LUISS

Department of Law

Course of Metodologia della Scienza Giuridica

Debiasing in the Digital World: unleashing the potential of AI to overcome the blind spots of human choices

Prof. Filiberto E. Brozzetti
SUPERVISOR

Prof. Antonio Punzi
CO-SUPERVISOR

Gianmarco Casciotta - 162343
CANDIDATE

Academic Year 2022/2023

Human beings don't like change; it scares them.

But we cannot prevent them from coming.

Either we adapt to change, or we fall behind.

Abstract

In our daily lives, each of us is surrounded by technologies and collaborates with dozens of them to simplify and facilitate our tasks. At the same time, however, technologies, especially those with "intelligent" systems, can also pose serious risks.

This dissertation aims to enable readers to become more aware of the "Digital World" in which we live and, above all, will attempt to provide a cutting-edge solution to the growing number of challenges that technology creates.

Chapter 1 begins this exploration with a historical journey from the development of the Internet to artificial intelligence ("AI"). The chapter sets the stage for the evolution from the "Gutenberg's Galaxy" to our contemporary "Zuckerberg's Galaxy." Within this Digital World, we find ourselves moving from mere citizens to "quantified selves." However, beneath the surface lie several questions and concerns. Can humans and AI coexist harmoniously, or are we on the brink of a clash? Are "Hollywood fears" about AI justified? What are the risks that technology creates?

Chapter 2 delves into the cognitive model of human decision making. It unveils the complexities of human choice, from traditional rational choice theory to insights from behavioral economics. As the rational choice model begins to crumble under the weight of Allais' paradox, scholars such as Simon introduce us to the concept of bounded rationality and lay the groundwork for the new discipline of Behavioral Economics. In fact, it is from the 1970s onward that scholars such as Kahneman and Tversky showed how the idea of the perfectly rational "Homo Oeconomicus" has turned from reality to utopia.

For these reasons, in Chapter 3, a new approach to solving the risks that new technologies create on human beings will be proposed: the use of AI as a debiasing tool. Indeed, it can induce humans to make more reasoned, informed, and rational choices in a user-friendly digital environment. This proactive approach will not only enable us to better manage the unintended consequences of technology but could also pave the way for a more sustainable and resilient future in which AI emerges as a beacon of hope in our battle against the blind spots of human choices in the face of technology risks.

To cite this Dissertation:

MLA	Casciotta, Gianmarco. "Debiasing in the digital world: unleashing the potential of AI to overcome the blind spots of human choices." (2023).
APA	Casciotta, G. (2023). Debiasing in the digital world: unleashing the potential of AI to overcome the blind spots of human choices.
ISO 690	CASCIOTTA, Gianmarco. Debiasing in the digital world: unleashing the potential of AI to overcome the blind spots of human choices. 2023.

Acknowledgements

Nella speranza che questo lavoro susciti l'interesse del lettore, desidero cogliere l'occasione per esprimere la mia sincera gratitudine a tutte le persone che hanno contribuito in modo significativo alla sua realizzazione.

In particolare, desidero rivolgere un profondo ringraziamento al mio Relatore e Maestro, il Prof. Brozzetti, per aver creduto in me, per la Sua infinita disponibilità, per il Suo sostegno, i Suoi preziosi consigli e per le conoscenze trasmesse durante tutta la stesura dell'elaborato e la cui stima per me costituisce motivo di orgoglio.

Un ringraziamento speciale va anche al mio Correlatore, il Prof. Punzi, per tutti gli insegnamenti preziosi che mi ha fornito. Con Lei ho affrontato il mio primo esame ed oggi, con il completamento di questo lavoro, ho sostenuto l'ultimo e il più importante.

Infine, desidero esprimere la mia riconoscenza a tutti i docenti e al personale della LUISS. Grazie alla vostra competenza, professionalità e disponibilità, avete contribuito in modo significativo alla mia crescita professionale e personale.

Table of contents

List of figures

Introduction

In an era characterized by the steady advancement of technology, we find ourselves at a crucial time, full of unique opportunities and significant challenges. The rise of technology and the emergence of "intelligent" systems have ushered in a new era, redefining the parameters of our daily existence. This journey, which began with the modest origins of information technology and culminated in the era of artificial intelligence (AI), has been truly remarkable.

In our daily lives, each one of us is surrounded by "intelligent" system: waking up and looking at our smartphones, working and using our PCs, sending emails, turning to Siri for a question, asking Alexa to put on a timer or our favorite song, setting Waze for a road trip, searching TripAdvisor for reviews of the restaurant where we are going to eat, posting a photo on social media, reading the news, are just some of the actions we perform daily without even realizing it and that are now part of our being digital in a new "Digital World."

The dozens of technologies we use every day have simplified our lives in many ways. Nevertheless, as we traverse this uncharted territory, it becomes increasingly important to understand the risks that await us.

In this Digital World can humans and AI coexist harmoniously, or are we on the brink of a clash? Many people including Elon Musk, Stephen Hawking and current U.S. President Joe Biden have shown concern about AI and the impact it may have on our lives. Are these "Hollywood fears" about AI justified?

Surely technology not only produces countless benefits, but also produces risks. What are these risks that technology creates? Does the human being have knowledge of them? Does he have the proper tools to cope with them?

This dissertation aims to make readers more aware of the "Digital World" in which we live and, more importantly, will attempt, in light of the considerations that will be made primarily in the second and third chapters, to provide a cutting-edge solution to the growing number of challenges that technology creates.

As we embark on this intellectual journey through the realms of technology, cognitive biases and AI-driven debiasing, we aim to uncover the nuances of the interplay between the limits of human decision-making and the risks that artificial intelligence creates.

In a world where humans are grappling with their own limitations, can machines, with their rationality, help guide us toward more informed, reasoned, and unbiased choices?

CHAPTER 1.

The advent of technology and "intelligent" systems: opportunities, risks, and implications in everyday life

Summary: **1.1** Brief history of technologies and AI development — **1.2** Welcome to the new "Digital World"— **1.3** The user as demiurge of the Digital World: from Gutenberg galaxy to the Zuckerberg galaxy — **1.4** From citizen to a "quantified self" within the "Digital World"— **1.5** Humans and AI: clash or harmony? — **1.6** Some clarification amid "Hollywood fears": is AI really intelligent? — **1.7** The risks of technology

1.1 Brief history of technologies and AI development

"*Roads? Where we're going, we don't need roads*" said Dr Emmett Brown (nicknamed "Doc") to Marty McFly in the Spielberg movie Back to the Future before they climbed into the DeLorean time machine heading for 21 October 2015. In fact, the doctor had just told Marty that he had to come with him into the future because there will be a problem with his and Jennifer's children. After they all get into the DeLorean, Marty comments that they will not have enough road to reach the 88-mph needed for time travel. Doc smiles and converts the vehicle into a flying car and speeding off into 2015.[1]

Doc Brown might have been eight off, but he was completely right. 2023 is here, and for everyday life we don't need roads: all we need is an electronic device and a steady internet connection.

Before the 1900 Paris World's Fair, some artists were asked to imagine the society of the year 2000 and draw it. Of course, these drawings are full of mechanics, gears, wheels, etc. However, no one foresaw, and indeed it was also difficult to draw, the expansion of electromagnetism, which was in its infancy at the time, electronics, and computer science. Then again, Niels Bohr, the Nobel laureate in Physics and father of the atomic model, once said, "*Prediction is very difficult, especially if it's about the future.*"[2]

---

[1] For the scene I'm talking about see here: https://www.youtube.com/watch?v=G3AfIvJBcGo.

[2] See e.g., here: https://blogs.cranfield.ac.uk/cbp/forecasting-prediction-is-very-difficult-especially-if-its-about-the-future/#:~:text=Niels%20Bohr%2C%20the%20Nobel%20laureate,model%20out%2Dof%2Dsample.

But who could have imagined all the changes and innovations that have taken place in the last 30 years?

The evolution of technology has been extraordinary and has had a significant impact on human habits.

First, the digital devices with which we work and communicate have changed: we have gone from large mainframes in the 1950s to personal computers in the 1970s and 1980s, then to laptops, tablets, and smartphones that have become an integral part of our daily lives.

On the one hand, the size of these tools has decreased; on the other, their processing capacity has increased enormously, enabling the development of more complex software and advanced algorithms to solve complex problems. Tasks that once required room-sized computers and days to complete can now be effortlessly executed by a humble laptop within mere seconds. Even more astonishing is the fact that the smartphones we casually employ today possess computing capabilities that surpass the very machines that drove Neil Armstrong to the moon in 1969.[3]

Second, the emergence of the Internet has revolutionized the way we communicate and access information. From a limited network used by academics in the 1960s, the Internet has become a global infrastructure accessible to billions of people around the world.[4]

This has paved the way for new opportunities for communication, commerce, and knowledge sharing.

Finally, Artificial Intelligence (hereinafter also "AI") is becoming increasingly pervasive in our devices and services. Machine learning algorithms enable computers to analyze large amounts of data and draw conclusions or provide predictions. This technology is used in many applications, such as speech recognition, computer vision, autonomous vehicles, and many others.

Although AI has only recently actually been developed, its concept has much older origins: the human desire to create an entity that can mimic its own behaviors has plagued humanity since the earliest times. It is certainly a utopia that has accompanied mankind in every place and age,

---

[3] Kendall, G., (2019, July 9). Apollo 11 anniversary: Could an iPhone fly me to the moon? *Independent* (blog), https://www.independent.co.uk/news/science/apollo-11-moon-landing-mobile-phones-smartphone-iphone-a8988351.html.

[4] In the 1960s, the U.S. Department of Defense developed ARPANET, a network of connected computers that was the precursor to the Internet. Initially, ARPANET was used primarily by scientists and academics to share information and research resources.

uninterruptedly and everywhere. Literature, myths, and legends are full of enigmatic characters characterized by typically human attitudes and thoughts.[5]

In antiquity, it is the myth of Prometheus, a god from the Titan family, who creates thinking and feeling clay beings without divine permission and is bitterly punished by Zeus for it.

In the Middle Ages, we find the story of the Golem, an artificial being made of clay, which is mute and not capable of reason, but possesses great strength and can carry out orders.

Literature also uses the myth of the artificially created being.[6] Perhaps the most famous example from this period is Mary Shelley's novel Frankenstein or The Modern Prometheus (1818).[7]

In the twentieth and twenty-first century, the robots live mostly in sci-fi novels such as the ones by the US-American author Philip K. Dick.[8]

In recent years, American sci-fi blockbuster films have heavily drawn on the mythological figure of the artificial human, which now appears as a robot that cooperates with humans on earth and on spaceships. Apart from these, there is also the idea of a fully digitalized world which sci-fi films and novels have taken up. The vision is almost always dystopian: there are worlds completely dominated by machines like in the film The Matrix[9] or futuristic nightmarish societies such as the one in the film Demolition Man[10], in which people act and interact based on digital instructions and even sexual contact may only take place through the mediation of digital media.[11]

Outside the literary, film and theater fields, the current landscape began to emerge from 1956, when at Dartmouth College in Hanover, New Hampshire, a group of scholars gathered with the one and only purpose of creating a machine that could simulate every aspect of human learning and intelligence.

However, the history of AI goes back long before that date, including cybernetics, the first electronic calculators and even concepts and designs developed centuries earlier. Cybernetics, born in the 1940s, studied communication and control processes in both animals and machines.

---

[5] Nida-Rümelin and Weidenfeld, 2022., pp. 1-2.

[6] *Ibid.*

[7] In this tragic story, a Swiss scientist creates an artificial human. This artificial man arouses so much disgust and fear due to his size and ugliness that he cannot connect with human society and, on the contrary, accumulates more and more rage and hatred within himself. In the end, he kills the bride of his creator and himself.

[8] Films such as Blade Runner, Minority Report or Total Recall have been made from this author's books and stories.

[9] It is a film directed by Wachowksis, USA, 1999.

[10] It is a film directed by Marco Brambilla, USA,1993.

[11] Nida-Rümelin and Weidenfeld, 2022, p. 2.

Warren McCulloch and Walter Pitts in 1943 proposed the first model of artificial neurons based on knowledge of biological neurons, propositional logic, and Alan Turing's theory of computability.[12] Cybernetics aimed to understand the mechanisms of self-regulation and control found in living organisms and feedback machines, which can adapt to the environment by modifying their behavior. Later, it was shown that any computable function could be processed by a network of connected neurons. These developments led to important achievements in AI, such as Donald Hebb's demonstration in 1949 that a simple rule for updating connections between neurons could enable learning.[13]

The roots of AI stretch far back into antiquity, but its "official" beginning was in 1956. Two young mathematicians, John McCarthy and Marvin Minsky, had persuaded Claude Shannon, already famous as the inventor of information theory, and Nathaniel Rochester, the designer of IBM's first commercial computer, to join them in organizing a summer program at Dartmouth College.[14]

The goal was stated as follows:

"The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."[15]

It was precisely during the seminar that John McCarthy coined the term "Artificial Intelligence."[16]

In the decades following the seminar, AI was characterized by a range of approaches and goals.

Some researchers focused on simulating human cognitive processes, while others aimed to achieve the best possible performance for programs, independent of imitating human processes.

In the 1980s, AI became an industry, with the development of commercial expert systems and a focus on chip design and human-machine interfaces. In addition, there was a return of the neural network approach, which had suffered a temporary decline. AI evolved further in the

---

[12] McCulloch and Pitts, 1943.

[13] Hebb, 1949.

[14] Russell, 2019.

[15] McCarthy et al., 2006.

[16] For further information see e.g., here: https://home.dartmouth.edu/about/artificial-intelligence-ai-coined-dartmouth.

following years, adopting approaches based on existing theories, supporting claims with proven theorems and experimental evidence, and focusing on well-defined real-world problems.

Today, AI focuses mainly on specific problems and the design of practical applications, including health care, automation and robotics, autonomous vehicles, finance, virtual assistance, and speech translation/recognition.[17] This had a profound impact on various aspects of society, including innovation. AI, with its ability to process vast amounts of data and perform complex tasks, has revolutionized industries, and opened new possibilities for innovation.

These machines are typically programmed to analyze data, recognize patterns, solve problems, and adapt their behavior based on new inputs or experiences, aiming to emulate human intelligence and enhance efficiency, accuracy, and productivity across various domains. AI encompasses a wide range of techniques and approaches, including machine learning, natural language processing, computer vision, expert systems, and robotics, among others, with the goal of creating "intelligent" systems that can think, learn, and act in ways that simulate or surpass human intelligence in specific contexts.

Through the application and integration of AI into various industries and sectors has the potential to revolutionize business operations, enhance productivity, and drive economic growth.

Finally, in the last years we have witnessed the development of so-called Generative AI. Unlike traditional AI, which focuses on recognizing patterns and making decisions based on existing data, generative AI goes beyond mere analysis by generating entirely new content, such as images, text, audio, or videos.[18]

In conclusion to this brief digression, one aspect is crucial to highlight: technology and AI will continue to evolve and change our lives. This development will certainly improve our lives but will also bring risks, so it will be up to our generation to ensure responsible and ethical use of them.

---

[17] Jordan, 2022.

[18] For further information see e.g., here: https://en.wikipedia.org/wiki/Generative_artificial_intelligence or here: https://research.ibm.com/blog/what-is-generative-AI.

1.2 Welcome to the new "Digital World"

Knowledge, photographs and videos, our emails and what we tell on the Internet, but also our clicks, our conversations, our purchases, our bodies, our finances, our sleep thanks to the technological evolution are becoming digital data.[19]

In the interconnected and to a large extent virtual world, data is of great importance and value.[20]

This evolution has transformed individuals from mere citizens to active users of a "Digital World" in which the only thing that matters is the data we produce. Today human beings interact, work, live in the Digital World.

This transformation would not have been possible without the birth of the Internet, big data, and algorithms. Before addressing these factors, however, it is necessary to understand what is meant by "data" and why they become "digital data."

The term "data" refers to a single unit of information or a numerical or nonnumerical value, which may be represented in textual or numerical form. Today in the digital age in which we live, it is essential to convert data into digital data because most technologies and daily activities are based on processing digital data.[21] This conversion is accomplished by translating the information into binary form, so that it can be processed, stored, and transmitted by electronic devices such as computers and other digital systems.[22] In fact, data can be expressed in various forms, such as text, images, sound, video, etc. However, to allow computer to manipulate this information, data must be converted into a form understandable by the binary system used by computers.[23]

In 1995, Nicholas Negroponte in his book "Being Digital", identified digital data and digitization as what would delineate the line between two incommensurable cultural eras.[24]

---

[19] White, 2017.

[20] Norta et al., 2016, p. 19.

[21] See, for example, the definition of "digitalization" given by the Oxford online dictionary here: https://www.oxfordlearnersdictionaries.com/definition/english/digitization.

[22] The term "binary form" refers to a system of communication or representation that uses only two symbols or states to convey information. These symbols are typically represented by the numbers 0s and 1s. The binary language is the foundation of all digital computing systems.

[23] The binary system is a number system like the decimal system: it allows any calculation like the decimal system. Every number in the decimal system corresponds to a number in the binary system.

[24] Negroponte, 1995.

The translation or description of anything into binary language and the representation of elementary thought operations as algebraic operations enable the reunification of language, logic, and mathematics, of computation, description, and knowledge.[25]

The transformation of every information into a set of digital data has only gained full meaning in the present day, with the birth of the Internet in the early 1990s, in which everything represented or contained is digitized.[26] Indeed, the Internet has profoundly influenced our technological culture, changing the way we communicate, learn, work, and do business.

One of the main effects is that every description or representation of the world thus becomes digital and no longer analog. The 1990s marked the boundary between two eras: the "analog era" and the new "digital era." Those born in the latter are referred to as the "digital natives" or "digital generation."[27] While individuals born in the analog era may be called at best "digital immigrants" because they did not grow up with computers, the Internet, or other digital devices as an integral part of their upbringing.[28]

Another effect of the advent of the Internet, thanks mostly to the enormous technological development we have witnessed over the past 30 years, is that it has completely changed our lives by making us increasingly dependent on technology and Internet. Every day we use the Internet to shop for groceries, to buy clothes, to calculate calories of a food, to check the weather, to catch a cab or bus, to call and text, etc.

This increase in internet use has been highlighted by GWI.[29]

Their research reveals that the "typical" global internet user now spends almost 7 hours per day – 6 hours and 58 minutes to be precise – using the internet across all devices.

For context, if we assume that the average person spends roughly 7 to 8 hours per day sleeping, the typical internet user now spends more than 40 percent of their waking life online.

---

[25] The binary system is also the basis of two-valued truth logic: true and false or its algebraic Boolean translation. The binary system is thus also the basis of the laws of thought represented by Boole in his algebra. For example, the operations of conjunction, disjunction, or negation, which universally characterize human thought, can be expressed by logical relations between symbols or elements in which 1 and 0 represent the truth values of the relation.

[26] Romeo, 2012.

[27] Selwyn, 2009.

[28] Prensky, 2005.

[29] GWI is the leading audience targeting company for the global marketing industry (for further information see here: https://www.gwi.com/book-demo?utm_source=kepios&utm_medium=referral&utm_campaign=2021+Kepios+Global+Audiences). Their research is used by DATAREPORTAL for his report (see here: https://datareportal.com/reports/digital-2022-time-spent-with-connected-tech).

The amount of time we spend online continues to climb too, with the daily average increasing by 4 minutes per day (+1.0 percent) over the past year.

