani-dello-xinjiang/

B. WEBER, *Swift and slashing, computer topples Kasprov,* in New York Times, May 12, 1997. Last visited May 11, 2024. https://www.nytimes.com/1997/05/12/nyregion/swift-and-slashing-computer-topples-kasparov.html

Bill C-27, *An Act to enact the Consumer Privacy Protection Act, the Personal Information and Data Protection Tribunal Act and the Artificial Intelligence and Data Act and to make consequential and related amendments to other Acts, short title: Digital Charter Implementation Act, 2022*, November 22, 2021, https://www.parl.ca/LegisInfo/en/bill/44-1/C-27

C. AGATA, *Intelligenza Artificiale, Big Data e nuovi diritti*, in Rivista Italiana di informatica e diritto, vol. 4, no. 1, 2022, pp. 94-97

C. ARADAU, M. BUNZ, *Dismantling the apparatus of domination?: Left critiques of AI,* in Radical Philosophy, no. 2.12, 2022, pp. 10–18

C. NARDOCCI, *Intelligenza Artificiale e Discriminazioni,* 2021, Convegno annuale dell'associazione "Gruppo di Pisa", Il diritto costituzionale e le sfide dell'innovazione tecnologica, 18 e 19 giugno 2021

C. O'NEIL, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy,* 2016

Canadian Charter of Rights and Freedoms, entered into force April 17, 1982*, https://laws.justice.gc.ca/eng/const/page-12.html*

Charter of Fundamental Rights of the European Union (2000/C 364/01), proclaimed on 7 December 2000 and became legally binding when the Treaty of Lisbon entered into force on December 1, 2009

China Academy of Information Technology CAICT, JD Explore Academy, White Paper on Trustworhty artificial Intelligence, july 2021. Last visited July 7. http://www.caict.ac.cn/english/research/whitepapers/202110/P020211014399666967457.pdf

Chinese Ministry of Foreign Affairs, *Global AI Governance Initiative,* October 20, 2023. Last visited August 28, 2024. https://www.fmprc.gov.cn/eng/xw/zyxw/202405/t20240530_11332389.html

CISV, Associazione Italiana per l'Intelligenza Artificiale, Università degli studi di Bari, ONG 2.0, *L'intelligenza Artificiale per lo Sviluppo Sostenibile*, 2021

Code of Ethics for the New Generation of Artificial Intelligence, *新一代人工智能伦理规范》发,* 26 september 2021. Last visited July 7, 2024. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html

Committee of Minsters, *Council of Europe adopts first international treaty on artificial intelligence*, 17 May 2024. Last visited June 4, 2024.

https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence

COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI), *Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law*, July 2023

COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI), *Draft Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law,* December 2023

Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the regions, *Artificial Intelligence for Europe,* COM/2018/237 final*,* Brussels, April 25th 2018

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and Committee of the Regions, 2030 Digital Compass: the European way for the Digital Decade, COM(2021) 118 final, March 9, 2021

Congress.gov. S.2551, 117th Congress, *AI Training Act*, Introduced July 29, 2021

Convention on the Rights of the Child, adopted 20 November 1989 by General Assembly resolution 44/25, entered into force on September 2, 1990

Corrections and Conditional Release Act (S.C. 1992, c. 20) CCRA, Assented to 1992-06-18, https://laws-lois.justice.gc.ca/eng/acts/c-44.6/

Cosa sono i Big Data e come vengono utilizzati?, in BNova, february 24, 2022. Last visited july 4, 2024. https://www.bnova.it/data-science/cosa-sono-i-big-data/

Council of Europe, Committee on Artificial Intelligence (CAI*), Revised Zero Draft [Framework] Convention on Artificial Intelligence, Human Rights, Democracy, and the Rule of Law, Strasbourg*, 6 January 2023

D. LESLIE, C. BURR, M AITKEN, J. COWLS, M. KATELL, M. BRIGGS, *Artificial Intelligence, Human Rights, Democracy and the Rule of Law, A Primer*, Council of Europe and The Alan Turing Institute, June 2021

D. MILWARD, *Sweating it Out: Facilitating Corrections and Parole in Canada Through Aboriginal Spiritual Healing,* 2011, p. 47