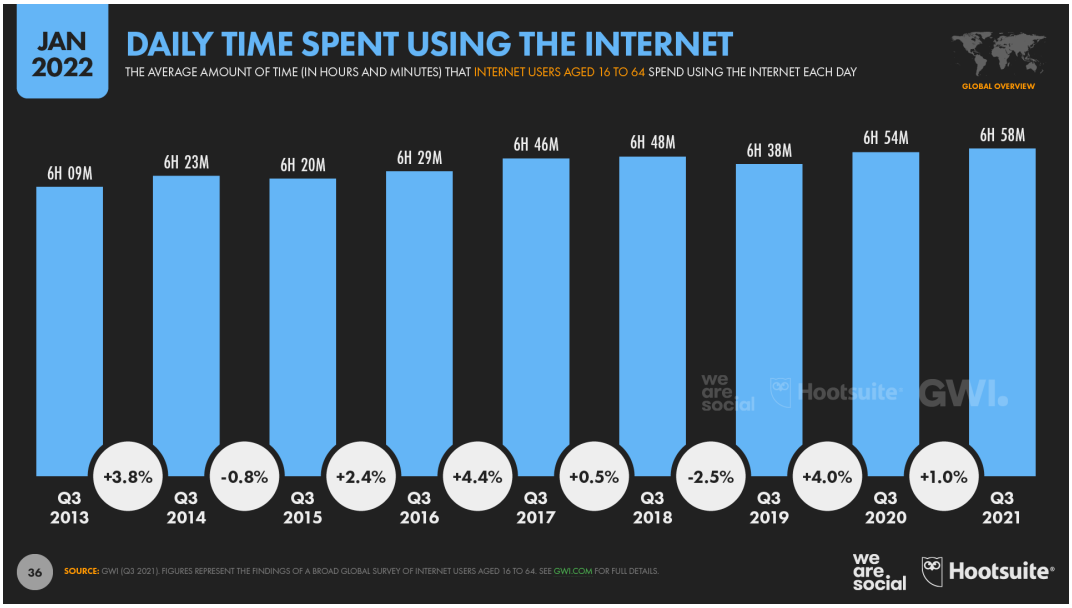

Figure 1. Global overview on daily time spent by people using the Internet

That may not sound like a big increase but added up across all the world's internet users, those 4 extra minutes per day should equate to more than 5 billion additional days of internet use in 2022. In total, the latest numbers suggest that the world should spend more than 12½ trillion hours online in 2022 alone.[30]

This data has grown and will continue to grow partly because there will be more and more people of the "digital generation" and so there will also be more devices through which to connect to the Internet.

[30] Obviously, there are considerable differences in behaviors by geography.

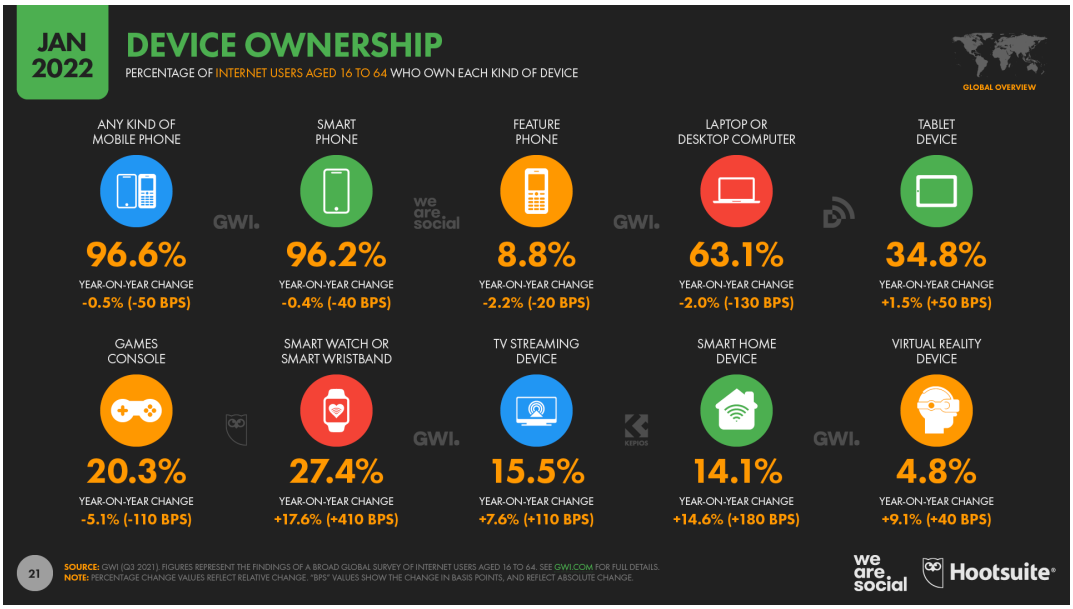

Figure 2. Global overview on device ownership

This massive use of Internet and new devices brings with it several consequences. The main one of these is that we create huge amounts of digital data every day. Every event, communication, click, and data input are evaluated, aggregated, estimated, and classified by large service and content providers, data brokers and those who sell products online. These entities use a wide range of technological solutions to track what we do and what we want.[31]

All this information and all our interactions with digital technologies, both intentional and unintentional, contribute to the creation of large volumes of data that constitute so-called "big data."[32] Big data has contributed to the computerization process of society, which feeds gigantic databases of information that had never been recorded, made accessible and easily manipulated. They are processed by data analysis and prediction algorithms that consider every available piece of information in order to conduct quantitative surveys of individuals, users, and groups to understand their needs and predict their future behavior.[33]

From the behaviors consciously or unconsciously displayed online, from the history of searches, of previous purchases of an individual, but also from the orientations and conduct of other and

[31] McCune, 1998.

[32] Big data refers to extremely large and complex sets of data that cannot be effectively processed or analyzed using traditional data processing applications and tools. The term "big data" is characterized by three main attributes (known as the three V's): volume, velocity, and variety. In addition to the three V's, big data is often associated with a fourth V: veracity.

[33] White, 2017.

distinct individuals, which show, however, similar, or consistent, or emulative profiles with respect to the target considered, producers and sellers are able to predict the tastes, preferences and even the needs of individual consumers, regardless of their explicit manifestation of will. This means that it is possible to intercept demands that the market has not yet even had a chance to generate or develop, to predict an individual's desires and needs for products or options before he or she even knows about them.[34]

Today, a flood of data pours onto the Internet. Every day 3.3 billion queries are made on the 30 trillion pages indexed by Google; on Facebook, more than 350 million photos and 4.5 billion likes are distributed; 3 billion Internet users exchange 144 billion e-mails. If you computerized all communications and writings from the dawn of humanity until 2003, it would take 5 billion gigabits to put them in memory. Today, we generate this volume of information in just two days.[35]

Nevertheless, data, in its raw form, holds limited value because they are an immense collection of numbers, amounting to billions and billions, lacking significance without the crucial element of interpretation within a specific context. For this reason, algorithms give computers mathematical instructions by which to sort, process, aggregate, and represent information. An algorithm means a succession of instructions that define the operations to be performed on the raw data to obtain certain results.[36] Whenever we solve a problem or perform a task, we actually carry out an algorithm. The word "algorithm" is over 1,000 years old,[37] but until just over 50 years ago, algorithms were the exclusive subject of mathematicians and engineers. Since the algorithms have become software programs and are executed by computers, they have assumed a special role in many areas and moments of our life.[38]

The advent of Internet, technology and devices development, abundant data storage, the huge production of big data, the use of algorithms, better processing capacity and AI on the one hand have made our world "digital", on the other hand have made us users of a "data-driven world." Humans are captivated by the power of data. We now have access to vast amounts of information, right at our fingertips, and its allure is undeniable.

---

[34] Brozzetti, 2019.

[35] Jordan, 2022.

[36] Oxford Reference. (n.d.). Algorithm. In *Oxford Reference online*. Last accessed July 18, 2023, https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095402315;jsessionid=3965554A8D650F7D570A6A1F555CDCC2.

[37] It derives from the 9th Century mathematician Muhammad ibn Mūsāal'Khwārizmī, latinized 'Algoritmi'.

[38] Talia, 2019.

1.3 The user as demiurge of the Digital World: from Gutenberg galaxy to the Zuckerberg galaxy

The massive use of Internet, as highlighted in the previous section, also caused an anthropological transformation.

McLuhan's "Gutenberg man,"[39] Sartori's "homo videns,"[40] "homo electronicus" in short has become, maybe without realizing it, "homo digitalis" himself mesh of the network, ganglion of the fabric, junction of information.[41] This concept refers to the profound transformation that humanity is experiencing as we increasingly engage with and rely on digital technologies.

Homo electronicus refers to the previous stage of human development, characterized by our adaptation to electronic devices and technologies. This stage represents the period when electronic devices, such as televisions, radios, and computers, became an integral part of our lives. It highlights our ability to utilize and interact with these technologies, shaping our behaviors and communication patterns.

Conversely, the term *homo digitalis* represents the next phase of human evolution, where digital technologies, particularly the internet and mobile devices, have become ubiquitous and deeply intertwined with our daily lives. In the last 30 years the advent of the Internet and the advancements in technology have led also to a significant shift in how *homines digitales* interact with each other, access information, and carry out various activities. Nowadays the *homo digitalis* depends on digital platforms, social media, online communities, and the vast array of digital tools and services available to us. This transformation is seen as an evolutionary shift, changing the way we perceive, interact, and exist in the world. It also reflects how our cognitive abilities, social structures, and individual identities are being reshaped by the digital realm.

Many authors describe this transformation as the transition from the "Gutenberg Galaxy" to the "Zuckerberg Galaxy."[42]

The Gutenberg Galaxy is a concept introduced by Marshall McLuhan, a Canadian philosopher and communication theorist, in his book "The Gutenberg Galaxy: The Making of Typographic Man." Published in 1962, McLuhan argued that the invention of the printing press by Johannes

---

[39] McLuhan, 1962.

[40] Sartori, 2014.

[41] Brozzetti, 2019.

[42] For example, 30 years ago, no one would have thought of a cell phone without a keyboard, video calling, voice messaging, voice assistants (such as Siri), smart working, streaming platforms, and all the other innovations we cannot do without today.

Gutenberg in the 15th century and the subsequent proliferation of print culture had profound effects on human society, cognition, and communication. The printing press, according to the author, transformed the way people perceive and understand the world. Before the advent of the printing press, knowledge was primarily transmitted orally or through handwritten manuscripts. The printing press revolutionized this by enabling the mass production of books and making them widely available. As result of this, information became more standardized and accessible, leading to increased specialization and compartmentalization of knowledge. Overall, the printing press played a crucial role in the formation of modern societies (especially regard culture, cognition, and social structures).

The "Zuckerberg galaxy" is a playful term combining the names of Mark Zuckerberg, the co-founder of Facebook, and the concept just analyzed of the "Gutenberg Galaxy." The term is used to describe the digital and social media-dominated landscape of the modern era. Social media, such as Facebook, Instagram, X app (or Twitter for the more nostalgic), Google, YouTube, Netflix, Amazon, and others, have become integral parts of people's lives, shaping the way they communicate, interact, and access information.

These platforms provide a sense of belonging, allow for the formation of online relationships, and facilitate the exchange of ideas on a global scale. Information spreads rapidly, widely and they can reach a vast audience within seconds. At an average of 2 hours and 27 minutes per day, social media accounts for the largest single share of our connected media time, at 35 percent of the total.

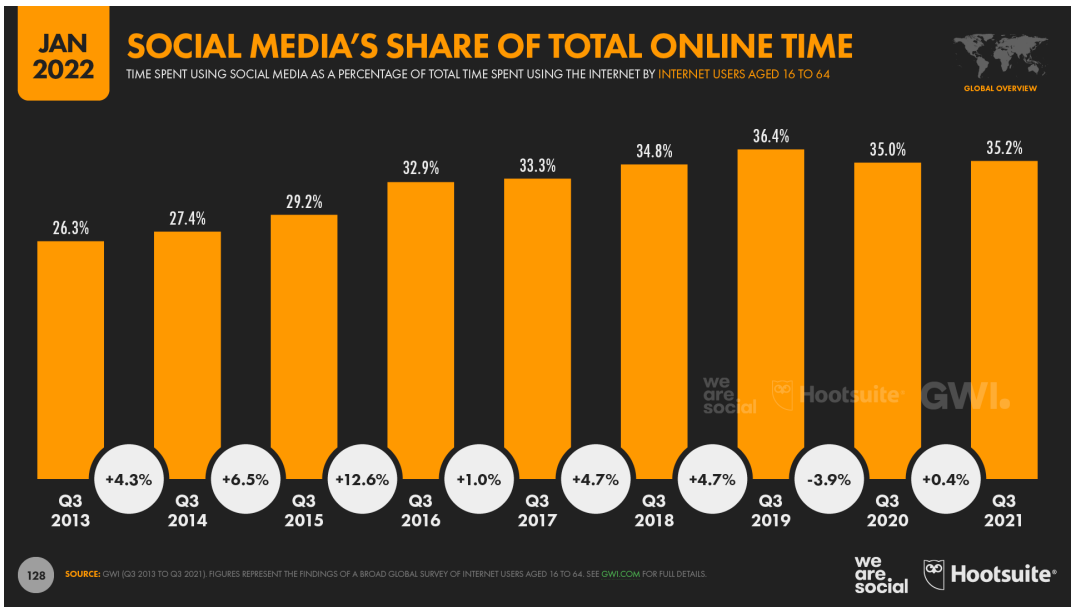

Figure 3. Global overview on social media's share of total online time

One of the main problems with this revolution was pointed out by Castells, in a book in his famous triptych,[43] and concerns the fact that this latest revolution – the Zuckerberg Galaxy – developed while the first – the Gutenberg Galaxy – was still in progress.

From electronic communication, which informs distant subjects from one another, integrated electronic communication has already budded: the network, which puts directly connected individuals in communication with one another.[44]

This has created a radical change in the ways in which an individual participates in this "galaxy." Before, mass media transmitted communications and information to a mass audience, which had a mere receptive role. An individual was the classic product of mass society within which he or she had no defined entity but were simply an indistinguishable individual among the masses.

Today, in contrast, the user of the "Digital World" can actively participate in creating the "galaxy" – the system – by adding his or her own knowledge and they become sources and vectors of information to others. The user, the *homo digitalis,* like the Platonic demiurge, creates the Digital World through the data he or she voluntarily or involuntarily produces. The global village of printing and telecommunications, of information technology that was becoming telematic, however recognizable within determined boundaries is transfigured in the swarm of social networks, in the cluster formation of virtual communities as active as they are ephemeral. It is no longer a space of virtual rooms, databases, places in any case institutional of the new mass information, but it is a space of flows.[45]

An example of this new approach can be seen in the Google's PageRank algorithm aimed at discovering the quality of information. Before Google, the early search engines (e.g., Lycos and Alta Vista) gave better rankings to sites whose pages contained the most times the keyword requested by the user. Google's PageRank algorithm opposed this practice with another strategy: hierarchy to information is assigned based on the links a site receives from another site. So, the most visible information is not the most viewed, but rather the information to which active Internet users have chosen to give recognition by referring back to it via many links.[46] The quality of a piece of information is thus assured by the most active Internet users, who contribute in this way to creating the system for themselves.

---

[43] Castells, M. (1996). The Rise of the Network Society, The Information Age: Economy, Society and Culture, Vol. I. Wiley.

[44] Bentivegna, S., & Artieri, G. B. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale*. Gius. Laterza & Figli Spa.

[45] Castells, 1996.

[46] In the digital universe, this principle has taken the name "collective intelligence" or "wisdom of crowds."

1.4 From citizen to a "quantified self" within the Digital World

Traditionally, a citizen is a person who, being part of a particular country, enjoys certain rights and responsibilities. A person, if he or she meets the requirements established by law, can obtain citizenship of the country. The concept of citizenship in recent centuries has been of fundamental importance to the individual. By giving citizens a set of rights and freedoms, citizenship guarantees them dignity and equality within a society. Nevertheless, modern society also knows that people start from different situations and face different obstacles during their lives. For this reason, it is the task of the modern state on the one hand to ensure that all citizens are treated equally, and on the other hand to recognize diversity in order to take measures to overcome these inequalities and provide effective opportunities for all, so that every individual has a chance to realize himself or herself.[47]

After past generations struggled to achieve these principles, they are now being swept away by the Digital World.[48]

As I mentioned in the paragraph 1.2, data must be converted into a form understandable by the binary system used by computers. Through this act of "translation", the complexities of our information, our real experiences and our diversities are simplified into simple numbers, leaving out all the nuances that cannot be encapsulated in this binary structure. So, in the Digital World, our existence and life are simplified into a binary world made of 0s and 1s. This process creates what has been called the "quantified self."[49]

The tendency of the *homo digitalis*, thanks to the huge technological development, is to quantify certain aspects of his life to try to improve or simply to keep himself in check. For example, many of us keep track of our body weight, caloric intake, steps walked, or physical activity performed. We are gradually viewing ourselves as mere machines to be tracked and calibrated, neglecting our intricate human nature filled with hopes, dreams, and passions.

This shift in perspective can significantly impact how we perceive success and the pursuit of a fulfilling life.

---

[47] This principle in Italy is expressed in Article 3 of the Constitution and is referred to as the "principle of substantive equality." However, the same principle is also found in the Constitutional Charters of other countries (e.g., Spain and South Africa).

[48] The modern concept of citizenship began to develop from the 17th century, with the emergence of national sovereignty and the change of political and social structures during the Modern Age. Some of the key turning points in the process of defining modern citizenship were the Treaty of Westphalia in 1648, which ended the Thirty Years' War and introduced the principle of territorial sovereignty; the American Revolution; the French Revolution; The Declaration of the Rights of Man and of the Citizen, adopted during the French Revolution in 1789; and World War I and World War II.

[49] Han, 2014. You can see also Gary Wolf TED talk here: https://www.ted.com/talks/gary_wolf_the_quantified_self.

Moreover, the reduction of the individual to data makes it even easier to ignore the unique nature, value, and dignity of the individual, and simply lump them into an aggregate that can then be used to achieve some larger goal by those who control and can dispose of our data.[50] In general, this leads us to think of people as interchangeable, which is a step on the road to a utilitarian mindset that demeans the individual in favor of the collective and denies the essential rights we all value as human beings.

The tendency to quantify certain aspects of our life are not themselves wrong or ill-advised. In fact, it is right to care about one's health and make the most of the potential of technological innovations in this field.

Instead, the danger is in letting the results of quantification guide our decision-making to an excessive degree and nudging ourselves into focusing too much on those aspects of our lives that are easily put into numbers and not enough on those that cannot. Indeed, by quantifying various aspects of our lives, thanks to the wealth of data and analytics, we often fail to grasp the underlying nuances and complexities of our choices. Instead of seeking understanding, we resort to mere numerical measurements and records, devoid of context. So, it often happens that the user of the "Digital World" tends to forget about aspects that cannot be converted into numerical-quantifiable format. This is referred to as modern cognitive bias.[51]

This risk has also increased due to the use of AI.

Blind reliance on the use of these systems in various areas, such as recommending products online or selecting job openings, affects our ability to make autonomous and responsible decisions.

On the other hand, these practices are fine and recommended if this information are incorporated into a decision-making process that includes other less quantifiable aspects of health, such as fatigue, aches, and pains, and simply how we feel on a day-to-day basis, as well as life in general.

In this case, however, a question immediately arises, which I will try to answer in the third chapter of this paper: are we sure that this kind of shared decision-making is actually "shared"? Can we manage not to be influenced by "intelligent" systems in our lives?

---

[50] Mayer-Schönberger, and Cukier, 2013.

[51] White, 2017.

1.5 Humans and AI: clash or harmony?

Humans or Artificial Intelligence?

The question posed in these terms, for some might the Hamletic doubt between being or not being; for others, the idea of a comic battle between two archenemies.

In reality, this relationship is anything but that. But, before delving into this relationship, it is necessary to understand better what AI is.

AI refers to all efforts to recreate some degree of human reasoning through non-human elements or devices.

Over the course of history, humans have attempted to recreate artificial life for millennia. The ultimate goal has always been to replicate, improve or even surpass human characteristics.

Robert Geraci, a professor of religious studies, tries to explain this trend based on the story of Adam and Eve. The myth states that humanity, from that point on, has lived in a state of falling from grace, and for this reason Geraci interprets the quest for artificial life as an attempt to escape that imperfection.[52]

John Jordan, on the other hand, prefers a more secular interpretation and believes that these attempts stem from the importance that the concept of "frontier" occupies in Western (especially American) culture.[53] Man, by nature, always tends to cross a physical frontier to seek innovation.[54]

Like a kind of Heraclitan tension, human beings constantly aspire to a greater degree of complexity and diversity. This tension is reflected in the principle of evolution: the more unattainable a thing seems to man, the more the desire to possess it grows in him, leading him to despair and inner torment.

The desire for advancement in the technological sphere is therefore based on the search for innovative solutions and the constant challenge to overcome current limitations. Humans set ambitious goals, seeking to overcome difficulties and make significant improvements in society. For this reason, since the earliest times, humans have been constantly engaged in improving existing technologies and seek new ones, and AI is one of them.

---

[52] Geraci, 2010, p. 31.

[53] Jordan, 2022, pp. 36-37.

[54] Jordan in his book cites as examples the fights in America against indigenous peoples for the conquest of western territories and the space race during the Cold War.

Baudrillard argues, "*technology is an extension of the body. It is the functional sophistication [the "artifice"] of a human organism that enables it to match nature and invest it triumphantly.*"[55]

Paraphrasing this definition, I therefore believe that a technology is such when it has three different characteristics:

(i) the application
(ii) of scientific knowledge
(iii) for practical purposes.

Thus, technology is something that can be used for solving practical problems, optimizing procedures, making decisions, and choosing strategies aimed at certain objectives.[56]

It is appropriate to dwell briefly on the third element of the definition above: the ultimate practical purpose.

The desire to seek new solutions and improve what already exists is not for an ephemeral and abstract purpose but is aimed at finding practical solutions to practical problems. It is in fact this element that has driven and drives technological innovation and progress in society.

AI falls under the above definition of "technology" because it is a tool that is applied for a practical purpose: enabling or facilitating a certain operation.

So, going back to the question at the beginning of this section, how does AI relate to humans? Is it a clash that must end with a winner and a loser?

In light of the observations above, the answer is no.

In fact, we live in a phase of history in which we take for granted the presence of complex machines that work for us and with us, and often we do not even realize how much technology is present in our daily lives and how it affects them.

Technology has been a part of our lives since ancient times, so much so that to separate humans from it would be a mistake. From the use of ancient prehistoric tools to the advent of writing, from the mechanization of printing to the digitization of information, from the invention of the wheel to airplanes, the long history of conceptual technologies has been at the heart of humanity's evolution.

---

[55] Baudrillard, 2010.

[56] Treccani. (n.d.). Tecnology. In *Vocabolario Treccani online.* Last accessed July 8, 2023, https://www.treccani.it/vocabolario/tecnologia/.

Such technologies have profoundly transformed who we are, what we know, our ways of thinking, and our representations of ourselves. For these reasons it would be naïve to think that AI is our enemy or something that can replace us.

For AI, too, as with all new technologies throughout history, it will be up to us to learn how to use it, live with it and exploit its potential.

Already some years ago John Kelly, former director of research at IBM, said: "*The goal is not to replicate the human brain [...]. It is not about replacing human thinking with machine thinking. Rather, in the age of cognitive systems, it is about having humans and machines work together to produce better results, each contributing their superior skills to this partnership.*"[57]

[57] Cited by Carlo Ratti in: Ratti, C. (2015). *Gli innovatori*. Aspenia, 68, pp. 44-49.

1.6 Some clarification amid "Hollywood fears": is AI really intelligent?

Many people remain concerned about AI and the impact it can have in our lives.

For example, in 2014 Elon Musk in an interview said:

"*I think we should be very careful about artificial intelligence. If I were to guess like what our biggest existential threat is, it's probably that. So we need to be very careful with the artificial intelligence. Increasingly scientists think there should be some regulatory oversight maybe at the national and international level, just to make sure that we don't do something very foolish. With artificial intelligence we are summoning the demon. In all those stories where there's the guy with the pentagram and the holy water, it's like yeah he's sure he can control the demon. Didn't work out.*"[58]

Also, Professor Stephen Hawking has warned that the creation of powerful artificial intelligence will be "*either the best, or the worst thing, ever to happen to humanity.*"[59]

Not only Musk and Professor Hawking, but also U.S. President Joe Biden also recently said: "*Artificial intelligence promises risks to our society, our economy and our national security, but also incredible opportunities.*"[60]

Leaving aside Elon Musk's wizards and spirits, it is appropriate to dwell briefly on these statements because many other people, like them, fear that AI will cause risks and may surpass humans in ability.

So, is all this alarmism justified?

Certainly, AI like all new technologies when they come to the public has created concerns and fears. Its potential is certainly risky so much so that the independent high-level expert group on artificial intelligence, set up by the European Commission,[61] has defined AI as a "disruptive

---

[58] McFarland, M., (2014, October 24). Elon Musk: 'With artificial intelligence we are summoning the demon'. *The Washington Post* (blog), https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/.

[59] Professor Stephen Hawking on several occasions expressed some concerns about AI systems. Among the various interviews, see for example: Hern, A., (2016, October 19). Stephen Hawking: AI will be 'either best or worst thing' for humanity. *The Guardian* (blog), https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge.

[60] "Remarks by President Biden on Artificial Intelligence" (available here: https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/07/21/remarks-by-president-biden-on-artificial-intelligence/).

[61] "Ethics guidelines for trustworthy AI" (available here: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai).

technology" because it can radically change not only established technologies, but also the rules and business models of a given market, and often business and society in general.[62]

This qualification while absolutely correct, in my opinion still does not justify all the fears. These I believe can be attributed mainly to two causes: the first is that AI systems are one of the most pioneering innovations of the past half-century, and because of their rapid development and use, the legislature has not yet figured out the best regulatory approach. The second comes from the error that is often made to attribute an identity to AI, recognizing it as an intelligent entity.

1.6.1 How to deal with the rapid development of AI?

Normally, any new technology that is born and later enters the world, goes through a whole series of intermediate steps.[63]

The first and most complex phase is the "development phase" and involves mainly computer scientists, programmers, and engineers. After figuring out "How can the new technology work?" the second phase of "How can you make money with it?" starts and involves mainly investors. Finally, before technology enters the public eye and becomes part of our lives, it is necessary to interact with other people who do not appear as engineers, scientists, or investors: lawmakers.

The new technology must comply with a country's existing legislative framework in order to be used in accordance with predetermined principles.

With the advent of AI, we have skipped this last step. The legislature unfortunately has arrived unprepared and confused about the Digital World.

To get an idea of this, one only has to listen to the questions put to TikTok CEO Shou Zi Chew during his hearing before the U.S. Congress: many of them were so vague, speculative, and irrelevant that one doubted the competence of those asking them.[64] The same situation occurred in 2018, during Mark Zuckerberg's hearings before the U.S. Congress, and then in the European Parliament.[65]

---

[62] Cambridge Dictionary. (n.d.). Disruptive technology. In *Cambridge Dictionary online*. Last accessed July 18, 2023, https://dictionary.cambridge.org/dictionary/english/disruptive-technology.

[63] Jordan, 2022, pp. 15-16.

[64] You can find the hearing here: https://www.youtube.com/watch?v=_E-4jtTFsO4.

[65] For U.S. Congress you can find the hearing here see here https://www.youtube.com/watch?v=u-FlWZ1BOcA and for EU Parliament here: https://www.youtube.com/watch?v=bVoE_rb5g5k.

These two examples demonstrate the gulf separating the technological expertise of lawmakers from that of Big Tech managers and, more importantly, the former's lack of understanding of the digital phenomenon.

This poor preparation and understanding of the phenomenon, especially when it concerns the regulation of new technologies, is very dangerous for mainly two reasons.

The first one is that the market moves too fast.

Consider for example Threads, Zuckerberg's new app, that has reached 100 million monthly active users in just five days after launch.[66] The same milestone had been reached by ChatGPT[67] in two months, TitTok in nine months, and Instagram in two and a half years.[68] As can be seen, the longer the years go by, the more a technology reaches users in less time.

The second one is that new technologies use our personal data to offer us a certain service and this constitutes a danger when the users don't fully understand it.

It is certainly true that this rapid development and use of technology can bring risks,[69] but it is also true that attempts to curb a new technological frontier – especially when an innovation reaches a large segment of the population in a short time – are inconclusive.

This is why I believe that a possible solution is not to be afraid of new technologies and restrain them, but rather to establish uniform rules among different states. This should be done with an approach in which pragmatism prevails over ideology. So, before we ask ourselves whether and how we "would like" or "would not like" to govern innovation, we should first ask ourselves whether, by placing "a reef" on it, we will succeed in "stemming the sea."

In other words, we should consider with extreme caution the adoption of rules that would end up serving little or no purpose except to curb innovation itself.

---

[66] Threads is an American social media platform and social networking service owned and operated by Meta Platforms (for more information see here: https://en.wikipedia.org/wiki/Threads_(social_network)).

[67] ChatGPT is an AI chatbot developed by OpenAI and launched on November 30, 2022 (for more information see here: https://en.wikipedia.org/wiki/ChatGPT).

[68] Duarte, F., (2023, July 13). Number of ChatGPT Users (2023). *Exploding Topics* (blog), https://explodingtopics.com/blog/chatgpt-users#.

[69] See *Section 1.7 et seq.*

1.6.2 AI systems have an intelligence and thanks to this will overcome human abilities

Intelligence is such a hard word to describe.

Intelligence could be defined as "the faculty, peculiar to the human mind, to understand, think, and make judgments and solutions based on the data of even intellectual experience."[70]

According to this definition, the processes, and various faculties to which we (consciously or unconsciously) refer whenever we utter the word "intelligence" are considered attributes peculiar and exclusive to human beings. These capabilities would be as typical for humans as they are atypical for other living beings, such as animals and plants (and *a fortiori* for inanimate objects).

Today, some on the contrary criticize this anthropocentric view typical of the Middle Ages and, even more so, of Humanism and Renaissance, and believe that thanks to enormous technological developments another kind of intelligence exists: artificial intelligence of the machines.

Humans are imbued with the anthropomorphic idea that calculating machines are intelligent and that their creators have succeeded in infusing a spirit into their mechanisms. This idea has been mainly fueled by the film industry[71] and theatre[72]. However, if you get out of science fiction, in research laboratories no one really believes that AI (let alone algorithms) has this kind of intelligence.

It is indeed a mistake to attribute an identity to AI, recognizing it as an intelligent entity.

Marvin Minsky, one of the founding fathers of AI, explains this trend by stating that: if one thoroughly understands a machine or a program, he finds no urge to attribute "volition" to it. If one does not understand it so well, he must supply an incomplete model for explanation.[73]

---

[70] Garzanti Dictionary. (n.d.). Intelligence. In *Garzanti Dictionary online*. Last accessed August 4, 2023, https://www.garzantilinguistica.it/ricerca/?q=intelligenza.

[71] Television and cinema have made many portentous stories about AI and technology very famous (think of the impact that movies such as Star Wars, Terminator, 2001: A Space Odyssey, Robocop, Blade Runner, and Wall-E of Disney have had).

[72] Theater, while having a minor impact, also played a role in this story. R.U.R., for example, was a play that debuted in Prague in 1921 and introduced the word "robot" to the world to criticize "mechanization" in industries and the ways in which it can dehumanize people.

[73] Minsky, 1965.

I believe that there are mainly two reasons why AI cannot be considered intelligent. The first is explained by the very definition of AI and the second by the "problem of the context" or adaptability.

AI is not in itself intelligent, but it is a machine that has a behavior associated with the idea of intelligence. As Floridi and Cabitza stated,[74] AI is not an autonomous agent with its own identity, but it is a mere machine because it is a tool that makes something possible when it would not be for our mind or body.[75]

So, AI is a technology in the broadest sense: it is a new form of ability to act and not a new form of intelligence and, like other tools, we use it to extend our capabilities. AI can affect the environment around us, but it cannot deal with it on its own.

Therefore, AI, being a technology, cannot be considered as a real entity with its own identity, consciousness, and autonomy, but as a simple form of automation that reads certain inputs and generates a certain output helping the agent who uses it.

This meaning of AI can also be interpreted by the definition given in art. 3 of the AI Act, according to which AI is a system that, based on human assigned tasks (inputs), generates results (outputs) that affect the environment with which the system interacts (therefore also users).[76]

AI cannot be considered intelligent also because the original project of AI, as machines mimicking the operations that characterize our intelligence, crashed many years ago on the problem of context.[77]

It is true that a computer developed by IBM, Deep Blue, defeated Russian chess champion Garry Kasparov in 1997, but only because it focused all its immense machinery and several megawatts on a micro universe: the chessboard, the thirty-two pieces and the rules of the game. Indeed, as complex as the game of chess may be, this represents but a tiny fraction of the environments in which a human brain operates flexibly (consuming only a few watts). Kasparov's brain, which also lost the game, can also understand a poem or a joke, catch an ironic

---

[74] Floridi and Cabitza, 2021.

[75] Already Aristotle in the IV century BC, had defined the machine as: "everything that allows us to produce an effect beyond our natural abilities through technique and to our benefit and: more we say that it is in the car with the part of our ability and allows us to overcome difficulties".

[76] 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 (*the text is available here*: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206).

'Artificial intelligence system' (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

[77] Cardon, 2018.

nuance, and translate from Russian to English and vice versa. Deep Blue could do none of these and its intelligence was focused only on one specific context. Therefore, calling Deep Blue (or any other machine) intelligent is nonsense if it cannot adapt its reasoning to all situations and contexts.

Today, the new AI goes beyond contexts, indeed exploits them to the fullest: big data represent huge collections of contexts. "Intelligent" machine translators do not translate per se but make a statistical estimate of the best possible translation, comparing it with all other translations in memory. To learn, the computer needs to absorb as much text and related translations as possible in the languages under consideration. The machine, therefore, does not reason abstractly and understand the meaning of what it is doing, but, based on the vast amount of data provided by the thousands of contexts, it can estimate the statistically most likely matches in another language.[78]

Anyone who takes issue with a cell phone or computer because, for example, it did not perform a simple operation has a problem with rationality and reality. He ascribes to a machine property that it does not have: a computer or a cell phone is not an interlocutor, but a tool, certainly it is more complex than other tools, but it is still a device that can be described with the help of physics, without desires, beliefs and especially without thought.

Computers can certainly successfully simulate human thought; indeed, they are able to execute more accurately and quickly many of the thought processes of humans, but despite this often-perfect simulation, they are unable to form their own idea of things, have no consciousness or insights.

According to some, however, the shift from software systems to machine learning systems, which are able to develop their own rules on the basis of the rules given in advance, means that AI does not merely simulate human thinking, but produces its own thinking.

However, I believe that this thesis should not be accepted because the thinking of human beings follows different logic than that of AIs.

The AI used by companies like Amazon or Netflix, when they recommend us a book to buy or a movie to watch, is not intelligent and does not really care what our interests and tastes are. The algorithm is merely using our previous buying or viewing behavior to recommend other products that we might be interested in and therefore willing to spend money on.[79]

---

[78] Cardon, 2018.

[79] White, 2017.

The way an AI acts responds to a utilitarian ethic.

The output provided to us is not the result of critical thinking or moral evaluation but is the result of an optimizing calculation that pursues maximizing the outcome as its ultimate goal. From a mathematical point of view, this amounts first to establishing a value function that can evaluate all consequences of actions according to the extent to which they realize the expected value, then to calculating the expected value of different decision options on the basis of the probabilities they have of realizing themselves, and finally to selecting the action option with the highest expected value. In performing the calculations therefore, AIs are not influenced by moral feelings or perceptions and do not make their own judgments. Therefore, I believe that precisely because of the lack of these elements, AIs cannot be considered intelligent.

On the other hand, human intelligence derives precisely from the fact that what characterizes us as beings endowed with reason is our ability to take a position in a value sense. Indeed, this position taking is based on the capacity for judgment, thoughtfulness, and deliberation. These capacities can never be replaced by an algorithmic rule. For example, a person in deciding may be willing to give up certain advantages in order to achieve in the (even distant) future certain goals and perhaps gain greater utility. I may, for example, decide to begin university studies in the hope that obtaining a degree will guarantee me a better job position in the future than I could obtain without having completed the same course of study.

Philosopher John Searle, opposing Alan Turing's thesis, also argues that AI cannot be equivalent to the human mind because since the human mind possesses intentionality, and the computer does not, the computer cannot have a mind.[80]

The thesis advocated by Turing, and thus by the so-called "strong AI," states that a computer can achieve the same results as a human mind, that is, the ability to think, to have cognitive states, to understand speeches and questions to answer. The program consists of symbols and computational rules that enable the machine to perform a determined process of manipulating symbols with which it composes answers.

Searle, on the contrary, formulates an objection that the human mind cannot be reproduced solely in syntactic terms, since this disregards its main quality, namely intentionality, which refers back to semantics. Intentionality is the main component of the human mind and is closely related to the event of consciousness. Event of consciousness and intentionality are considered

[80] Searle, 1980.

primitive properties and relate to a human being's ability to formulate his or her goals and to feel emotions.

Searle, therefore, argues that AI cannot be equivalent to human intelligence because it is not enough to process symbol manipulation programs according to syntactic rules to generate mental activity. The fact is that the human mind understands, processes, and expresses itself through a language whose words, on the one hand, are invested with meaning and, on the other hand, determine how a response will be given. In support of his thesis, Searle envisioned a thought experiment called "the Chinese room."[81]

In this experiment, he imagines a person who does not speak Chinese, let alone know the ideograms, alone in a closed room. This person is given, by passing them under the door, paper cutouts with Chinese ideograms and is asked to replicate them in turn, again with ideograms. To do this, he receives some manuals and dictionaries. Although the subject does not understand Chinese, he is still able to establish a rule of association and put the ideograms in an order with meaning. Outside the room there is in fact a native Chinese speaker who, after receiving the corresponding answers, concludes that there must be someone in the room who speaks Chinese himself.

It is clear that something fundamental is missing in this imaginary situation: an understanding of the Chinese language. In Searle's reasoning, the room represents the computer, and his experiment shows that even if the answers provided by a system are equivalent to those a native Chinese speaker would give, we cannot claim that the system understands Chinese. Understanding a language and speaking it presupposes a multiplicity of cognitions. A person who speaks Chinese uses certain expressions to refer to particular objects. By means of certain utterances he pursues definite purposes. Based on what he hears he forms certain expectations, etc. These are all properties that the Chinese room, and thus an AI, does not have, since it neither pursues purposes nor possesses expectations that prove the ability to speak and understand the Chinese language. In other words: the AI simulates understanding a language it does not actually understand.

[81] *Ibid.*

1.7 The risks of technology

As I have pointed out in previous pages, technology today solves most of our daily problems.

We live in an era where technology revolves mainly around the expansion of the full potential of the Internet. Technological innovation has accelerated the transformations of daily life, proving over the years to be one of the greatest tools of change.

Nowadays when we faced with everyday problems, no longer look for solutions in a user manual, encyclopedia, or dictionary, but turn directly to the Internet. We also have an internet service or an App for every daily need: from grocery shopping, to ordering medicine, to booking a cab or a vacation, etc. Moreover, today the Internet not only connects people, but has also managed to connect individuals with things (Internet of Things, or IoT).

The past 30 years have shown that in the face of these innovations, the benefits are considerable in terms of cost, availability of services, flexibility of schedules, and personalization. However, new technologies can generate risks.

What follows is a brief analysis of the main risks that technology, especially technology with AI systems, can cause.

1.7.1 Privacy and data security issues

The concept of Earth as a collection of 208 distinct nations has given way to a singular global village, a tightly packed space where privacy has become an elusive luxury.

In our pursuit of a more interconnected and harmonious world, we've overlooked the lurking perils concealed within these alluring technological advancements. What we once held dear (our locations, actions, and even thoughts) now exist as an open book, shared thoughtlessly across an array of blogs and social platforms, often by our own hands.

The huge production of data and the increasing reliance on technology and digital systems[82] certainly raises issues for privacy and data security.

Unfortunately, our folly lies in assuming that this deluge of information that we produce is ephemeral, vanishing into the digital ether, or worse yet, dismissing our data's significance entirely. Regrettably, we remain oblivious to the reality that these colossal stores of information,

[82] As highlighted in *Section 1.1.*

bolstered by the trailblazing capabilities of AI, are meticulously amassed, traded, and consumed just like any tangible commodity.

In this landscape, privacy has crumbled.

Our lives, with all their intricacies, uncertainties, and emotions, are laid bare for all to see. Our every revelation, every question, and every emotional struggle finds its way onto the vast canvas of blogs and social networks. Paradoxically, the same technologies that promised to unite us have left us exposed, fostering a world where the boundaries between public and private blur into insignificance.

The heart of our miscalculation lies in underestimating the endurance of our digital footprint. We believe that our online confessions and fleeting thoughts will dissipate, yet they're captured in the caverns of cyberspace, adding to the vast reservoirs of data. This information, which might seem innocuous on its own, emerges as a precious resource when woven into the intricate tapestry of AI-driven insights. In the interconnected and to a large extent virtual world, data is of great importance and value. Combinations of huge amounts of data create new data.[83]

Algorithms, relentlessly sifting through these troves, decipher patterns, preferences, and trends that offer unprecedented power to those who wield them.