Data Headhunters, *Fuzzy Logic vs Probability: Handling Uncertainty in Data,* January 5, 2024. Last visited July 4, 2024. https://dataheadhunters.com/academy/fuzzy-logic-vs-probability-handling-uncertainty-in-data/

Draft Framework Convention on artificial intelligence, human rights, democracy and the rule of law, March 2024. Last visited June 4, 2024. https://rm.coe.int/-1493-10-1b-committee-on-artificial-intelligence-cai-b-draft-framework/1680aee411

E. GLOVER, *AI Bill of Rights: What You Should Know,* in Builtin, 2024. Last visited July 3, 2024. https://builtin.com/artificial-intelligence/ai-bill-of-rights

E. HILL, J. WOLFE, E*wert v. Canada: Shining Light on Corrections and Indigenous People*, in The Supreme Court Law Review: Osgoode's Annual Constitutional Cases Conference, Vol. 94, 15, 2020, p. 391-413.

E. ISIN, E. RUPPERT, *Data's empire: Postcolonial data politics*, (eds) Data Politics, 2019, pp. 207–227

E. KLEIN, S. PATRICK, *Envisioning a Global Regime Complex to Govern Artificial Intelligence*, 2024, p. 5

E. M. CAMPBELL, D. F. SITTIG, K. P. GUAPPONE, R. H. DYKSTRA, J. S. ASH, *Overdependence on technology: An unintended adverse consequence of computerized provider order entry*, in AMIA Annual Symposium Proceedings, 2007, pp. 94-98. Last visited August 27, 2024. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2710605/

E. STRADELLA, *Le fonti nel diritto comparato: analisi di scenari extraeuropei (Stati Uniti e Cina)*, in "DPCE Online", vol. 51, 2022. Last visited July 3, 2024. https://www.dpceonline.it/index.php/dpceonline/article/view/1569

E. YUDKOWSKY, *Artificial Intelligence as a Positive and Negative Factor in Global Risk*, In *Global Catastrophic Risks,* 2008, p. 1

European Commission of 2021 on *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (artificial intelligence act) and amending certain union legislative acts*, COM/2021/206 final, April 21, 2021

European Parliament of 2023 *on Artificial Intelligence Act on amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). Amendment 122, Proposal for a regulation, Recital 76

European Parliament, *Legislative Train, A Europe fit for the digital age*, August 20, 2024

Executive Order No. 13859, Federal register, vol. 84, N. 31, 11 February 2019

Foreign, Commonwealth, and Development Office, Department for Science, Innovation, and Technology; and the AI Safety Institute, *AI Safety Summit 2023*, November 2023, https://www.gov.uk/government/topical-events/ai-safety-summit-2023

G. GOMEZ, *Intelligenza artificiale, profilazione e nuove forme di discriminazione*, in Teoria e Storia del Diritto Privato, 2022, pp. 11-19.

G. KAKAR, *Cognitive dysphoria: Evaluating the paradigm shift of artificial intelligence technology in digital colonialism.* In Indian Journal of Artificial Intelligence and Law Vol 2, 2021

G. LONGO, *Big Data e Intelligenza artificiale: che futuro ci aspetta?,* S&F Scienzaefilosofia.It, no. 20*,* 2018, pp. 15ss

G. SANGUINETTI, *Machine Learning: accuratezza, interpretabilità e incertezza*. Ithaca: Viaggio nella Scienza XVI, 2020, pp 71-78

General Office of the CPC Central Committee and the General Office of the Council, *Opinions on strengthening the ethical governance of science and technology, 关于加强科技伦理治理的意见,* 2022. Last visited July 7, 2024. https://www.gov.cn/zhengce/2022-03/20/content_5680105.htm

Global Partnership on Artificial Intelligence GPAI, launched in June 2020, https://gpai.ai/projects/

Government of Canada, *Consumer Privacy Protection Act, 2023,* https://ised-isde.canada.ca/site/innovation-better-canada/en/consumer-privacy-protection-act

Government of Canada, *Directive on automated decision-making*, April 1 2019, Annex 1. Last visited May 10, 2024. https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

H. CHAMBERS, *Canada: AI and Data Act - Key takeaways*, in DataGuidance, 2023. Last visited July 14, 2024. https://www.dataguidance.com/opinion/canada-ai-and-data-act-key-takeaways

H. MARCUSE, *Negations: Essays in Critical Theory,* 1968, p. 168.

H. SHEIKH, C. PRINS, E. SCHRJIVERS, *Mission AI, The New System Technology,* Research for Policy, 2023, pp. 15-19

H. SHILLER, *Communication and Cultural Domination,* 1976

H.R. 3611, *Algorithmic Justice and Online Platform Transparency Act,* 117th Cong., 28 May 2021

H.R.6216, *National Artificial Intelligence Initiative Act of 2020*, 116th Congress.