This colossal fusion of information and AI is the cornerstone of a new economy—one where our personal narratives are leveraged for profit. Our digital selves, unwittingly constructed through posts, likes, and shares, are packaged and sold. Advertisers, corporations, and entities amass these fragments to construct a strikingly accurate mosaic of who we are, what we yearn for, and what makes us tick. In this marketplace of intangibles, personal data stands as a prized currency, exchanged with a fervor akin to the most valuable tangible goods.

Many of the digital services that we use every day seem free, but in fact we pay for them with the data we produce while using a service. This type of relationship is very peculiar because we are not, as a rule, buying a service (as is the case with water, electricity, and all other utilities), nevertheless, we are paying for it with our data.

We may not like this, but it is the price we must be willing to pay.

Surely, however, one might wonder if the users of this new Digital World are aware of this.[84]

---

[83] Norta et al., 2016, p. 19.

[84] This topic will be explored further in the third chapter of this dissertation.

As we traverse this altered world, the cautionary tale reverberates: transparency and interconnectivity bring not only unity but also vulnerability.

A single action, a passing thought, can have lasting repercussions in the ever-watchful digital realm. The very technologies that promised to bridge our divides now call upon us to tread with vigilance, to weigh the allure of connection against the price of exposure.[85]

AI could achieve, and even surpass, the very imagination of the philosopher and jurist Jeremy Bentham[86]: in the future, the extraordinary technologies at our disposal could reduce humans to living inside a gigantic panoptic, an ideal prison in which inmates, continuously guarded by a single jailer, are unable to understand whether or not they are being watched; in this invisible and insubstantial recluse, machine learning would be its shackles and data, instead, the key: it is up to us alone to decide whether we consciously manage our data and escape from this prison or, instead, surrender it to those who wish to manipulate us.

1.7.2 The dark side of AI decision making

Today, AI extends beyond the erosion of our privacy; it now encroaches upon the very process of decision-making. Through intricate analysis of our historical behaviors and those of ostensibly comparable individuals worldwide, machines assume the mantle of authority, dictating not only the news on our screens but also curating our social interactions and connections.

Digital platforms like Facebook or Instagram recommend other users with similar interests to connect with. To do this, data collected about us is used: what posts we make, which users we like, and who we interact with. On this basis, it builds algorithms known as "recommendation systems," which can provide you with exclusive recommendations. They wield influence over our consumption choices, steering us towards specific products and services. In doing so, this digital dominion subtly shapes not only our viewpoints but also the intricacies of our relationships and the tapestry of our societal bonds. By replacing human-curated judgement with data-backed judgement, AI ultimately narrows our field of vision and reduces our social and economic choices in retail, dating, entertainment, education, health care, and job opportunities.

In this way, the salient tradeoff in the AI age is not privacy, but choice itself.[87]

---

[85] White, 2017.

[86] Bentham, J. (2012). The panopticon. In *Offenders or Citizens?* (pp. 13-15). Willan.

[87] See e.g. here: https://qz.com/1153647/ai-isnt-just-taking-away-our-privacy-its-destroying-our-free-will-too.

The influence of AI on our choices has an even greater impact because of the lack of transparency in decision-making processes.

Many legal issues focus on the mindset of the decision maker, such as the intention of the manager who decides to fire an employee, and the process by which a person reaches a decision, such as the factors a judge considers when determining the appropriate sentence for a person convicted of a crime. The legal system has developed processes for examining the behavior of humans but is struggling to find an analytical framework for examining AI decision-making.[88] Critics note that AI uses a decision-making process that is a "black box," meaning that it relies on algorithms so complex that the people affected by the decisions made by the systems cannot understand them and the government is unable to regulate them properly.[89]

1.7.3 Biased algorithms

Bias is another fundamental concern associated with "intelligent" systems.

While we might perceive algorithms as mere mathematical constructs, ostensibly neutral in nature, empirical studies have starkly demonstrated that algorithms are not impervious to the taint of human bias.

Bias can be intentionally introduced into algorithms by the people who design them. Programmers possess the capability to encode bias by leaning upon data that is inherently skewed against specific racial or religious groups, perpetuating historical prejudices. Such biased programming maneuvers can contort algorithmic outcomes into forms that unjustly discriminate against minorities or women. The very instruction given to algorithms to bestow disproportionate weight upon factors that are proxies for gender or racial biases can culminate in algorithms that amplify these toxic inclinations.

For example, in the article published on October 11, 2018, by Reuters news agency, it was mentioned that within Amazon, for several years, a research team had been developing experimental software with the purpose of examining candidates' resumes to evaluate them for possible employment.[90]

---

[88] See e.g. here: https://www.americanbar.org/groups/litigation/publications/litigation_journal/2020-21/fall/artificial-intelligence-and-legal-issues//.

[89] The increasingly widespread use of AI in a wide range of industries increases doubt about whether the technology is protected from scrutiny because of the complexity of the algorithms or trade secrets. The companies that design and use these algorithms consider them proprietary. Requests for disclosure of algorithms and information on how calculations are made are generally rejected on the grounds that these formulas are confidential business data that companies have the right to protect.

[90] See here: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

According to the article, the algorithm, trained using data from applicants and previous employees, had learned to identify phrases and words that favored resumes deemed good, but at the same time penalized those that contained terms associated with women or women-only educational institutions.[91]

Furthermore, the implicit biases embedded in programmers' psyches can exert inadvertent influence over algorithmic design choices. Even well-intentioned designers can unconsciously incline towards data that favor particular groups while disadvantaging others, unconsciously perpetuating skewed perspectives.[92]

Similarly, when insufficiently comprehensive data underpins the system's training, the outcome invariably reflects this dearth, further perpetuating skewed results.

When AI infiltrates decisions like hiring, promotions, or pay raises, it morphs into a tool capable of unlawful employment discrimination. Even an ostensibly impartial algorithm for ranking promotion candidates might subtly encode identifiers for race or utilize variables correlated with race, like educational attainment or residential location. In instances where AI learns from past successes, if those successes historically favor a particular race, the algorithm inadvertently perpetuates this bias.

For example, on May 23, 2016, the investigative newspaper "ProPublica" described software ("COMPAS") used in some U.S. courts to estimate the likelihood of a defendant becoming a recidivist.[93]

The article stated that those scores had a bias against African American defendants, after comparing the "false positive" and "false negative" rates of different ethnic groups.[94] The conclusion was: "*Black defendants were often predicted to be at a higher risk of recidivism than they actually were*" and "*white defendants were often predicted to be less risky than they were.*"

Even machine translations may be subject to unintentional bias, since they are based on signals extracted from natural data. At the time of writing (August 2023) Deepl.com translates the English sentence: "The president met the senator, while the nurse cured the doctor and the

---

[91] Amazon had made no official statement on the matter, except to point out that the software in question had never been used to evaluate applicants.

[92] An illustrative example resides in facial recognition algorithms trained predominantly on images of White men, which subsequently falter when tasked with identifying women and people of color. This stems from the fact that the dataset is inherently skewed and fails to capture the diversity inherent in society.

[93] See here: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[94] The article caused great alarm in the media, but not all scholars agreed with its authors' conclusions, and then the whole affair turned into a legal, political and academic controversy.

babysitter" as follows: "Il presidente ha incontrato il senatore, mentre l'infermiera ha curato il medico e la babysitter". Although the English version contains no indication of the gender of president, senator, nurse, and all other workers, the Italian translation attributes them.

Addressing these biases in AI systems requires both vigilance in dataset selection and programming practices, as well as a commitment to ongoing evaluation and improvement to ensure that technology serves as an unbiased tool rather than an amplifier of societal inequities.

1.7.4 Job displacement?

As I have pointed out in Section 1.6.2 humans are afraid that AI will surpass humans in certain activities and therefore has the potential to disrupt industries and lead to job displacement. Despite the enormous potential that new technologies have, I do not believe that an entire job replacement will occur.

A reading of economic history suggests that when a task is mechanized or automated, workers still find new ways to stay involved as employees. For example, in 1970, one-third of women in the U.S. workforce were secretaries; with the introduction of personal computers and word processing software, the need for secretaries dramatically decreased, but the overall number of women employed increased.[95]

Instead, I believe that AI will be used for jobs that people do not want to do: the dull, dirty, and dangerous ones.[96]

Experience confirms this: it has gone well in using robots to defuse bombs, in rescue operations during a natural disaster, in performing repetitive tasks on the assembly line, even in cleaning the living room at home of dust. So, some jobs are close to disappearing precisely because they involve performing tasks that intelligent machines can do with ease by replacing human labor.

According to some, this will give humans much more free time to explore and express their interests and talents.

Kevin Kelly, one of the founders of Wired, wrote about this: "*We need to let robots take over. They will do jobs we have been doing, and do them much better than we can. They will do jobs we can't do at all. They*

[95] Jordan, 2022, p. 120.

[96] These characteristics are summarized in the expression "three Ds".

will do jobs we never imagined even needed to be done. And they will help us discover new jobs for ourselves, new tasks that expand who we are. They will let us focus on becoming more human than we were."[97]

In recent months, concerns about job displacement have begun to intensify again in the face of so-called "generative AI." These "intelligent" systems can generate text, images, video, music or other media in response to requests from a user.[98]

Journalists, programmers, screenwriters,[99] and many other categories of workers are concerned that these systems will overtake them in their tasks and thus take away their jobs.

These concerns, while legitimate, to date I do not believe are well-founded. Indeed, it would be enough to try using these tools and one would realize that, at least for now, they cannot yet surpass the man in novelty, creativity, passion, and critical thinking with which he can complete a job.

"Intelligent" systems delude that they are creative, but reality they are not. As mentioned above, these systems can only process something different or creative because they learn from the data that the human gives them and only based on the goals that the human gives them.

So, a programmer can give these "intelligent" systems and millions of pictures and images and then train them to do a certain task (e.g., transform photographs into a certain artistic style) but creativity is and will always remain the prerogative of the human being.[100]

1.7.5 More cybercrimes

As technology advances, so do the tools and methods used by cybercriminals. Activities such as hacking, identity theft, phishing, and ransomware attacks pose substantial risks to individuals, businesses, and critical infrastructure systems.

---

[97] Kelly, K., (2012, December 24). Better Than Human: Why Robots Will - And Must - Take Our Jobs. *Wired* (blog), https://www.wired.com/2012/12/ff-robots-will-take-our-jobs/.

[98] For further information see e.g., here: https://en.wikipedia.org/wiki/Generative_artificial_intelligence or here: https://research.ibm.com/blog/what-is-generative-AI.

[99] In May, thousands of U.S. screenwriters crossed arms to demand greater protections for working conditions, greater transparency on copyright and in streaming, and protection from possible threats from AI. For further information see e.g., here: https://tg24.sky.it/spettacolo/cinema/2023/09/15/sciopero-sceneggiatori-hollywood.

[100] On this issue see e.g., what Marina Geymont says here: https://www.raiplay.it/video/2022/10/755H-Geymonat-7f6ef247-37cc-44d9-96f7-02f4396363bc.html.

1.7.6 Health risks and the environmental impact

Finally, society's growing dependence on technology and daily increase in the use of technological devices[101] can have a significant human health and environmental impact.

As far as human health is concerned, various studies suggest that the use of electronic devices has effects on the human brain: from social and emotional development, through altering sleep patterns[102] to "lazy thinking."[103]

But the negative effects related to incorrect use of technological devices are also others: from problems with the musculoskeletal system due to positions assumed for too long (back pain, neck pain, "mouse tendonitis") to dermatological ones (reduced tone, decreased brightness and increased dryness of the skin). No less serious is the risk of obesity related to prolonged use of cell phones, as well as tablets and computers, as they promote sedentariness, a major risk factor for overweight, cardiovascular disease and diabetes.[104]

Instead, regarding the environmental impact of electronic devices, there are mainly two areas to be analyzed: from a "macro" point of view is the production phase, while from a "micro" point of view is the individual's use of them.

Building a device requires a considerable amount of fossil fuels, materials (including toxic ones), rare minerals, and water. In extracting raw materials to manufacture the components, there is a large environmental impact that also involves substantial use of electricity. But we must also consider that these final devices are then packaged and transported over long distances. So, before we start using a lot of energy to operate them, already the production and distribution of them causes harmful effects on the environment.

The individual's use of technological devices also has effects on the environment. From the increased consumption of electricity within our homes or offices causing global warming, pollution, and depletion of limited resources to the disposal phase of such devices.[105]

---

[101] See the statistics in *Section 1.1.*

[102] What affects it is the type of blue light emitted by the device screen.

[103] E.g., it is no longer necessary to do mathematical calculations by hand, memorize phone numbers or transcribe notes, because all these operations can be performed faster and more neatly by electronic devices.

[104] For further information you can see: https://www.grupposalutepiu.it/salute/smartphone-pc-prevenire-i-danni-alla-vista/; https://press.rsna.org/timssnet/media/pressreleases/14_pr_target.cfm?ID=1989.

[105] This is referred to as WEEE (Waste Electrical and Electronic Equipment), and according to the European Commission it is the fastest growing waste in Europe. See e.g., the Directive 2012/19/EU of the European Parliament and of the Council of 4 July 2012 on waste electrical and electronic equipment (WEEE).

CHAPTER 2.

The limited rationality of human beings between heuristics and biases

Summary:

2.1 The importance of choice — **2.2** The standard economic approach to decision making: rational choice theory — **2.3** The ration choice model begins to crumble: Allais' paradox — **2.4** The roots of behavioral economics: Simon and the concept of bounded rationality — **2.5** The new economy of the second 20th century: exploring the impact of behavioral economics on human decision making — **2.6** Beyond behavioral economics: some cognitive biases that influence our daily lives

2.1 The importance of choice

Understanding the human mind since ancient times has always been a chimera for humans. To know the brain is in fact to know oneself. Despite the considerable advances made over the years in the field of neuroscience and the strides made possible by technological progress, the investigation of human cognitive processes continues to be a field steeped in considerable complexity. In fact, it is crucial to consider and understand how our minds process and organize the information and how this affects our decisions in everyday life.

Choice concerns the way individuals allocate their time of responding among available response options.[106] Humans live in a world characterized by limited resources where there is a constant need to deal with situations in which making choices is necessary. It could be a choice between different goods or different behaviors. Not choosing can also be considered a choice and often with relevant implications.[107]

Therefore, important questions arise when talking about choices: how do people take decisions? Which are the factors that influence their choices? When a person has to decide, is he or she able not to be influenced by external factors? Can weigh a choice and decide rationally?

In the following paragraphs, I will examine how certain contributions within the realms of economics and psychology have endeavored to address these inquiries throughout history.

---

[106] Fisher, 1997.

[107] For an extensive description see Cass Sunstein (2015): "Choosing not to choose: Understanding the value of choice."

2.2 The standard economic approach to decision making: rational choice theory

Over the centuries, economics has undergone profound changes that have redefined its role within society. In Aristotle's time, it was regarded simply as the administration of the household economy, confined to the private sphere and subordinate to disciplines such as politics and ethics.[108] Wealth was seen only as a means of obtaining specific goods and services, and not as an end, since the pursuit of profit was considered against nature.

During the Middle Ages, with the fall of the Roman Empire, the economy went through a phase of considerable retrenchment. Currency was abandoned in favor of barter, and agriculture dominated at the expense of manufacturing. Landowners received tribute in kind. Because money was considered impure and a source of sin, the economy was based on barter, and activities paid for with currency were discredited by society. This was mainly because the economy did not enjoy autonomy but was completely subject to morality.

Around the year 1000, with the beginning of the late Middle Ages, the Western world experienced a revival supported by population growth and the return of monetary circulation. This encouraged the development of commercial activities within cities and the formation of new artisans. In this historical context, the role of merchants became increasingly significant, prompting them to undertake long journeys to acquire new resources. At the same time, banking activities revived, thanks to the emergence of trade contracts that financed shipments in exchange for a share of the earnings. During the late Middle Ages, trade became so important that the economy was transformed into mercantilism, a system of economic policy characterized by state intervention through protectionist policies to support exports and limit imports through the application of duties.

This historical transformation led to a radical change in the conception of economic theory. From the simple administration of the domestic economy, there was a shift to economic policy, where the state took a proactive role in implementing policies to increase the country's wealth and power.

In the face of this evolution, the economic needs of the state changed, giving rise to Neoclassical Economics.

---

[108] The word "economics" comes from the Greek word "oikonomia" (*οικονομια*) meaning management of a household.

The term "Neoclassical Economics" dates back to 1900, when the American economist Thorstein Veblen used it in his paper "The Preconceptions of Economic Science."[109] Two main periods can be traced within Neoclassical Economics: an early stage and a postwar stage.

In Classical Economics[110] and the early stage of Neoclassical Economics, it was widely accepted to talk about cognitive and affective states. The main author of this phase was the economist William Stanley Jevons who said that the subject of economy should have been "*maximising pleasure and reducing pain*."[111] Early neoclassical economists found no compelling reasons to embrace alternative methods for assessing the soundness of the underpinnings of their economic theories. They placed their trust in the method of introspection to analyze individuals' decisions and were convinced that introspection upheld the principles of hedonic psychology.

After the Second World War when the dissatisfaction of several economists towards the results achieved by the early stage of Neoclassical Economics brought change to the approach of the study of decision making. Postwar neoclassical economists wanted to root their discipline in a solid methodological ground and at the same time to improve the predictive power of their theories. They claimed that economy should refer to conscious states, so they rejected the idea that introspection was a scientifically acceptable means to explore such states.

The basic concepts of pleasure and suffering as foundation of choice, were substituted by a theory of preferences. People's sensations of pleasure and suffering are not observable, while their choices can be observed directly. After assuming that people's choices reflect their preferences it was possible to test empirically what people prefer. By substituting the concept of "utility" with the one of "preference", postwar neoclassical economists explicitly intended to separate economy from psychology.[112] However, it is important to highlight that they did not deny that people might be motivated by pleasure, pain and/or other mental states. Postwar theorist simply chose to remain agnostic about questions like motivation and preference formation arguing that such issues were outside the scope of economy.[113] As a result, they formulated a comprehensive theory that overlooked some crucial nuances of human behavior, arriving at the so-called marginalist revolution.

---

[109] Veblen, T. (1900). The preconceptions of economic science. *The quarterly journal of economics*, *14*(2), 240-269.

[110] Classical economics refers to one of the prominent economic schools of thought that originated in Britain in the late 18th century. The link between economic and psychological principles can be traced back to the work of the philosopher and economist Adam Smith (1723-1790). Although he did not have a theory of decision making in the modern sense, his vision of human nature appeared multifaceted, and he was deeply interested in the psychological underpinnings of human behavior.

[111] Jevons, W. S. (1879). *The theory of political economy*. Macmillan, p. 37.

[112] Robbins, L. (1984). *An essay on the nature and significance of economic science*. New York University Press, p. 85.

[113] Robbins, 1984, p. 86.

With the advent of the marginalist revolution between 1870 and 1890, economic theory underwent a transformation that toned down its social aspect, leading it to distance itself from the social sciences and adopt an approach more akin to the natural sciences and economic positivism. Until then, economic theory had retained a social element by recognizing that economic phenomena were closely intertwined with and influenced by a variety of social factors. In contrast, with the isolation of social factors from the scope of economic study, a radical shift occurs in the process of economic analysis. The latter is based on analyzing economic events in isolation, as if their occurrence is not dependent on or influenced by external factors. The new paradigm underlying this theoretical conception is based on the concept of the complete rationality of the individual. This individual, guided by a budget constraint,[114] equipped with accurate information and aware of available alternatives, is able to make decisions that maximize his utility and, consequently, his level of well-being.[115]

Economic science in its new dimension focuses on the question of optimal resource allocation within a context in which individual needs are unlimited, but the resources available to meet them are limited. In such a situation, it is necessary to set priorities to optimize the allocation of available resources and consequently maximize overall welfare. In the case of a single individual, who is faced with inexhaustible wants and needs but limited resources (such as income), it is not possible to meet all his or her needs in an unlimited way. Therefore, each individual should rationally determine which needs he considers most crucial and likely to procure the greatest satisfaction. This will enable him to allocate resources in such a way as to maximize his well-being, measured in terms of utility. Consequently, the object of economic theory narrows down to problems related to resource allocation. Aspects related to ethics, equity and justice are eliminated from this perspective because it is not the task of economics to guarantee these values. Instead, these issues become the purview of the social sciences.