High-Level Expert Group on Artificial Intelligence, *A definition of AI: Main capabilities and scientific disciplines*. European Commission, 2019, p. 6

House of Commons of Canada, BILL C-27, *Personal Information and Data Protection Tribunal Act*, First reading, June 16, 2022, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

House of Commons of Canada, BILL C-27, *The Artificial Intelligence and Data Act (AIDA),* First reading, June 16, 2022, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading

I. DE FEO, A. AFFERNI, *AI Act: il Regolamento sull'Intelligenza Artificiale adottato dal Parlamento UE,* 14 March 2024. Last visited May 28, 2024. https://www.dirittobancario.it/art/ai-act-il-regolamento-sullintelligenza-artificiale-adottato-dal-parlamento-ue/

I. FALEFUAFUA TAPU, T. KAMAILELAULI'I FA'AGAU, *New Age Indigenous Instrument: Artificial Intelligence & Its Potential for (De)colonialized Data*, in Harvard Civil Rights-Civil Liberties Law Review Vol. 57, 2022, pp. 739-746

I. GENNA, *The regulation of foundation models in the EU AI Act,* in International Bar Association, April 12, 2024. Last visited May 29, 2024. https://www.ibanet.org/the-regulation-of-foundation-models-in-the-eu-ai-act

I. POOL, *Colonialism's and postcolonialism's fellow traveller: the collection, use and misuse of data on indigenous peopl*e, in T. KUKUTAI, J. TAYLOR (eds), Indigenous Data Sovereignty: Toward an agenda, 2016, pp. 57-76

Interim Measures for the Administration of Generative Artificial Intelligence Services, 生成式人工智能服务管理暂行办法, 13 July 2023. Last visited July 7, 2024. https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm

International Covenant on Civil and Political Rights (ICCPR), adopted 16 December 1966 by General Assembly resolution 2200A (XXI), entered into force on March 23, 1976

International Covenant on Economic, Social and Cultural Rights (ICESCR), adopted 16 December 1966 by General Assembly resolution 2200A (XXI), entered into force on January 3, 1976

J. DE VRIES, *The industrial revolution and the industrious revolution*, in The Journal of Economic History, Vol. 54, No. 2, 1994, p. 249-270

J. FLAMING-JONES, EU Policy. *EU AI Act nearing agreement despite three key roadblocks – co-rapporteur*, Euronews.next. last visited (May 21, 2024. https://www.euronews.com/next/2023/10/23/eu-ai-act-nearing-agreement-despite-three-key-roadblocks-co-rapporteur

J. HAUGELAND, Artificial intelligence: the very idea, Cambridge (US), London, 1985, p. 2

J. MCBRIDE, A. CHATZKY, *Is 'Made in China 2025' a Threat to Global Trade?, 2019.* Last visited July 5, 2024. https://www.cfr.org/backgrounder/made-china-2025-threat-global-trade

J. MCCARTHY; *What is Artificial Intelligence*, 2007, available at: http://www-formal.stanford.edu/jmc/whatisai.pdf

J. R. SEARLE; *Minds, Brains and Programs, in The Behavioral and Brain Sciences*, Vol. 3, 1980, Cambridge University Press

J. TACHEVA, S. RAMASUBRAMANIAN, *AI Empire: Unraveling the interlocking systems of oppression in generative AI's global order,* in Sage Journals, Big Data & Society, 2023. Last visited September 4, 2024. https://journals.sagepub.com/doi/10.1177/20539517231219241#bibr58-20539517231219241

J. VIERA MAGALHAES, N. COULDRY*, Giving by taking away: Big tech, data colonialism and the reconfiguration of social good*. International Journal of Communication, 2021, Vol. 15, pp. 343–362.