Homo Oeconomicus is a rational agent who has consistent and stable preferences; he is entirely forward-looking and pursues only his own self-interest. When given options he chooses the alternative with the highest expected utility for himself.[116]

---

[114] In economics, a budget constraint represents all the combinations of goods and services that a consumer may purchase given current prices within his or her given income (for further information see e.g., here: https://en.wikipedia.org/wiki/Budget_constraint).

[115] Krugman, P. R., Wells, R. (2018). *Microeconomics*. Macmillan Learning.

[116] Oxford Dictionary. (n.d.). Homo Oeconomicus. In *Oxford Dictionary online*. Last accessed August 9, 2023, https://www.oxfordreference.com/display/10.1093/oi/authority.20110803095943203;jsessionid=3D815EC47510A54A5B09832AD432A53B.

This gives rise to the "rational consumer" model, which allows for reasoning in terms of utility maximization given certain information and budgetary constraints.

The theoretical model of the rational consumer, first developed by Daniel Bernoulli in the 18th century and later developed by John von Neumann and Oskar Morgenstern in the 20th century, defines how, based on personal tastes and preferences, individuals generate their own utility function.[117]

The utility function is influenced by all goods and services consumed, defined by economic theory as the consumption basket. The relationship between that personal basket and the utility generated defines the utility function. The latter graphically that has a positive slope that, however, tends to decrease as the number of units consumed of a given good or service increases.

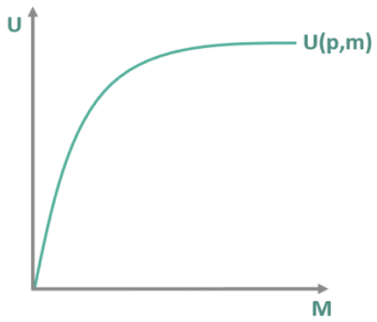

Figure 4. Utility function[118]

To maximize one's individual well-being, it is necessary to understand how utility varies as the quantity consumed increases.

For the purposes of the model, it is therefore important to focus on marginal utility, which is the change in total utility produced by the consumption of an additional quantity of a good or service. The marginal utility curve has a negative slope because the consumption of an additional unit contributes less to an individual's welfare than before.

[117] Each individual, according to traditional economic theory, produces through his or her personal preferences a different utility function.

[118] Image source here: https://www.wallstreetmojo.com/utility-function/.

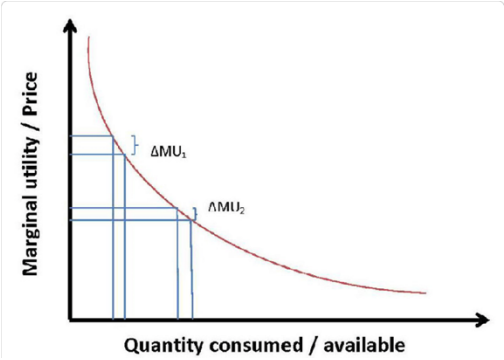

Figure 5. Marginal utility curve[119]

Thus, a perfectly rational consumer will continue to consume until an additional unit exhibits negative marginal utility, which will result in a reduction in total utility relative to the previous unit. Since in order to take advantage of an additional unit of a particular good or service it is necessary to take on an additional cost, then in the face of limited income it is automatically necessary to reduce the quantity consumed of another good or service.

Individual income is the budget constraint, as the cost of the consumption basket cannot exceed total income. Should a consumption basket exceed disposable income, then it could not be consumed. The budget line then is the segment that shows all the consumption baskets that can be purchased by employing one's entire income.

Given a given budget constraint, which defines the consumption basket accessible to the individual, and considering the utility function, it is possible to find the optimal consumption basket, that is, the combination of goods or services that maximizes utility given a certain income.[120]

From this analysis it is possible to infer how, according to traditional rational economic theory, individuals are rational in their consumption choices, perfectly knowing their tastes and acting consistently with them; based on the information they possess, they are able to make choices designed to maximize their utility without being influenced by additional external elements.

---

[119] https://www.researchgate.net/figure/Curve-of-Diminishing-Marginal-Utility-which-doubles-as-the-axiomatic-Demand-Curve-Here_fig5_304811331.

[120] The analysis of this *Section* was carried out using the following as sources:

- Besanko et al., 2020;
- Krugman et Wells, 2018.

2.3 The ration choice model begins to crumble: Allais' paradox

Thus, "*if you look at economics textbooks, you will learn that homo economicus can think like Albert Einstein, store as much memory as IBM's Big Blue, and exercise the willpower of Mahatma Gandhi.*"[121]

Effectively, this is the image of the economic agent that the economic-mathematical models present to us.

An individual who knows what he wants, knows how to get it (no matter how complex the calculations he has to do to find the optimal solution) and rationally and consistently does whatever it takes to achieve his goal and to maximize his or her own utility, without being influenced by other factors.

But is this always true? Are we made that way? Are we completely rational?

Although Neoclassical Economics had achieved a predominant position in the 20th century due to its formal and axiomatic approach, criticism was not slow to emerge. Indeed, some economists began to consider that their discipline would benefit from a closer integration with psychology. The idea became widespread that the theoretical models developed up to that point had neglected the human factor and its implications in decision making.[122]

A prime example of this new perspective is the experiment conducted by French economist Maurice Allais with participants at an international congress devoted to the theory of rational expectations, held in Paris in 1952.[123] Subsequently, Allais described the main events of that experiment in an article published the following year.[124]

According to Allais, any study of rationality in economics must take into account the following elements of complexity: "(i) the distinction between monetary and psychological values; (ii) the distortion of objective probabilities and the appearance of subjective probabilities; (iii) the mathematical expectation of psychological values (the mean of the probability distribution of psychological values) and (iv) the dispersion (variance) as well as general properties of the form of the probability distribution of psychological values."[125]

---

[121] Thaler and Sunstein, 2009.

[122] Neoclassical Economics had favored the search for universally valid rules, but it had neglected the analysis of less rigid patterns that were more in keeping with the complexity of human reality.

[123] Maurice Félix Charles Allais (31 May 1911 – 9 October 2010) was a French physicist and economist, the 1988 winner of the Nobel Memorial Prize in Economic Sciences (for further information see e.g. here: https://en.wikipedia.org/wiki/Maurice_Allais).

[124] Allais, 1953.

[125] Allais 1953, p. 504.

Allais' experiment consisted of asking "people considered perfectly rational" to choose between two different scenarios:

SCENARIO 1	SCENARIO 2
A. *You have a chance to earn €1 million with certainty.* B. *You have a lottery in which there is an 89% chance of winning €1 million and an 11% chance of winning nothing.*	C. *You have a chance to earn €5 million with certainty.* D. *You have a lottery in which there is an 89% chance of winning €5 million and an 11% chance of winning nothing.*

According to a rational choice, the choice of situation (A) in the first Scenario should have imposed consequently the choice of (C) in the second. The results of the experiment conducted by Allais, however, showed that the majority of people had indeed chosen (A) in the first case, but (D) in the second.[126]

This is what constitutes the Allais' paradox: the inconsistency in choosing between the two scenarios even though the probabilities are identical. In fact, according to traditional economic theory, people should evaluate decisions based on their expected utility (i.e., the probability of gain multiplied by the value of gain). In Allais' paradox, people seem to give more weight to the certainty of gain rather than the higher probability of higher gain.

This paradox led to a greater understanding of human choice patterns and the limits of rationality in economic theory, also paving the way for cognitive research and interpretation of the many anomalies found in rational choice.

Scholars began to consider whether it was realistic to assume that individuals were so capable of conducting extremely intricate decision-making processes, or whether models of rational behavior should be interpreted more in a normative sense. This would have seen them as decision support tools, suitable for use by experts but not necessarily for ordinary decision makers.[127]

[126] Allais 1953, p. 527.

[127] Egidi, 2006.

2.4 The roots of behavioral economics: Simon and the concept of bounded rationality

The roots of behavioral economics can be traced back to the work of Nobel laureate Herbert Simon in the 1950s and 1960s.[128] He is remembered for criticizing the idea of the completely rational economic agent and introducing the concept of "bounded rationality."

According to Simon, humans are unable to behave as rational subjects because of limitations inherent in their rationality. These limitations come from two elements: the context (the decision environment and the time in which a choice is made) and the limits of the solutions achievable by the agents (i.e., information, available time, and subjective analytical capabilities).

The consequence of these limitations is that the decision maker, based on the processing of these limiting factors, develops cognitive and symbolic processes that lead him or her to come to conclusions that may be wrong or inconsistent with preferences, thus resulting in solutions that do not maximize one's expected utility.

The limitingly rational decision maker approach, introduced by Simon, becomes relevant when the decision maker is faced with situations where it is impossible to identify an optimal choice or where the computational cost is too high. In such circumstances, the individual is inclined to look for an alternative that provides satisfaction rather than devoting himself to finding the optimal solution. This approach is called "satisficing,"[129] which is opposed to the optimization inherent in perfect rationality theory.

The concept of "satisficing" refers to procedures by which the existence of satisfactory decision alternatives is made possible by dynamic mechanisms for adjusting aspiration levels to reality, both on the basis of available information regarding the environment and taking into account the time resources allocable for such operations.[130]

To confirm his hypothesis, Simon used as example the decision-making strategies employed in the game of chess.[131]

---

[128] Herbert Alexander Simon (June 15, 1916 – February 9, 2001) was an American political scientist, with a Ph.D. in political science, whose work also influenced the fields of computer science, economics, and cognitive psychology. His primary research interest was decision-making within organizations, and he is best known for the theories of "bounded rationality" and "satisficing" (for further information see e.g. here: https://en.wikipedia.org/wiki/Herbert_A._Simon).

[129] A portmanteau of the terms "satisfy" and "suffice."

[130] Simon, 1972, pp. 168-169.

[131] The adoption of the game of chess, as a kind of mirror reflecting some properties of the decision-making processes employed in the real world, had already been proposed by von Neumann and Morgenstern in their joint work on game theory, and it is no coincidence that it was also used by IBM in the elaboration of Deep Blue (see *Section 1.6.2*).

Simon points to the fact that player regularly focus on far fewer strategies than are possible with each move:

"*Studies of the decision-making of chess players indicate strongly that strong players seldom look at as many as one hundred possibilities - that is one hundred continuations from the given position - in selecting a move or strategy. […] Chess players do not consider all possible strategies and pick the best, but generate and examine a rather small number, making a choice as soon as they discover one that they regard as satisfactory.*"[132]

The generation and evaluation of alternatives often occur through habit-driven processes and repetition of decision-making procedures that are ingrained in the subject's "cognitive programming."

The tactical short-range considerations just recalled, as well as the possible cognitive limitations that exist on a personal basis, would be the same as those that occur in the decision-making process as a whole: when agents decide, in short, they are either unable to consider all possible alternatives, or for reasons of time and energy to be expended they do not want to do so, thus falling under the operational razor of what the psychological literature calls "aspiration levels", or thresholds of sub-optimal decision-making.[133]

The concepts developed by Simon represented a marked departure from earlier positions in traditional economic thought. These traditional positions assumed that once the individual's capacity for action was accepted as an axiom, there was no further need to consider the subject's actual cognitive and volitional abilities. However, Simon opposed this perspective and outlined the path for a new behavioral approach in economics in the latter 20th century.

[132] Simon, 1972, p. 166.

[133] Arnaudo, 2012, p. 42.

2.5 The new economy of the second 20th century: exploring the impact of behavioral economics on human decision making

Simon's proposed notion of rationality, examined in the previous section, focuses on both the procedural aspect of subjective decisions and the decision-making environment in which these deliberations take place. It is precisely these two elements, as already pointed out, that form the characteristic basis of that new line of research that has emerged under the name "Behavioral Economics" (hereinafter also "BE").

Several authors attempted to define BE. Camerer and Loewenstein described it as an approach for understanding decision making and behavior that integrates behavioral science with economic principles:

"*behavioral economics increases the explanatory power of economics by providing it with more realistic psychological foundations [...] At the core of behavioral economics is the conviction that increasing the realism of the psychological underpinnings of economic analysis will improve the field of economics on its own terms - generating theoretical insights, making better predictions of field phenomena, and suggesting better policy.*"[134]

Richard Thaler gave a similar definition in the "Yearly Guide for Behavioral Economics":

"*I view behavioral economics to be economics that is based on realistic assumptions and descriptions of human behavior. It is just economics with more explanatory power because the models are a better fit with the data.*"[135]

Although different definitions have been provided, most of the experts in the field agree on a fundamental concept: the aim of BE is to provide an adequate model of human behavior. In fact, at the heart of BE is the attempt to adapt the concept of bounded rationality to neoclassical studies, which, by contrast, assumed a perfectly rational economic agent.

Behavioral economists do not reject modeling practices of rational action per se but seek to refine them to reduce the discrepancy between observable reality and theoretical models. They recognize that in certain contexts economic agents behave as perfectly rational individuals, while in other situations they are influenced by interdependent preferences, emotions, and cognitive limitations. These factors can lead to suboptimal or even contradictory choices with respect to the rational choice model.

Over the past 50 years, growing dissatisfaction with traditional economic models has turned BE into one of the most relevant and discussed fields in economics. This has also been made

---

[134] Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2004). *Advances in behavioral economics*. Princeton university press, p. 3.

[135] Samson, A. (2016). The behavioral economics guide 2016 (with an introduction by Gerd Gigerenzer), p. 23.

possible by multidisciplinary collaboration among scholars from different disciplines, such as psychology and philosophy, who have helped develop this new area of research.

A multidisciplinary approach made it possible to approach the topic from different perspectives, overcoming some of the paradigms that have characterized the development of traditional economic theory.

It is no coincidence, in fact, that the use of the label "behavioral economics," although used since the 1950s,[136] is normally reserved for a course of study and research that began only in the early 1970s, traceable to a few well-identified researchers: these were two Israeli-born psychologists, Amos Tversky and Daniel Kahneman, who were joined shortly thereafter by a U.S. economist, Richard Thaler.[137]

The groundbreaking work of Kahneman and Tversky for convenience is divided by the authors themselves into three distinct research programs[138]:

i. the Prospect Theory (a model of choice under risk and with loss aversion);

ii. the framing effects with their implications for rational agent models;

iii. and the heuristic and bias program.

Before going into the merits of these three different research programs, however, it is necessary to focus on one aspect that serves as their premise: Kahneman's subdivision between "System 1" and "System 2" to describe the characteristics of the thought processes that people use in their daily choices.

Having done so, the dissertation will continue with a brief description of the first two research programs, and then focus more on the third (the heuristic and bias program).

---

[136] Several researchers including Allais and the Hungarian George Katona were avowedly skeptical of the axiomatic structure regarding the rationality of human behavior that economic studies were taking on.

[137] In general, the influence of Simon's work is felt. It was in the context of theorizing about the limited cognitive abilities of agent subjects that the most famous and crucial behavioral studies were born.

[138] Kahneman, 2003, p. 1449.

2.5.1 The dual cognitive system

Figure 6. Angry woman photo (Kahneman, 2017).

What did you think as soon as you saw this picture?

At first glance, an average person sees her dark hair and angry expression. In addition, what you see has extended into the future. You have a sense that this woman is about to say very rude words, probably in a loud, shrill voice.

All this flow of thoughts came to us automatically and effortlessly. Our reaction to the picture simply happened. This is a case in point of what Kahneman in his book "Thinking, fast and slow" defines as "fast thinking" or intuitive thinking.[139]

Now look at the following problem:

25 x 56

It is easily recognizable that this is a multiplication problem and probably some people have ability to solve it in their heads while others with pen and paper. Some also have a vague intuitive knowledge of the range of possible outcomes, although very few have arrived at an exact solution. Performing this calculation, as well as other reasoning, is an effort. Reaching an answer requires carrying out a deliberate, demanding mental process that follows precise rules.

[139] Kahneman, D. (2017). *Thinking, fast and slow.*

This pattern of reasoning is not only a mental event, for it has effects on the body as well: muscles become tense, blood pressure rises, heart rate increases, and pupils dilate.

Unlike the example in Figure 6 above, in this case you experience what Kahneman refers to as "slow thinking."

Kahneman defines these two different modes of thinking and deciding with the concept of "System 1" and "System 2."

They are two different operating systems, which govern all decisions, and they correspond to the everyday concepts of reasoning (System 1) and intuition (System 2).

Both systems have different characteristics[140]:

i. System 1, the one used for Figure 6, is the intuitive one. This works quickly, automatically, with little or no effort and is much more powerful than we ourselves are aware of;

ii. System 2 in contrast, the one used in the multiplication example, is analytical, systematic that is activated when we encounter mental activities that require focus and concentration.

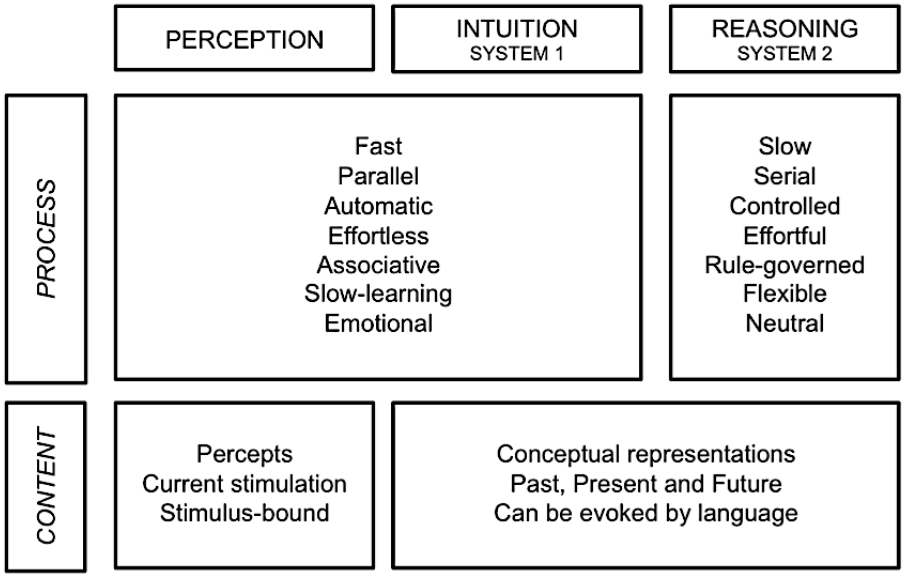

Figure 7. The scheme shows the dual system model of decision-making (see Kahneman, 2003).

[140] Kahneman, 2003.

An example of the operation of these two systems can be seen in the figure 8 below.

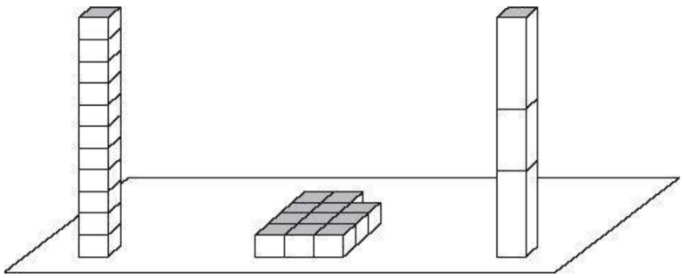

Figure 8. Picture of "different" towers (Kahneman, 2017).

In this case you know that the two towers on the left and right are equally tall and are more similar to each other than the array of blocks in the center.

However, no one immediately knows that the number of blocks in the left tower is equal to the number of blocks arranged on the floor. In fact, to confirm this hypothesis, it is necessary to count the two sets of blocks and compare the results.

This is an activity that only System 2 can perform.

As I have pointed out in these few examples, it can be summarized that System 1 works automatically by itself and System 2 applies the law of least effort, that is, it relies on System 1, when it understands that it is a task easily performed by the latter.[141]

System 2 is normally in a comfortable low-effort mode in which only a tiny fraction of its capacity is engaged. System 1 continuously generates suggestions for System 2: impressions, intuitions, intentions, and feelings. When approved by System 2, impressions and intuitions are transformed into beliefs and impulses into voluntary actions. When everything runs smoothly, most of the time, System 2 adopts System 1's suggestions with little or no modification. One generally believes one's impressions and acts on one's desires, and that is usually fine. However, when System 1 encounters difficulties, it calls on System 2 to support more detailed and specific elaboration that can solve the problem of the moment.

So, System 2 has the role of supervising, directing, and modifying the thoughts and actions "proposed" by System 1 or intervening when a question arises for which System 1 does not offer an answer.

[141] Kahneman, 2017.

Other examples of this process can be seen below.[142]

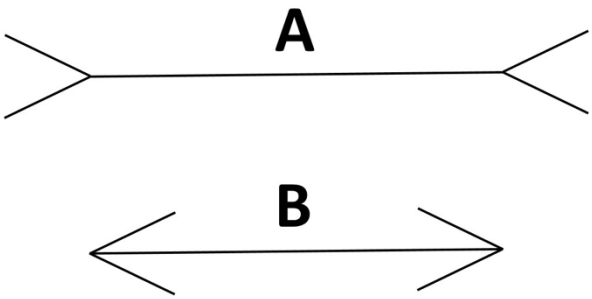

Figure 9. Two horizontal lines image (Kahneman, 2017).

There is nothing special about this image: two horizontal lines of different lengths. The lower line is obviously longer than the upper one.

This is what we all see, and we naturally believe what we see. Someone more observant, however, might recognize that the horizontal lines are actually the same length.[143] As you can easily confirm by measuring them with a ruler you know that the lines are equally long.

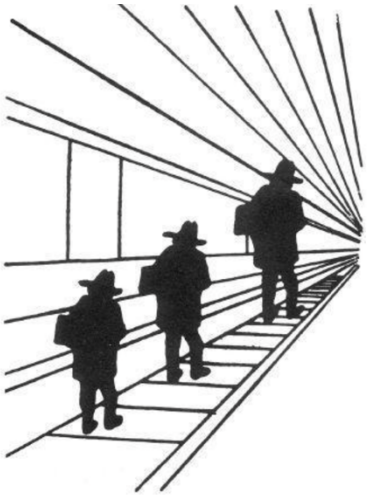

Figure 10. Picture of the three men (Kahneman, 2017).

[142] The following examples are taken from the book Thinking, Fast and Slow (Kahneman, 2017).

[143] This is the famous "Müller-Lyer optical illusion". It was devised by Franz Carl Müller-Lyer (1857–1916), a German sociologist, in 1889 (for further information see e.g. here: https://en.wikipedia.org/wiki/Müller-Lyer_illusion).

In this case, is the figure of the man on the right larger than the figure on the left? The most obvious answer that immediately comes to mind is yes. However, if you use a ruler to compare the two figures, you will find that they are actually the same size. Our impression of relative size is dominated by an illusion that automatically interprets the image as a three-dimensional scene. However, it should be kept in mind that the image is printed on a flat paper surface.

In addition to optical illusions, there are also cognitive illusions. Try reading what is written in rows and columns in the figure 11 below.

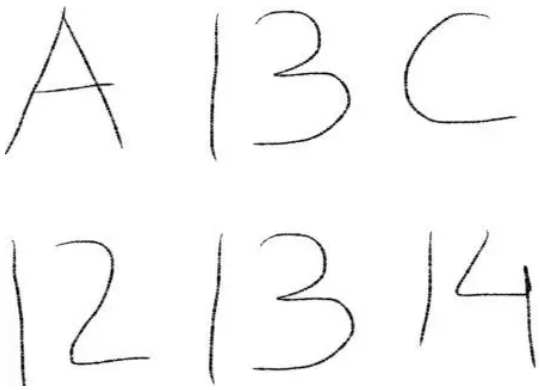

Figure 11. Cognitive illusion example (Kahneman, 2017).

The row probably reads A-B-C, while the column reads 12-13-14. This is in fact the instinctive choice of System 1. Yet, if you look closer, the central font used in the figure is the same. So, it could just as easily read A-13-C, but, from experience, everyone knows that after A comes B.

Another example that is often used is:

"*A bat and ball cost €1.10. The bat costs one dollar more than the ball. How much does the ball cost?*"

A number came to your mind. The number, of course, is 10 (10 cents). The distinctive mark of this easy puzzle is that it evokes an answer that is intuitive, appealing, and wrong. If the ball costs 10 cents, then the total cost will be €1.20 (10 cents for the ball and €1.10 for the bat), not €1.10. In fact, the correct answer is 5 cents.

2.5.2 Prospect Theory (and loss aversion)

After this brief introduction on Kahneman's division between "System 1" and "System 2," it is appropriate to briefly dwell on another contribution, by Kahneman and Tversky dating back to the 1970s, by which neoclassical economic rationality suffered a severe backlash: Prospect Theory.

From a conceptual point of view, this theory is based on the descriptive analysis of empirical results obtained through the application of questionnaires and experiments on different individuals in order to test whether or not the fundamental principle of expected utility theory is respected in practice.[144]

According to the rational choice model developed by Neoclassical Economics, individuals make decisions with the goal of maximizing expected utility. It is assumed that utility from the consequences of choices is completely determined by the final state of resources, thus being independent of the reference point.[145]

Prospect Theory, instead, represents a new a behavioral model that illustrates how people make decisions in contexts characterized by risk and uncertainty, i.e., situations in which the possibility of gain or loss exists. The theory assumes that individual's reason in terms of expected utility relative to a reference point, rather than relying on absolute outcomes. The following are examples of some of the operational strategies adopted by Kahneman and Tversky in their experiments.[146]

> Problem 1*: In addition to whatever you own, you have been given €1,000. You are now asked to choose one of these options:*
>
> *A) 50% chance to win €1,000 or B) get €500 for sure*
>
> Problem 2*: In addition to whatever you own, you have been given €2,000. You are now asked to choose one of these options:*
>
> *C) 50% chance to lose €1,000 or D) lose €500 for sure*

[144] The methodology remembers Allais' experimental approach and it's now defined as "experimental economy." Experimental economics is a branch of economics that studies human behavior in a controlled laboratory setting or out in the field, rather than just as mathematical models. It uses scientific experiments to test what choices people make in specific circumstances, to study alternative market mechanisms and test economic theories (see here: https://www.investopedia.com/terms/e/experimental-economics.asp, last accessed August 16, 2023).

[145] See *Section 2.2* for more details.

[146] The following examples are taken from the book Thinking, Fast and Slow (Kahneman, 2017).

It can easily be confirmed that in terms of the end state of wealth-all that matters for rational choice theory-problems 1 and 2 are identical. In both cases, one has a choice between the same two options: one can have the certainty of being €1,500 richer than one currently is or accept a gamble in which one has the same chance of being €1,000 richer or €2,000 richer.

According to the neoclassical approach, then, the two problems should elicit similar preferences. However, experiments have shown different results. In the Problem 1, a large majority of respondents preferred the sure thing. In the Problem 2, most of the people preferred the gamble.

You are offered a gamble on the toss of a coin.

- *If the coin shows tails, you lose €100.*
- *If the coin shows heads, you win €150.*

Is this gamble attractive? Would you accept it?

To make this choice, you must balance the psychological benefit of getting €150 against the psychological cost of losing €100. How do you feel about it?

Although the expected value of the gamble is obviously positive because you stand to gain more than you can lose, you probably dislike it (most people do).

The rejection of this gamble is an act of System 2, but the critical inputs are emotional responses that are generated by System 1. For most people, the fear of losing €100 is more intense than the hope of gaining €150. For this reason, losses matter larger than gains and that people are loss/risk averse.

These are just some of the examples that have pointed out the weakness of the neoclassical rational choice model.

Bernoulli's theory, according to Kahneman and Tversky, is too simple because it lacks the so-called "reference point," the prior state against which gains and losses are evaluated.

In Prospect Theory therefore, it is not enough to know only the state of wealth to determine its utility, but it is also necessary to know the reference point.

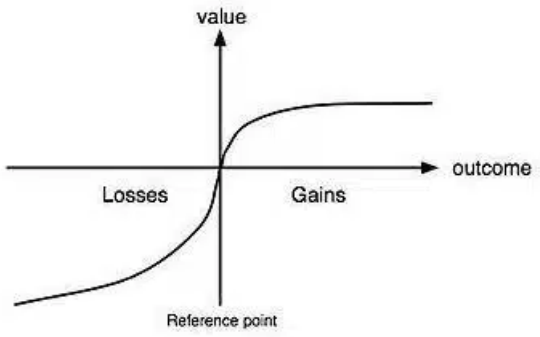

Figure 12. The graph shows the value function described by Kahneman and Tversky in the Prospect Theory (Kahneman, 2003).

The graph shows the psychological value of gains and losses, which are the "carriers" of value in Prospect Theory (unlike Bernoulli's model, in which states of wealth are the carriers of value).

The graph has two distinct parts, to the right and to the left of a neutral reference point. A salient feature is that it is "S-shaped", which represents diminishing sensitivity for both gains and losses.

Furthermore, the two curves of the "S" are not symmetrical. The slope of the function changes abruptly at the reference point: the response to losses is stronger than the response to corresponding gains. This is loss aversion.[147]

2.5.3 Framing effect

The second research's program of Kahneman and Tversky focused on the so called "framing effect."[148].

According to the principle of invariance, an essential aspect of rational choice theory, preferences should not be affected by variations of irrelevant options or outcomes.

Instead, Kahneman and Tversky showed in their experiment how this principle is systematically violated in certain circumstances and how people's decisions are affected by the frame in which a problem is formulated. Their most famous experiment is the so-called "Asian disease" shown below[149]:

[147] Kahneman, 2017.

[148] Tversky and Kahneman, 1981, 1989.

[149] *Ibid.*

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
In the first version of the problem, the possible options were the following: - *If Program A is adopted, 200 people will be saved* - *If Program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved* *Which of the two programs would you favor?*
In the second version of the problem, the possible options were the following: - *If Program A' is adopted, 400 people will die* - *If Program B' is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die* *Which of the two programs would you favor?*

Although the two versions produce the same outcome, they differ only in that the former is formulated in terms of the number of lives saved, while the latter is formulated in terms of lives lost. This, for a rational economic agent, should cause no problems. That is, an individual should choose program A or program B in both versions.

However, the results of their experiments proved otherwise.

While in the first version of the problem, most of the respondents preferred the program A, conversely, in the second version, most people's preference was for the program B. The authors interpret this outcome claiming that for the respondents of the first version of the problem, the certainty of saving people was disproportionately more attractive. Conversely, the certainty of deaths in the second version was disproportionately more aversive.[150]

[150] In another famous experiment, was showed that people's choice between surgery and radiation therapy was changing by describing outcome statistics in terms of survival rates or mortality rates. When the rate was proposed in a frame of survival, the chance that patients choose the surgery option was substantially higher than when a mortality frame was used. For further information see: McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England journal of medicine*, *306*(21), 1259-1262.

2.5.4 Heuristics and bias

A key point in Tversky and Kahneman's analysis is the interaction between System 1 and System 2. In most situations the two systems act in a coordinated way. However, in some cases System 1, which is fast and automatic, conflicts with System 2, which is slower, and reason based. These conflict situations were the subject of the Heuristic and Bias research program conducted by the two Israeli psychologists.[151]

They highlighted that the human cognitive system could rely on a limited amount of resources to solve problems. When the amount of information is too high or complex people are "forced" to rely on mental shortcuts and simplified strategies in order to make decisions. These shortcuts are defined as "heuristics" and they ignore some of the information, with the goal of making decisions more quickly and simply.[152]

Usually, these strategies work properly but in certain circumstances they can lead to systematic mistakes in evaluation. These mistakes are called "cognitive biases."[153] Thus, by bias we define all those judgments or biases that are not based on evidence and hard data but on the information held, which are processed on the basis of particular heuristics.

The problematic aspect is that cognitive biases can sometimes cause perceptual distortions and lead to the formation of opinions and feelings that do not correspond to reality, inaccurate judgments, illogical interpretations, and irrationality.

Tversky and Kahneman's work focuses on three heuristics that have been found to be widely and systematically used during a series of controlled experiments. These are, specifically, the heuristics of: representativeness, availability, and anchoring.

In ideal continuity (albeit unstated) with Allais' paradox (see Section 2.3), the authors noted how "*several of the severe errors of judgment reported earlier occurred despite the fact that subjects were encouraged to be accurate and were rewarded for the correct answers*" and even the judgments of subjects skilled in probability and statistical calculations "*are liable to similar fallacies in more intricate and less transparent problems.*"[154]

---

[151] Tversky & Kahneman, 1974.

[152] Tversky & Kahneman, 1974 and Kahneman, 2011.

[153] Tversky & Kahneman, 1974 and Kahneman, 2017.

[154] Tversky & Kahneman, 1974, p. 1130.

a. Representativeness

When a decision maker has to formulate a solution or assess the probability of an event happening, or has to assign a person to a group, he often draws from his memory stereotypical information. Thus, people make their choices based on the similarity between A and their idealistic image of B, in other words, on how representative A is of B. This strategy is sometimes successful, but very often, leads to mistakes and decisions based on stereotypes instead of probabilistic assumptions.

"Representativeness heuristics" has been brilliantly showed in the "Linda" experiment[155] in which the researchers provided the experimental subjects with the description of a fictitious character called Linda. The description was the following:

> *"Linda is 31 years old, single, outspoken, and very bright.*
>
> *She majored in philosophy.*
>
> *As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations."*

Following, the experimental subjects were asked to estimate the likelihood that Linda would belong to one of the 8 categories listed below:

> *1) Linda is a teacher in elementary school.*
>
> *2) Linda works in a bookstore and takes yoga classes.*
>
> *3) Linda is active in the feminist movement.*
>
> *4) Linda is a psychiatric social worker.*
>
> *5) Linda is a member of the League of Women Voters.*
>
> *6) Linda is a bank teller.*
>
> *7) Linda is an insurance salesperson.*
>
> *8) Linda is a bank teller and is active in the feminist movement.*

[155] Kahneman & Tversky, 1982 and Tversky & Kahneman, 1983.

The two critical items in the list were 6) ("*Linda is a bank teller*") and 8) ("*Linda is a bank teller and is active in the feminist movement*"). The other six possibilities were unrelated and miscellaneous.

As might be expected, 85 percent of respondents ranked the conjunction item 8) higher than 6), indicating that Linda resembles the image of a feminist bank teller more than she resembles a bank teller.

However, this classification could be wrong. In fact, a rational economic agent (as outlined in Section 2.2) would not have hesitated to say the opposite because the answers given in the experiment violate the conjunction rule, which says that the conjunction of two events (bank teller and feminist) cannot be more probable than any of the two events alone (bank teller or feminist):

$$P(A \cap B) \leq P(B)$$

$$P(A \cap B) \leq P(A)$$

The bias was reasonable because the description of Linda was more representative of the conjunction of the 2 options (number 8) than of just one of them (number 6). This phenomenon is commonly known as the conjunction fallacy[156] and as well as in Tom W's experiment[157], demonstrates how the human mind implements heuristics, exploiting similarities that allow it to make a quick and seemingly right choice, but which is found to be rationally wrong.

b. Availability

How much should you worry about hurricanes, nuclear power, terrorism, mad cow disease, alligator attacks, or avian flu? And how much care should you take in avoiding risks associated with each? What, exactly, should you do to prevent the kinds of dangers that you face in ordinary life?

In answering questions of this kind, most people use what is called the availability heuristic.[158] Individuals tend to assess the probability of an event frequency based on the ease with which they recall examples relevant to it. Thus, often overestimating the possibility of that event to happen and underestimating another actually more frequent one. For example, people often

---

[156] Tversky & Kahneman, 1983.

[157] "*Tow W is a graduate student at the main university in your state. Please rank the following nine fields of graduate specialisation in order of the likelihood that Tow W is now a student in each of these fields. Use 1 for the most likely, 9 for the least likely: Business administration; 2) Computer science; 3) Engineering; 4) Humanities and education; 5) Law; 6) Medicine; 7) Library science; 8) Physical and life sciences; 9) Social science and social work.*" See Kahneman, 2017.

[158] Thaler & Sunstein, 2009, p. 25.

tend to overestimate the incidence of events causing vivid and emotional deaths, such as hurricanes or earthquakes, while underestimating the likelihood of occurrence of less vivid but statistically significant events such as deaths caused by asthma attacks. Another example is that fatal car crash is a more likely event then an airline crash. Nonetheless, the fear of death due to a plane crash is taken more seriously even though driving on roads leads to far more accidental deaths. To be precise, the probability of being involved in an air crash is only 1 in 11 million which staggers against a 1 in 5000 chance of a road accident.[159]

c. Anchoring

Tversky and Kahneman in another experiment manipulated a wheel of fortune.

This wheel was numbered from 0 to 100, but was designed to stop exclusively on 10 or 65. After implementing this modification, they would ask the students to spin the wheel and then they were asked to write down the number at which the wheel stopped (of course it was always 10 or 65).

Next, they had to answer two questions: whether they thought the percentage of African nations among UN members was higher or lower than the number resulting from spinning the wheel and what the actual percentage of African nations within the UN was.

It is important to note that the spin of the wheel of fortune could not provide any useful information to answer these questions. In theory, participants should have simply ignored the result obtained from the wheel and answered according to their own knowledge. However, the experiment showed that participants did not ignore this number; rather, the average estimates of students who observed the numbers 10 and 65 were 25% and 45%, respectively.

This phenomenon is known as the "anchoring effect." [160]

It is a heuristic that comes into play when people keep in mind a specific value for an unknown quantity before estimating it. As a result, estimates are biased, keeping close to the number that was considered.[161]

---

[159] Tyagi, 2015.

[160] Kahneman, 2017.

[161] This heuristic also happens when this heuristic also happens when we have to buy something and the price, we are willing to pay is anchored to the listing price.

2.6 Beyond behavioral economics: some cognitive biases that influence our daily lives

At the conclusion of this second chapter and after this brief exploration of human cognitive processes and decision-making patterns, it becomes crucial to answer for clarity the questions posed at the beginning of the chapter.

How do people make decisions? Which are the factors that influence their choices? When a person has to decide, is he or she able not to be influenced by external factors? Can weigh a choice and decide rationally?

As recent insights from behavioral economics, highlighted in the preceding sections, reveal, individuals often deviate from the behavior of the idealized, rational "Homo Oeconomicus" presented by Neoclassical Economics.

This reality has a significant impact on the notion of human rationality.

Historically, the prevailing idea was that humans were rational agents, unaffected by external influences. However, this assumption has been dismantled by the findings gathered from behavioral economics and psychology.

Humans have a limited rationality mainly caused by the presence of cognitive and behavioral biases, essentially distortions in decision making. Decades of research have identified a multitude of recurring cognitive biases, but below I will outline some of the most prevalent and significant ones.

2.6.1 Endowment effect

The endowment effect was first identified in the 1970s by economist Richard Thaler and manifests itself in the way that people often demand a much higher price to give up ownership of an object than they would be willing to pay to buy it.

In other words, there is a higher valuation associated with owning an object than its objective value in the marketplace.

In a research experiment conducted by Thaler and Kahneman on the endowment effect, college students were randomly assigned to one of three conditions: seller, buyer, and chooser.[162]

[162] Kahneman et al., 1991.

The sellers were initially given a university mug and then asked at what price, between 0 and 9 dollars, they would be willing to sell it.

Buyers were asked if they would like to purchase the mug at a price within that range.

The selectors had the option of choosing between a cup and the cash amount, at each price.

The results of the experiment revealed that the sellers, who already owned the mugs, attributed twice the median value to the mugs than the other groups to give them up.

These results were also confirmed through other experiments in which factors such as the gender of the participants, the types and combinations of goods, and the sums of money involved, which ranged from small amounts to more significant amounts. In each case, people demonstrated evaluations and preferences that varied systematically and substantially according to the initial reference point and the direction of the exchange.

2.6.2 Status quo bias

Status quo bias refers to the psychological phenomenon where people tend to prefer things to stay the same or remain unchanged.

This bias can manifest in various aspects of decision-making and behavior, including personal choices, public policy, and social attitudes. Individuals may exhibit status quo bias when making choices between maintaining the current situation and adopting a new course of action, even if the new option might be objectively better.