Japanese Ministry of Foreign Affairs, *G7 Leaders' Statement on the Hiroshima AI Process*, October 30, 2023, https://www.mofa.go.jp/ecm/ec/page5e_000076.html

K. CRAWFORD, *Artificial Intelligence's. White Guy Problem,* in The New York Times, 2016. Last visited August 27, 2024. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

K. CRAWFORD, *The Hidden Biases in Big Data*, in Harvard Business Review, 2013. Last visited August 27, 2024. https://hbr.org/2013/04/the-hidden-biases-in-big-data

K. GOODLUC, *The Erasure of Indigenous People in U.S. COVID-19 Data,* in HIGH COUNTRY NEWS, 2020. Last visited September 8, 2024. https://www.hcn.org/articles/indigenous-affairs-the-erasure-of-indigenous-people-in-us-covid-19-data/

K. HAO, *Artificial intelligence is creating a new colonial world order,* MIT Technology Review, 2022. Last visited September 4, 2024. https://www.technologyreview.com/2022/04/19/1049592/artificial-intelligence-colonialism/

K. HAO, H. SWART, *South Africa's private surveillance machine is fueling a digital apartheid,*in MIT Technology Review, 2022. Last visited September 20, 2024. https://www.technologyreview.com/2022/04/19/1049996/south-africa-ai-surveillance-digital-apartheid/

L. ALZUBAIDI, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, in Journal of Big Data, 2021, pp. 1-4

L. BERTUZZI, *US obtains exclusion of NGOs from drafting AI treaty,* EURACTIV, 2023. Last visited June 4, 2024 https://www.euractiv.com/section/digital/news/us-obtains-exclusion-of-ngos-from-drafting-ai-treaty/

L. RINALDI, *L'intelligenza artificiale come nuova frontiera dei diritti fondamentali*, Tesi di Dottorato, Università degli studi di Trento, 2023, pp 140-142

L. VERACINI, *Containment, Elimination, Endogeneity: Settler Colonialism in the Global Present,* Rethinking Marxism, Vol. 31(1), 2019, pp. 118–140.

M. AOURAGH, P. CHAKRAVARTTY, *Infrastructures of empire: Towards a critical geopolitics of media and information studies*, in Sage Journals, Media, Culture & Society, Vol. 38(4), 2016, pp. 559–575.

M. BORGOBELLO, *AI Act: ecco come regolerà l'intelligenza artificiale generativa*, in Agenda Digitale, February 5, 2024. Last visited May 29, 2024. https://www.agendadigitale.eu/mercati-digitali/ai-act-ecco-come-regolera-i-foundation-model/

M. CASTIGLI, *5 V dei Big data, cosa sono, quale ruolo rivestono*, in BigData4Innovation, 2023. Last visited July 4, 2024. https://www.bigdata4innovation.it/big-data/5-v-dei-big-data-cosa-sono-quale-ruolo-rivestono/

M. FAZLIOGLU, *Contentious areas in the EU AI Act trialogues*, IAPP, 2023. Kast visited May 21, 2024. https://iapp.org/news/a/contentious-areas-in-the-eu-ai-act-trilogues/

M. KWET, *Digital colonialism: the evolution of American empire,* In ROAR Magazine, 2021. Last visited September 3, 2024. https://roarmag.org/essays/digital-colonialism-the-evolution-of-american-empire/

M. KWET*, Digital colonialism: US empire and the new imperialism in the global south.* In Sage Journals, Race & Class, Vol. 60(4), 2019, pp. 3–26.

M. LIM, *History of AI Winters,* 2018. Last visited May 11, 2024. https://www.actuaries.digital/2018/09/05/history-of-ai-winters/

M. MINSKY, *The Society of Mind*, The Personalist Forum, vol. 3, no. 1, 1987, pp. 19–32.

M. SOMALVICO, *L'Intelligenza artificiale*, 1987, Rusconi, Milano

M. WALTER, C. ANDERSEN, *Indigenous Statistic*: *a Quantitative Mesearch Methodology,* 2013, p. 9

MCCARTHY, *Mathematical Logic in Artificial Intelligence,* Daedalus, Vol. 117, No. 1, 1988, pp. 297-311

N. BETRO, *Xinjiang: sorveglianza speciale per gli uiguri,* in Il Caffè Geopolitico, 2021. Last visited September 21, 2024. https://ilcaffegeopolitico.net/172007/xinjiang-sorveglianza-speciale-per-gli-uiguri

N. COULDRY, U. A. MEJIAS, *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*, in Sage Journals, Television & New Media, Vol. 20(4), 2019, pp. 336–349.