Status quo bias can be influenced by factors such as fear of uncertainty, loss aversion, cognitive effort required for change, and the familiarity of the current state. As a result, this bias can sometimes lead to suboptimal choices because people might resist change, even if that change could lead to positive outcomes in the long run (and so against their best interests).[163]

[163] Samuelson and Zeckhauser, 1988.

2.6.3 Confirmation bias

Confirmation bias is a cognitive bias where individuals tend to seek out or interpret the evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand.[164] In other words, people have a natural tendency to favor information that supports what they already think and ignore or downplay information that contradicts their beliefs. This bias can occur at both conscious and unconscious levels and can also lead to tragic results, as one is incapable of knowing when to change one's opinion, or way of behaving: think, for example, of an investor who ignores signs that his or her investment strategy is not the optimal one, continuing to trust his or her intuitions.

2.6.4 Overconfidence bias

Overconfidence bias, also known as the overconfidence effect, is a tendency for people to favor information that confirms their preconceptions or hypothesis regardless of whether the information is true.[165] This bias can lead people to believe that they are more skilled, competent, or accurate than they actually are, which in turn can impact decision-making and behavior.

A study of a sample of students asked them to estimate the expected time, at best and worst, to complete their dissertation. On average, students reported that they believed they could complete the paper in thirty-three days at best and forty-eight days at worst. In reality, the average number of days was found to be fifty-five days.[166] Also, Dan Ariely conducted an interesting experiment on students' ability to evaluate their own abilities and based on that manage deadlines. He gave three different classes of his students a set of drills to complete, to the first group he gave complete flexibility in setting delivery times, to the second group he mandated completion by the end of the semester, and finally to the third class he instituted undeferrable deadlines. Those who were given complete choice of delivery timeframes performed the worst, while the third group performed the best. In the absence of deadlines set in advance, students tended to procrastinate, achieving lower results.[167]

---

[164] Nickerson, 1998.

[165] Cambridge Dictionary. (n.d.). Overconfidence bias. In *Cambridge Dictionary online*. Last accessed August 18, 2023, https://dictionary.cambridge.org/dictionary/english/overconfidence.

[166] Krugman & Wells, 2013.

[167] Ariely, 2010.

2.6.5 Self-serving bias

The self-serving bias is a cognitive bias where individuals tend to take personal responsibility for their desirable outcomes yet externalize responsibility for their undesirable outcomes.

In other words, people tend to take credit for positive outcomes but distance themselves from negative outcomes by attributing them to external circumstances, other people, or bad luck. This bias can serve to protect one's self-esteem and maintain a positive self-image.

An example very dear to students is that of an exam. If one passes an exam, it is due to intellectual gifts, great study done and the ability to meet and exceed goals. Conversely, if the exam went badly, people tend to attribute the failure to external factors such as lack of time or bad luck.

2.6.6 Optimism bias

Optimism bias is a cognitive bias where individuals tend to believe that they are less likely to experience negative events and more likely to experience positive events compared to others. In other words, people often have an optimistic outlook about their future, expecting good things to happen to them while downplaying the potential for negative outcomes.[168]

For example, by most smokers, the chance of getting cancer is underestimated; in general, the probability of dying in a traffic accident is also underestimated, but at the same time there is a tendency to overestimate our chances of a working career or life expectancy. Another glaring example concerns the expectations inherent in the chances of successful marriages. The percentage of them that end in divorce stands at 40 percent, but by polling a couple that has married, it is likely that their estimate of the chances of such an occurrence is close to zero.

2.6.7 Social Desirability Bias

Social desirability bias is a cognitive bias that occurs when individuals respond to surveys, questionnaires, or interviews in a way that they believe will make them appear more socially acceptable or desirable, rather than providing honest or accurate responses.

[168] Weinstein, 1989.

2.6.8 Present bias

People are often faced with choices involving different time periods, that is, affecting not only the present, but also the future. For an individual acting rationally, such as a "Homo Oeconomicus," this process does not present great obstacles. Based on his or her own preferences and rational calculations that include factors such as estimates of future income and life expectancy, the individual will make an informed decision. According to the Neoclassical Economic model, a person's preferences are assumed to remain consistent over time, unchanged as prospects change.

For example, if you were asked on Monday to choose between receiving €10 on Saturday or waiting a day and getting €15 on Sunday, which option would you choose? What if, on the other hand, it was already Saturday and you were again asked the same choice: immediate €10 or €15 tomorrow, what would be your answer?[169]

Consistently and rationally, if you had decided on Monday to wait an extra day to receive €15 on Sunday, your decision should not change if the same choice were presented to you the following Saturday. You would be expected to maintain control over your decision and, even when faced with the possibility of receiving a smaller amount immediately (such as the €10), not give in to temptation.

However, it has been shown that human beings often seem to struggle with the concept of self-control. They may accumulate excessive debt, overindulge in nutrition, or fall into addiction to smoking. The most complex decisions for people often involve sacrificing immediate benefits for future benefits. For example, adopting a healthier diet today or quitting smoking requires a considerable commitment in terms of immediate satisfaction, while the benefits only manifest themselves in the future. This phenomenon is often described as "present bias," in which people tend to seek instant gratification and give disproportionate weight to the present. In other words, there is a strong inclination to prefer immediate benefits and postpone costs to prospects and this makes people dynamically inconsistent in their choices over time.

[169] Krugman et Wells, 2018.

CHAPTER 3.

Mastering the art of decision-making: harnessing the potential of AI in the debiasing process

Summary: **3.1** The illusion of being the architect of our own destiny — **3.2** The standard behavioral change tools — **3.3** Debiasing through nudge: the libertarian paternalism approach — **3.4** "Digital nudging" against us: the blind spots of human choices in the face of technology risks — **3.5** A cutting-edge debiasing tool: AI as the new Virgil in "digital hell" — **3.6** Why might AI as debiasing tool be the best solution? Machine rationality to deal with human irrationality

3.1 The illusion of being the architect of our own destiny

We live in a time when technology completely absorbs us. Physical reality and digital reality are often difficult to distinguish from each other.

This trend began about 30 years ago with the invention of the Internet.[170] Since then, there have been significant and systemic changes all over the world.

Electronic devices are becoming more and more powerful and more within reach, services are becoming more and more digital, homes and appliances are becoming smart, one can take courses or classes remotely, one can work from home... everything has changed.

Innovation obviously creates wonderful things: bringing family together, creating friendships and loves, finding organ donors, helping to create an increasingly informed world, creating jobs and new opportunities for personal growth.

However, there is also a downside.

How many times during our daily lives do some of us (the vast majority) ask ourselves questions like: why am I wasting time playing with my phone instead of studying or working? Why am I spending too much time on social networks? Why am I sitting all the time and not doing enough physical activity? Why do I prefer to watch a TV series or a movie instead of going outside?

Although it could be difficult to accept, the answer to these questions is and always will be one: it is your choice. And your choice is the result of a decision-making process.

---

[170] See *Sections 1.1* and *1.2*.

Decision making is a response to problems that include choices from among a set of alternatives. We are all decision makers. Everyday.

Only few decisions are singularly important, but numerous small decisions collectively have significant consequences that matter. If you regularly make mediocre decisions, you may never accomplish the things that are important to you, your family, or your career. We all learn decision making by doing it.[171]

Neoclassical Economic theory suggests that humans when they make decisions are rational and consistently do whatever they take to achieve their goal and to maximize their utility, without being influenced by other factors.[172]

But if a person's actions or omissions always depend solely on him or herself, and if we really are rational and skilled calculators, why do we often make wrong decisions? Why is it so difficult to change one's behavior?

These questions have been the starting point for countless research efforts that have flowed into the field of psychology and behavioral economics over the past six decades. What has been highlighted is that, because of their cognitive limitations, humans are not as rational as they seem and often, they make wrong choices and decisions.[173]

Heuristics, commonly defined as simple "rules of thumb" that people use to ease their cognitive load in making judgments or decisions,[174] can influence decision-making positively or negatively. Sometimes, they can be helpful in making simple, recurrent decisions by reducing the amount of information to be processed so people can focus on differentiated factors.[175] Other times, heuristic thinking can result in cognitive biases and introduce systematic errors when making complex judgments or decisions.[176] In such situations, common heuristics such as representativeness, availability, and anchoring, affect the evaluation of our alternatives, often leading to suboptimal decisions and bias.[177]

---

171 Keeney, 2004, p. 193.

172 See *Sections 2.2 et seq.*

173 See *Sections 2.4* and *2.5*.

174 Hutchinson and Gigerenzer, 2005, p. 98.

175 Evans, 2006.

176 Tversky and Kahneman, 1974 (see also *Section 2.5 et seq.*).

177 See *Section 2.5.4*.

But then what solutions can there be to face limited human rationality? How can we limit the "wrong" choices that each of us makes every day? How can people be helped to make decisions in line with their goals and well-being?

Once a behavioral phenomenon subject to heuristics or bias has been identified, research should be promoted to question its robustness and find new techniques to eliminate it.[178] This process in engineering is known as "destructive testing";[179] but when it targets a biased behavioral phenomenon is known as "debiasing."

Debiasing can be defined as that process through which one tends to reduce or mitigate biases, particularly in judgment and decision making. Thus, the ultimate goal of this process is to ensure that people make better decisions for their own well-being.

At this point another question arises: who is responsible for the development, promotion, and implementation of debiasing techniques?

The desirable answer would be the individual himself. However, it has already been pointed out that people have limited rationality and, for this reason, cannot always recognize their own biases and thus debiasing themselves. If we were to use the words of Kahneman and Tversky, we would have to say: System 2 is not always able to correct System 1.

Therefore, it has been necessary to find others who are able to help people make decisions and choices that maximize their well-being.

For centuries policymakers have played a role in the debiasing process because through adopting laws and incentives or providing more information to citizens, they try to direct human behavior toward optimal choices.[180] Today, however, this role is also assumed by those who, in the words of Thaler and Sunstein, choose the architecture of a choice.[181]

But are these debiasing techniques also sufficient for the risks the technology creates? Are we able to deal with them appropriately?

The following pages will briefly describe the standard tools used by policymakers; the recent debiasing technique of nudging; and finally, an innovative solution will be proposed to deal with the risks that technology creates.

---

[178] Fischoff, 1982.

[179] In destructive testing (or destructive physical analysis, DPA) tests are carried out to the specimen's failure, in order to understand a specimen's performance or material behavior under different loads (for further information see e.g., here: https://en.wikipedia.org/wiki/Destructive_testing).

[180] See *Section 3.2.*

[181] See *Section 3.3.*

3.2 The standard behavioral change tools

In the face of limited human rationality and choices that do not always ensure that the individual maximizes his or her own well-being, policymakers, through certain techniques that I refer to hereinafter as "standards," aim to steer citizens toward more informed and correct choices.

3.1.1 Regulation

If the citizen shows that he is unable to pursue his own good, then he should be helped and not left alone and abandoned[182] and the oldest tool to direct human behavior toward optimal choices is the law.

A state, through its legislative body, issues legal rules governing the relationships of individuals. These norms are coercive, mandatory for all members of the state. Among the means of guiding human behavior, the law is surely the most paternalistic because it overrides a person's own wishes (autonomy) in pursuit of his or her best interests.[183] Compulsory seat belt wearing in cars, helmet wearing on motorcycles, measures against fires or pollution, a ban on the sale of cigarettes to people under the age of 18[184] are just a few examples of the state's coercive paternalistic inclinations.

The question of when, if ever, a measure of paternalism may be justified remains one of the most contested in ethics. For example, the philosopher John Stuart Mill argued that intervention was justified only when trying to prevent a person from causing harm to others, not to himself.[185]

However, forms of paternalism may be justified when a person, because of his or her bounded rationality, lacks the capacity to make decisions for him or herself. It is indeed possible to prohibit certain options, especially when certain behaviors are harmful and expose people to high risks.

In these cases, the paternalism is direct and does not make subtle analyses of the actual effects of these prohibitions. It does not, for example, ask whether prohibitions, with relative criminal consequences, do not produce perverse effects to the contrary (increased use, the cost of

---

[182] Viale, 2018, p. 20.

[183] Oxford Reference. (n.d.). Paternalism. In *Oxford Reference online*. Last accessed August 26, 2023, https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100310127#:~:text=n.,his%20or%20her%20best%20interests.

[184] According to the WHO guidance, legislative interventions are one of the most effective tools for achieving positive effects on both nonsmokers and smokers. For further information see e.g., here: https://www.who.int/health-topics/tobacco#tab=tab_1.

[185] Mill, 1974, p. 17.

repression, and the creation of illegal economies) or whether instead it would be more economical and effective to use lighter forms of paternalism that appeal to our automatisms based on heuristics or social rules.[186]

3.1.2 Incentives

Behavioral research is clear in showing that offering a reward for a behavior can increase its frequency. If cost is a barrier to the target behavior, then offering an incentive can reduce the difficulty of the action.[187] Incentives can take both negative and positive forms.

Negative incentives include mechanisms designed to discourage specific behaviors. An example of a negative incentive is cigarette taxes.[188] That of smoking is a typical problem of the so-called battle between the current self and the older self. When it occurs in fact some people find it difficult to give voice to their future selves in current practices. For this reason, public policies help people with incentives, or as in this case disincentives, to strengthen the deliberative impact of the future self. Imposing significant taxes on cigarettes has in fact turned out in some countries to be an indirect method of reducing cigarette consumption.[189]

On the other hand, positive incentives encompass approaches like grants, discounted prices, subsidies, and even symbolic rewards, all designed to elevate the likelihood of desired behaviors.[190]

Not surprisingly, research has shown that incentives can exert a powerful influence on behavior and the larger the incentive or disincentive, the greater the amount of behavioral change.

Although incentives can produce large changes in behavior, they also come with several serious side effects.

The first is durability. Repetitive behaviors that are changed through incentives typically revert back once the incentive is removed.[191]

---

[186] Viale, 2018, p. 20.

[187] McKenzie-Mohr and Schultz, 2014.

[188] Ulen, 2013.

[189] In fact, it has been shown that the demand for cigarettes is somewhat price-elastic and that taxes on cigarettes can reduce cigarette consumption.

[190] Burgess and Ratto, 2003.

[191] Schultz and Kaiser, 2012.

A second limitation is the specificity of the change, and behaviors that are changed through incentives generally do not spill over into other domains. For example, offering a large incentive for the purchase of energy-efficient lightbulbs will generally not spill over into other energy-efficiency behaviors, like using a switchable power strip or turning off computers when leaving the office.[192]

Due to the side effects associated with the incentives, they should be used sparingly, and they typically work best in instances where cost operates as a barrier to the action.

3.1.3 Information and education

The third standard tool of behavioral intervention that has already been tried throughout history is to raise people's awareness and educate them to help them make decisions that will bring them well-being.

Over the years, governments have traditionally developed large campaigns to promote desirable behaviors, such as physical activity, or to reduce harmful behaviors, such as smoking. However, educating people and providing them with information about the risks and benefits of certain behaviors may not be enough. These campaigns have often produced little or no results. For example, the European Commission in 2009 (so immediately after the 2008 financial crisis), had identified the retail investment services market as one of the worst performing markets for consumers. This market had evolved and had become increasingly complex to be dealt with effectively by the consumer, who generally lacked adequate financial education and had little information. The results of various empirical tests confirmed that subjects, in their investment choices, manifested several biases. Therefore, it was found that Neoclassical pillars of consumer protection such as completeness of information and disclosure had opposite effects to the legislator's intentions. The exaggerated amount of information increased complexity and generated cognitive overloading, with obvious suboptimal performance, reasoning, and decision-making consequences.[193]

This is not to say that education or information should not be provided or used as a tool but maybe they should be considered only complementary to other tools because people often act automatically and are influenced by contextual factors that cannot be overcome using information alone.

---

[192] McKenzie-Mohr and Schultz, 2014.

[193] Viale, 2018, pp. 235-236.

3.3 Debiasing through nudge: the libertarian paternalism approach

The Homo Oeconomicus model, as highlighted in Section 2.5, has been challenged by the research program of Kahneman and Tversky. They carried out work mapping the anomalies of human judgment and decision-making against the formal norms of probability and utility. Their results showed that "real man" is very far from the model idealized by Neoclassical economists.

Human being is not an "econ," he is not like Spock from Star Trek[194]; rather, he is "human," he is more like Homer Simpson.[195] He has limited rationality, has many weaknesses in will, memory, attention, is very lazy, has a very limited capacity for calculation, is subject to contradictions, errors, and emotional perturbations.

If we were all like Homo Oeconomicus, capable of rational self-regulation, judging, estimating, and predicting events in a statistically correct way, deciding our expected utility in an optimal way, then there would be no need for state intervention. If Neoclassical utility theory were true, the citizen would be justified in being free and autonomous in his choices.[196]

Unfortunately, however, human beings are not "econs" but are more "humans." Therefore, complete citizen autonomy is not justifiable.

But then what kind of intervention can the state implement on his behalf? Above all, what space for intervention and what kind of intervention is it justified in taking?

From these questions began the work of economist Richard Thaler and jurist Cass Sunstein. Their goal was to find a new way to help citizens in their choices without using the traditional tools of regulation, incentives, and information-education. In their very famous book "Nudge," the authors propose a new way to influence people's choices to ensure that they make better decisions: the so-called "behavioral regulation" or "nudging."[197]

The earliest formulation of the nudge concept did not use that term. Instead, in their groundbreaking 2003 article, Sunstein and Thaler advocated for "libertarian paternalism".

They argued that, since deviations from rationality lead people to make suboptimal choices, appropriate interventions can make individuals better off by using an approach that: "*preserves*

---

[194] Star Trek is an American science fiction media franchise created by Gene Roddenberry, which began with the eponymous 1960s television series and became a worldwide pop-culture phenomenon. One of the main characters, Spock, is not a prisoner of emotions, has an iron will, is always attentive and present in analyzing choice problems, has a great memory and calculation ability. For further information see e.g., here: https://en.wikipedia.org/wiki/Star_Trek.

[195] The concepts and comparisons between Spock and "econ" and Homer Simpson and "human" are taken from the book: "Nudge: Improving Decisions about Health, Wealth, and Happiness" (Thaler and Sunstein, 2009).

[196] Viale, 2018, p. 151.

[197] Thaler and Sunstein, 2009.

freedom of choice but... encourages both private and public institutions to steer people in directions that will promote their own welfare."[198]

Hence, what renders an intervention libertarian paternalistic is the unique combination of its private welfare goal and its choice-preserving tools. As paternalism it aims to offset the irrational and self-defeating tendencies of citizens by "gently nudging" them to make rational decisions for their own good. As libertarianism, on the other hand, it aims to give the final say to the outcome of the deliberative and conscious processes of the citizen who can always oppose the gentle nudge.[199]

So, a nudge could be defined as "*any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives.*"[200]

Looking at this definition four elements can be distinguished:

i. the intervention in the context in which people live ("*any aspect of the choice architecture*");

ii. the influence of people's behaviors ("*altering people's behavior*");

iii. the systematic nature of biased choices that provide the opportunity to set interventions that have predictable outcomes ("*in a predictable way*");

iv. the absence of coercion, punishments or incentives ("*without forbidding any options or significantly changing their economic incentives*").

The works of Kahneman, Tversky, Thaler, Sunstein and others as well as paving the way for a new research agenda, has also shifted the attention of policymakers. Indeed, the study of human decision-making and the impact of social contexts, emotions, and other relevant factors on it has become crucial in public policy development. For over a decade, governments and other organizations have been increasingly turning to these "soft" behavioral interventions to achieve their policy goals and to promote private or public welfare. The general public but also, in several countries, policymakers have become increasingly aware of key insights from behavioral sciences.[201]

---

198 Thaler and Sunstein, 2003.

199 Viale, 2018, p. 24.

200 Thaler & Sunstein, 2008, p. 6.

201 Alemanno and Sibony, 2015, p. 2.

By now, nudges have already been implemented in nearly every major policy domain that concerns individual behavior, from health, safety, education, and finance through environmental protection and tax compliance, to public service delivery and more.[202]

In 2010, the United Kingdom established a Behavioral Insights Team, which now has an extensive track record.[203] In 2014, the United States created a Social and Behavioral Sciences Team ("SBST") of its own, and President Obama formally embraced the behavioral approach with an important Executive Order in 2015.[204] The goal was to use the office to filter all new bills and introduce some behavioral sensitivity into their final drafting.[205] Also, Australia and Germany established their own behavioral science teams in 2015.[206]

Uses of behavioral science, with particular emphasis on nudges, have attracted increasing interest all over the world, and perhaps especially in Europe.[207]

Regulators like nudges for a number of related reasons: first, nudges are based on a realistic view of human behavior that is intuitively appealing; second and related, policy makers may believe that nudges are politically more feasible than their alternatives; third, the great variety of behavioral tools means that they are more versatile than traditional instruments; fourth, and finally, nudges tend to entail relatively low implementation costs, imposing less strain on limited budgets and thereby appearing to be more efficient or cost-effective than competing traditional instruments.[208]

Thus, while traditional regulatory instruments affect behavior by imposing constraints (as mandates or bans do), using economic incentives (as in the case of taxes or subsidies), or disclosing unavailable or costly information, nudges rely on "softer" behavioral guidance tools.[209]

---

[202] Tor, 2022.