N. COULDRY, U. A. MEJIAS, *The Costs of Connection: How data is Colonizing Human Life and Appropriating It for Capitalism,* 2019

N. J. NILSSON, *Introduction to Machine Learngin,* Robotics Laboratory, Department of Computer Science, Stanford University, 1998, p. 1

N. J. NILSSON, *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, 2009

N. SANGMA, *Artificial Intelligence and Indigenous Peoples' Realities*, in Cultural Survival, 2024. Last visited September 8, 2024. https://www.culturalsurvival.org/publications/cultural-survival-quarterly/artificial-intelligence-and-indigenous-peoples-realities

National Institute of Standards and Technology (NIST), U.S department of commerce, *Artificial Intelligence Risk Management Framework (AI RMF 1.0),* January 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf

OECD*, AI Principles Overview*, adopted May 22, 2019, amended May 3, 2024, OECD member countries approved a revised version of the organisation's definition of an IA system available at https://oecd.ai/en/ai-principles

OECD, *Recommendation on Artificial Intelligence*, adopted May 22 2019, amended May 3 2024, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

P. BENANTI, *Anche l'intelligenza artificiale va decolonizzata*, in Africa, 2020. Last visited September 4, 2024. https://www.africarivista.it/paolo-benanti-anche-lintelligenza-artificiale-va-decolonizzata/173072/

P. L. DI VIAGGIANO, *Etica, Robotica e Lavoro: Profili D'Informatica Giuridica*, in Revista Opinião Jurídica, Vol. 16, no. 22, 2018, pp. 247-266

P. LICATA, *Generative AI: che cos'è e quali sono le applicazioni di business dei sistemi come ChatGPT*, in Digital4, june 19, 2023. Last visited May 16, 2024. https://www.digital4.biz/marketing/generative-ai-che-cosa-e-quali-sono-le-applicazioni-di-business/

P. MORO, *Intelligenza artificiale e tecnodiritto. Fondamenti etici ed innovazione legislativa,* in P. MORO (eds), *Etica, Diritto e Tecnologia. Percorsi dell'informatica giuridica contemporanea*, 2021, pp. 7-24

P. WANG, *On defining artificial intelligence, in Journal of General Artificial Intelligence , in Journal of Artificial General Intelligence*, vol. 10, no. 2, 2019, pp. 1–37

Presidency Conclusions, *The Charter of Fundamental Rights in the context of Artificial Intelligence and Digital Change*, October 21, 2020

R. BENJAMIN, Race After Technology: Abolitionist Tools for the New Jim Code, 2019

R. CHANDRAN, *FEATURE-Indigenous groups in NZ, US fear colonisation as AI learns their languages*, in Reuters, Media & Telecom, 2023. Last visited September 9, 2024. https://www.reuters.com/article/idUSL8N2UQ0EC/

R. LOVETT, V. LEE, T. KUKUTAI, D. CORMACK, S. R. CARROLL, J. WALKER, *Good Data Practices for Indigenous Data Sovereignty and Governance,* 2018, pp 26-35

R. PANETTA*, AI Act, requisiti e obblighi per i sistemi ad alto rischio: tutto quello che c'è da sapere,* in Agenda Digitale, 2024. Last visited May 21, 2024. https://www.agendadigitale.eu/industry-4-0/ai-act-requisiti-e-obblighi-per-i-sistemi-di-ia-ad-alto-rischio-tutto-quello-che-ce-da-sapere/

R.J. BRACHMAN, *(AA)AI Presidential Address: (AA)AI More than the Sum of Its Parts*, in AI Magazine, Vol. 27, no. 4, 2006, p. 19–34.

Report of Hu Jintao to the 18th National Congress of the Communist Party of China, 8 November 2012. Last visited July 5, 2024. http://cpc.people.com.cn/n/2012/1118/c64094-19612151.html#

Research Data Alliance, International Indigenous Data Sovereignty Interest Group, September 2019, *CARE Principles for Indigenous Data Governance*, The Global Indigenous Data Alliance. Last visited September 9, 2024. https://www.gida-global.org/care

Research Data Alliance, International Indigenous Data Sovereignty Intertest Group, 2017. Last visited September 9, 2024. https://www.rd-alliance.org/groups/international-indigenous-data-sovereignty-ig

S. BAROCAS, A. D. SELBST, *Big Data's Disparate Impact,* in California Law Review, 2016, pp. 67 ss.

S. DE SPIEGALEIRE, M. MAAS, T. SWEIJS, *Artificial Intelligence And the future of defense: strategic implications for small-and medium-sized force providers. What is artificial Intelligence*, 2017, pp. 25-42

S. MEZZADRA, B. NEILSON, *On the multiple frontiers of extraction: excavating contemporary capitalism*, In Cultural Studies, Vol 31 (2-3), 2017, pp. 185-204. Last visited September 3, 2024. https://www.tandfonline.com/doi/full/10.1080/09502386.2017.1303425?scroll=top&needAccess=true

S. PERREAULT, *Violent victimization of Aboriginal people in the Canadian provinces, 2009,* Component of Statistics Canada catalogue no. 85-002-X, Juristat, 2011. Last visited September 7, 2024. https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2011001/article/11415-eng.pdf?st=niePbvho

S. R. CARROL, D. RODRIGUEZ-LONEBEAR, A MARTINEZ, *Indigenous Data Governance: Strategies from United States Native Nations, in Data Science Journal*, Vol 18 (31), 2019, p. 2

S. RUSSEL, P. NORVIG, *Artificial intelligence, A Modern Approach* (4^{a} edition.), 2020

S. RUSSEL, P. NORVIG, *Artificial Intelligence, A Modern Approach* (2^{nd} edition), 2003, pp. 16-18

S. SILVA, M. KENNEY, *Algorithms, platforms, and ethnic bias: an integrative essay*, in Phylon (1960-) Vol. 55, No. 1 & 2, SUMMER/WINTER 2018, p. 11

S.R. CARROL, I. GARBA, O. L. FIGUEROA-RODRIGUEZ, J. HOLDBROOK, R. LOVETT, S. MATERECHERA, M. PARSONS, K. RASEROKA, D. RODRIGUEZ-LONEBEAR, R. ROWE, R. SARA, J. D. WALKER, J. ANDERSON, M. HUDSON, *The CARE Principles for Indigenous Data Governance*, in Data Science Journal, Vol. 19 (43), 2020, pp. 1–12. Last visited September 9, 2024. https://datascience.codata.org/articles/10.5334/dsj-2020-043

Senado Federal, Eduardo Bismarck, *Projeto de Lei nº 21 de 2020*, https://www25.senado.leg.br/web/atividade/materias/-/materia/151547

Senado Federal, Senador Rodrigo Pacheco (PSD/MG), *Projecto de Lei n° 2338*, de 2023, https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

Senado Federal, Senador Rodrigo Pacheco, *Projecto de Lei n° 2338*, de 2023, https://www25.senado.leg.br/web/atividade/materias/-/materia/157233

Senado Federal, Senador Styvenson Valentim, *Projeto de Lei n° 5051*, de 2019, https://www25.senado.leg.br/web/atividade/materias/-/materia/138790

Senado Federal, Senador Veneziano Vital do Rêgo, *Projeto de Lei nº 872* de 2021, https://www25.senado.leg.br/web/atividade/materias/-/materia/147434

Stanford University, Human-Centered Artificial Intelligence HAI, *Artificial Intelligence IndeReport 2024,* chapter 7: Policy and Governance, p. 9

State Council Notice on the Issuance of the Next Generation Artificial Intelligence Development Plan, *A Next Generation Artificial Intelligence Development Plan*, July 2017, p. 25. Last visited July 5, 2024. https://d1y8sb8igg2f8e.cloudfront.net/documents/translation-fulltext-8.1.17.pdf,

SUPREME COURT OF CANADA (SCC)**,** *Ewert v. Canada*, 2018 SCC 30 [2018] 2 S.C.R. 165, June 13, 2018

SUPREME COURT OF CANADA (SCC), *R. v. Ipeelee*, 2012 SCC 13, [2012] 1 S.C.R. 433, March 23, 2012

T. KUKUTAI, J. TAYLOR, *Data sovereignty for indigenous peoples: current practice and future needs*, in T. KUKUTAI, J. TAYLOR (eds), *Indigenous Data Sovereignty: Toward an agenda*, 2016, pp. 1-22

T. KUKUTAI, M. WALTER, *Recognition and indigenizing official statistics: Reflections from Aotearoa New Zealand and Australia,* in Statistical Journal of the IAOS 31, 2015, pp. 317–326

T. L. MCPHAIL, *eColonialism Theory: How Trends are Changing the World,* in The World Financial Review, 2014

T. M. MITCHELL, *Machine Learning,* published by McGraw-Hill, Maidenhead, U.K., International Student Edition, 1997, p. 2

T. MADIEGA, *EU Legislation in Progress, Artificial Intelligence Act*, EPRS - European Parliamentary Research Service, 2024

The Artificial Intelligence and Data Act (AIDA) – Companion document, https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document#fn20

The Bletchley Declaration by Countries Attending the AI Safety Summit, *1–2 November 2023*, UK Government, November 1, 2023, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023

The FAIR Data Principles, in FORCE11, developed in 2014 and published in 2016. https://force11.org/info/the-fair-data-principles/

The State Council People of China, *Made in China 2025* plan issued, 2015 https://english.www.gov.cn/policies/latest_releases/2015/05/19/content_281475110703534.htm

The White House Office of Science and Technology Policy (OSTP), *Blueprint for an AI Bill of Rights: Making Automated Systems Works for the American People,* October

2022, The White House, Washington, p. 5. Last visited July 3, 2024. https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

The White House, *Fact Sheet: Vice President Harris Announces New U.S. Initiatives to Advance the Safe and Responsible Use of Artificial Intelligence,* November 2023. Last visited July 5, 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2023/11/01/fact-sheet-vice-president-harris-announces-new-u-s-initiatives-to-advance-the-safe-and-responsible-use-of-artificial-intelligence/

The White House, JOSEPH R. BIDEN JR., *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,* October 30, 2023. Last visited July 4, 2024. *https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/*

Treaty on the Functioning of the European Union TFUE, signed on December 13, 2007 and entered into force on December 1, 2009,

U. VON DER LEYEN, *A Union that strives for more, My agenda for Europe, Political Guidelines for the next European Commission 2019-2024*, European Commission, 2019.

UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2022, https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

UNESCO, United Nations Educational, Scientific and Cultural Organization, *Recommendation on the Ethics of Artificial Intelligence*, November 24 2021, p. 10

United nations Charter, signed on June 26, 1945, in San Francisco, at the conclusion of the United Nations Conference on International Organization, and entered into force on October 24 1945.

United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), adopted by the General Assembly on 13 September 2007

Universal Declaration of Human Rights, proclaimed by the United Nations General Assembly in Paris on December 10, 1948

V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor,* 2018

W. MCCULLOCH, W. PITTS, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in the Bulletin of Mathematical Biophysics, vol. 5, no. 4, 1943, pp. 115–33,

What Is Machine Learning? Definition, Types, and Examples, 2024. Last visited May 16, 2024. https://www.coursera.org/articles/what-is-machine-learning

What You Should Know: Canada's Artificial Intelligence and Data Act, in Lumenova, April 2024. Last visited July 14, 2024. https://www.lumenova.ai/blog/canada-ai-and-data-act-what-you-should-know/

WHITE PAPER, On Artificial Intelligence - A European approach to excellence and trust, February 19, 2020, COM(2020) 65 final

World Economic Forum (WEF), *Personal Data: The Emergence of a New Asset Class,* 2011, p. 5

X. CHEN, *Algorithmic proxy discrimination and its regulations,* in Computer Law & Security Review Vol. 54, September 2024

Xi Jinping chaired the ninth collective study session of the Political Bureau of the CPC Central Committee and delivered a speech, October 31, 2018. Last visited July 5, 2024. https://www.gov.cn/xinwen/2018-10/31/content_5336251.htm