[203] Halpern, 2015.

[204] Obama, B. (2015). *Executive order — Using behavioral science insights to better serve the American people* (Executive Order 13707). Washington, DC: The White House. Retrieved from the White House: (available here: https://sbst.gov/download/Executive%20Order%2013707%20Implementation%20Guidance.pdf).

[205] Viale, 2018, p. 22.

[206] Sunstein, C. R. (2016). The council of psychological advisers. *Annual Review of Psychology*, *67*, 713-737.

[207] Whitehead et al., 2014.

[208] Tor, 2022, p. 236.

[209] Thaler and Sunstein, 2008.

For example, to counter some systematic tendencies of the human mind such as inertia (or the so-called status quo bias), present bias, and loss aversion[210] one can provide nudging interventions designed as "default options."

A "default rule" is a starting rule that will be in effect unless the party or parties facing the rule agree to change it.[211]

The paradigmatic initiative for its originality and realized success is the "Save More Tomorrow (SMT)" program by Richard Thaler and Shlomo Benartzi.[212] Both started with a question: how do we overcome the resistance and short-sightedness of American workers that lead them not to save and not to think about how to deal with the future when they are retired?[213] In fact, human beings often fail to represent themselves as individuals who will change and age as the years go by, but will always be in anthological continuity with the individual now. This lack of connection with the future is clearly at the root of many errors in perspective and behavior in the present, such as neglecting the need to save money and to provide social security protections.[214] U.S. workers, in fact, tended to disregard their pensions because they were averse to withdrawing amounts from their paychecks.

Thaler and Benartzi developed a new approach to address this issue without resorting to coercion.

Instead of asking workers to immediately enroll in a pension program, they created a mechanism in which workers were asked to enroll in a program that would be activated only in the near future. This way, they would not receive a decrease in their salary (which could have led them to reject the offer). After initial enrollment, they were automatically confirmed in the program unless they opted out (which was difficult due to status quo bias). Finally, only the amounts related to salary increases were deducted from their salaries each month, thus ensuring a constant revenue stream.

This mechanism helped reduce the loss aversion that results from noticing a decrease in salary, even though this is mainly due to inflation and is not perceived due to the monetary illusion, which leads the individual to consider only the nominal salary.[215]

---

[210] For both biases, see *Section 2.6*.

[211] Ulen, 2013, p. 10.

[212] Thaler and Benartzi, 2004.

[213] Viale, 2018, p. 161.

[214] Viale, 2018, pp. 188-189

[215] Viale, 2018, p. 161.

Another technique used in various states and leveraging the default option concerns organ donation after death.[216]

In Austria, for example, if a person makes no decision about consenting to the donation of his organs, they will be excised and transplanted into another person upon death. In contrast in Germany, if a citizen makes no decision about donations, upon his death his organs will not be explanted and transplanted. These are two opposite examples of default options regarding organ donation. The first, also referred to as "opt out," states that if you do not want to be a donor, you must declare it. The second, also called "opt in," has an opposite rule. If you want to donate, you must declare it.

The result of these two opposite options in Austria and Germany is that in the former country organ donations are approximately more than 90 percent, while in Germany less than 13 percent.[217]

The default option, however, is not the only nudge technique that has been introduced by states. Indeed, some authors speak of "cognitive" paternalism when referring to those nudges that attempt to appeal to the analytical, attention and reasoning capabilities of the human mind (the so-called System 2).

As highlighted in Section 2.5.4, humans often use heuristics when faced with complex tasks. In some cases, this can lead to correct decisions, in others to incorrect or biased decisions. Therefore, when faced with the complexity of a problem and limited human rationality, nudges can be useful to simplify and make more obvious information to strengthen human reasoning and judgment skills and enable them to make better and more informed decisions.

For example, to educate people to eat a balanced diet, the U.S. government adopted an initiative called "Choose My Plate."

Indeed, they had realized that the traditional food pyramid figure, in which the different amounts of necessary nutrients were illustrated within a pyramid in descending order, while scientifically correct was also unintuitive.

---

[216] Johnson and Goldstein, 2004.

[217] Thaler and Sunstein, 2008, pp. 178-179.

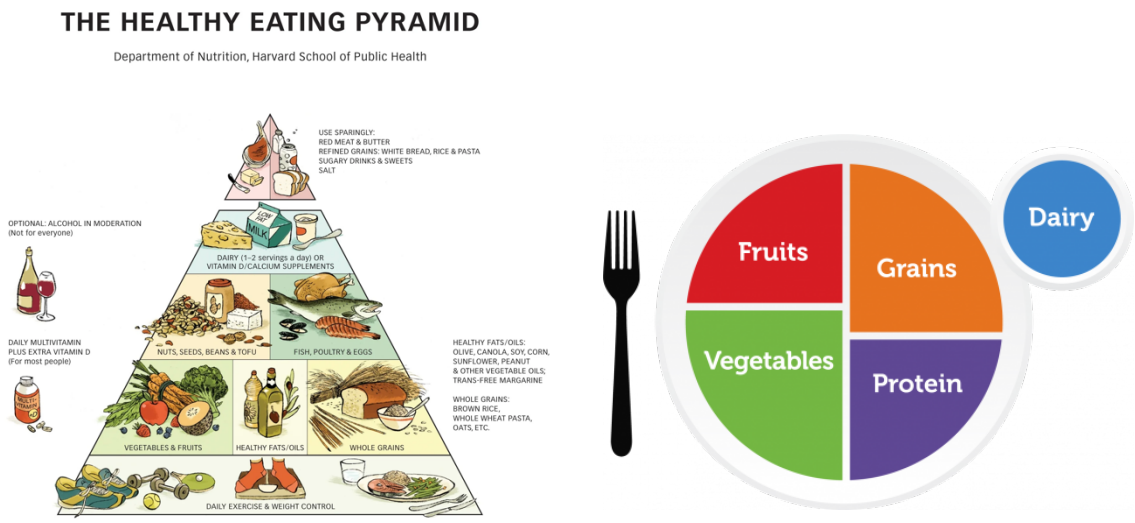

Figure 13. In the figure above are shown the "Food Pyramid"[218] and "MyPlate"[219] initiative.

Therefore, under the Obama administration the pyramid was replaced by a new image: a plate on which the daily proportions of the different nutrients were represented: fruits, vegetables, grains, and proteins. The information became clearer, more user-friendly, and easier to retrieve from memory when choosing or consuming food without resorting to coercion.

A further example of this kind is represented by the EU energy labelling.

Figure 14. Example of energy labellling

[218] Image source here: https://www.hsph.harvard.edu/nutritionsource/healthy-eating-pyramid/.

[219] See here: https://www.myplate.gov.

Eco-design legislation sets common EU wide minimum standards to eliminate the least performing products from the market. The energy labels provide a clear and simple indication of the energy efficiency and other key features of products at the point of purchase. This makes it easier for consumers to save money on their household energy bills and contribute to reducing greenhouse gas emissions across the EU.[220]

Simplification as a nudge technique, however, can be achieved not only through "visual effects," but also through simplified language.

For example, Regulation (EU) No 1286/2014 has introduced new disclosure requirements on entities that, in various capacities, are included in the process of creating and offering packaged retail and insurance-based investment products ("PRIIPs") on the market. The main idea of the PRIIPs Regulation is to ensure that those complex products are transparent and are easy to compare with others product. This goal was translated into the "KID": the key investor document. It is a document that can consist of a maximum of two pages (maximum three-sided A4 when printed), must have a standardized format, use visual indicators that make it easy for the investor to understand the costs he or she must incur, the risks he or she might incur, and the performance of the product he or she is buying.[221]

These and other provisions[222]were adopted by the European legislature to try to cope with the limited rationality of the retail investor during the investment phase.

[220] For further information see here: https://commission.europa.eu/energy-climate-change-environment/standards-tools-and-labels/products-labelling-rules-and-requirements/energy-label-and-ecodesign/about_en.

[221] See article 8 here: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R1286.

[222] Also, the Prospectus Regulation (EU) 2017/1129 introduced the Prospectus and the Prospectus Summary to provide useful information especially for retail investors (see here: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1129).

3.4 "Digital nudging" against us: the blind spots of human choices in the face of technology risks

In the contemporary technological landscape, a substantial portion of individuals' time is devoted to engaging with advanced computer systems. These systems range from personal computers and smartphones to numerous other digital gadgets that seamlessly integrate into daily routines both at home and work.

Technology positively revolutionizes daily life, especially through mobile device apps, which have the fundamental task of simplifying and enriching people's lives. Today, you can find an app for virtually any purpose-from photography to painting, from reading the news to translating, to finding your way around cities, and much more.

In addition, technology has given rise to new social ties and changes in relational dynamics. Communication and interpersonal relationships are increasingly developed through screens, thanks to networks that enable instant contact and erase physical distances. In this mobile space, everything is constantly evolving, and speed is crucial while distance becomes negligible.

The screen of a computer or cell phone is transformed into a stage for human actions and relationships, becoming the main environment in which people live and share their stories. Technology is no longer just a tool; it has also become a new framework for critical thinking that enables people to overcome cultural barriers that, as recently as 50 years ago, seemed insurmountable.

The increasing use of digital technologies in broad areas of our private and professional lives has a fundamental consequence: it means that people often make important decisions in digital choice environments.[223]

This brings with it risks and concerns. But why?

Because research in psychology has demonstrated that, because of their cognitive limitations, people act in boundedly rational ways,[224] and various heuristics and biases influence their decision-making.[225]

Nudges, as described in the previous section, attempt either to counter or to encourage the use of heuristics by altering the choice environment to change people's behavior. Commonly used

---

[223] Weinmann et al., 2016.

[224] Simon, 1972. See also *Section 2.4*.

[225] Kahneman, 2003, 2011, 2017. See also *Section 2.5* and *2.6*.

nudges include for example setting defaults or providing simpler and more understandable information. In various situations, the designers of the choice environments (the "choice architects"[226]) attempt to influence people's choices.[227]

All nudging interventions are part of the so-called "behavioral regulation." To achieve their policy goals, both governments and private organizations are increasingly turning to nudges in an attempt to shape individual behavior in most major policy domains including health, safety, education, finance, environmental protection, tax compliance, public service delivery and more.[228]

Although nudges should be used to help people make better choices,[229] in the digital environment so-called "digital nudges" do not always pursue this goal.

Digital nudging refers to all those features of virtual environments that condition people's behaviors in a predictable way.[230] Digital nudging is therefore used by web designers and developers in designing app and site interfaces to shape the behavior of the individuals they target.

The advantages of digital nudges are not limited to their potentially rapid response times or their access to current information. Additionally, as technology continues to advance and exceed expectations, digital nudges can use sophisticated algorithms to detect unique and distinctive behaviors of each individual and through machine learning and AI systems provide personalized interaction with the individual.

Studies show, for instance, how data on Facebook "likes" can predict different personal characteristics, such as demographics or even personality traits (e.g., extraversion or openness), with some accuracy.[231] Such predictions, in turn, can form the basis of more effective behavioral interventions that are adapted to the identified characteristics.[232]

---

[226] Thaler and Sunstein, 2008.

[227] For example, many organizations encourage people to engage in socially responsible behaviors, such as leading a healthy life, reducing waste or energy consumption, and planning for retirement. Likewise, many non-governmental organizations attempt to encourage people to donate funds or participate in charitable activities.

[228] Tor, 2022.

[229] Thaler and Sunstein, 2008.

[230] See e.g., here: https://pxritaly.com/it/blog/digital-nudging/#:~:text=Il%20Digital%20Nudging%20consiste%20nel,è%20a%20disposizione%20di%20tutti.

[231] Kosinski et al., 2013.

[232] Matz et al., 2017.

Digital nudges are distinct from their offline counterparts in their deployment of software and its user-interface design elements and are an increasingly pervasive feature of online environments.[233]

Although digital nudges share many features of their offline predecessors, they merit particular attention and analysis for two important reasons: first, the ubiquity of digital nudging across online platforms, social networks, other applications, and electronic devices makes it a nearly unavoidable feature of daily life, thereby bringing into sharper relief the promise and perils of nudges more generally. Second and more importantly, digital nudging raises unique issues compared to offline nudging, because of its potentially greater potency (e.g., due to the possibility of dynamic, personalized interventions), the opacity of the technological and behavioral mechanisms through which it shapes behavior (as when using AI and machine learning), and the central role of independent private actors (most notably, private intermediaries such as internet platforms) in its implementation.[234]

In the current technological environment, people spend much of their time interacting with sophisticated computer systems and this increases both the opportunities for and the incidence of digital nudging.[235]

For example, digital nudges have a huge impact in the communications industry.

Indeed, communication is one of the activities we do most frequently through e-mail, chat, social networks, etc., and it has played an extremely relevant role in our survival.[236] For this reason, we probably developed an evolutionary predisposition to communication that made this activity intrinsically rewarding, that is, accompanied by the release of dopamine neurotransmitters that result in pleasurable sensations.[237]

Digital services, knowing full well our need to communicate, exploit our "weakness" and use nudging techniques to make us spend as much time connected as possible.[238]

---

[233] Weinmann et al., 2016.

[234] Tor, 2023.

[235] *Ibid.*

[236] Throughout our evolutionary history, in fact, communicating and exchanging information with our peers was essential, for example, to understand the availability of food or any threats on the ground (Pasquinelli, 2012).

[237] Tamir and Mitchell, 2012.

[238] The result of this evolution, as highlighted in *Section 1.2*, is that we use electronic devices so much every day, and especially the use of social networks is increasing year by year. Of course, it is in the interest of digital platforms to increase the use of their services. In fact, even those social networks that are free (such as Facebook, Instagram, Snapchat, etc.) have huge earnings because advertisers pay them to have their ads shown to users. In fact, it is all run by an algorithm whose goal is to show users personalized content. In fact, we speak in these cases of an "attention economy."

For example, the "blue tick" in the instant messaging application (like WhatsApp) signals that a message has been received is one of the most glaring examples of these techniques.[239]

Figure 15. Example of "blue tick" in WhatsApp[240]

Through this expedient the one who sends the message can verify that it has been read. In turn, the receiver, by viewing the message, is aware that he or she is communicating to the other person that he or she has read the message and therefore feels obligated to respond quickly. Thus, it is a gimmick that exploits our propensity to communicate by adding information in the form of a graphic warning (the blue tick), which aims to create a kind of commitment on the part of the receiver and can manipulate our behavior.[241]

We incur digital nudges not only in digital platforms, but also in every website we use. One of the most glaring examples is the use of the default option by websites to accept cookies.[242]

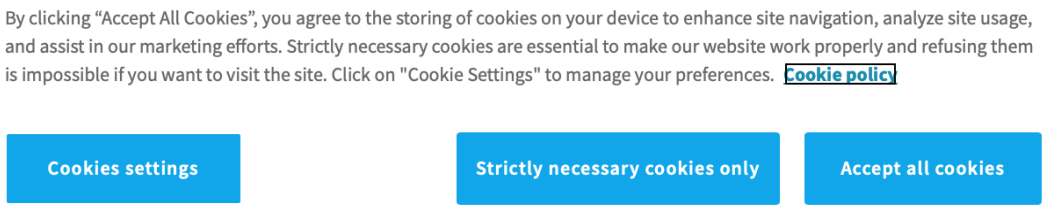

Figure 16. Example of a window menu asking us to accept cookies on a website

[239] Viale, 2018, pp. 217-219.

[240] Image source here: https://mobiletrans.wondershare.com/whatsapp/remove-blue-tick-from-whatsapp.html.

[241] Viale, 2018.

[242] HTTP cookies (also called web cookies, Internet cookies, browser cookies, or simply cookies) are small blocks of data created by a web server while a user is browsing a website and placed on the user's computer or other device by the user's web browser. Cookies are placed on the device used to access a website, and more than one cookie may be placed on a user's device during a session. For further information see e.g., here: https://en.wikipedia.org/wiki/HTTP_cookie.

Figure 17 and 18. Examples of a window menu asking us to accept cookies on a website

In fact, every time we use a web page a window, larger or smaller, appears asking us to accept them (and in some cases, if not accepted, the activity cannot be continued). Again, the architecture tends to favor choices that maximize the company's collection of information, this time at the expense of our privacy.[243] In fact, people rarely, because of their haste and inertia, have the patience to open the window menu that appears on the screen and select the settings on cookies.

Further examples of the use of digital nudging are smartphone devices, which often, once purchased, are set by default to automatically activate GPS navigation or when we sign a contract to download an application or book a flight. On these occasions, certain conditions, which are optional, are automatically pre-selected, such as sharing our data with third parties for marketing purposes.[244]

[243] Viale, 2018.

[244] See "privacy and data security issues" in *Section 1.7.1.*

In all these cases, to get out of the default situation, it is essential to activate our "System 2," and then pay attention to the contractual conditions, distinguishing between those that are required and those that are optional, and carefully selecting the options manually.

The power of influence of large technology companies through the application of "digital nudge" techniques extends to many other areas.

One obvious example is the way more and more people inform themselves through social media, such as the X app. These services use algorithms to select the content that is displayed on users' newsfeed, based on their digital actions and preferences.[245]

This approach carries several risks.

In the political arena, for example, if we interact primarily with content that reflects our preexisting beliefs, the algorithm will tend to present us with more and more similar content. This can reduce the diversity of opinions we are exposed to and limit our ability to examine different perspectives. Indeed, this phenomenon can eliminate contradiction and disincentivize critical reflection, as we mainly receive confirmations of our ideas. In the long run, this personalization and emphasis on confirmations of our beliefs could contribute to dependence to digital platforms, as constantly receiving positive confirmations of our opinions can be rewarding, which will increasingly stimulate the use of such services.

One of the greatest risks associated with digital nudges, as well as traditional offline nudges, is their difficulty of detection.

This aspect while it should not surprise us – since the very essence of the concept of nudge, and thus of libertarian paternalism, is to "gently nudge" toward a particular behavior without the individual being aware of it, in order to preserve his or her autonomy in decision-making and avoid a form of pure paternalism[246] – is not necessarily without danger.

---

[245] Such as "likes," comments, saved or shared content.

[246] Thaler and Sunstein, 2008.

3.5 A cutting-edge debiasing tool: AI as the new Virgil in "digital hell"

Our world has changed profoundly in the last 30. Human beings tend not to like change, it scares them. But we cannot prevent it from coming. Either we adapt to change, or we fall behind.

At the time of President Clinton, the possibility of regulating the Internet was considered. In those years, however, the Internet was seen more as a utility (it was very much related to the concept of communication) and for this reason it was preferred not to regulate it immediately so as not to curb its growth and encourage investment in it.

Over the years, however, it was understood that the Internet, and in general all the other digital services that had developed through it, were not merely a service but rather created a new "space."

Human beings, in fact, do not live on television, but watch it; they do not live on radio, but live on the Internet; they do not spend time on a newspaper, but spend it on Facebook or X app. In all these cases he is therefore in a space somewhere between online and offline life, what Floridi calls "onlife."[247]

So, the greatest innovation, the so-called "fourth revolution," was that we created a new space in which we live.[248]

This is the Digital World.

In this new "space", new technologies promise to vastly increase our economic productivity and bring information to our fingertips and improve our lives. Nevertheless, it has been pointed out in the previous section how new technologies, using "intelligent" systems, can exploit blind spots in our choices to affect our lives.

This may not please, but then what can be a solution?

The outdated solution of turning off a device or simply logging off no longer seems possible. We know, in fact, that being present in the Digital World is essential for professional and personal reasons. It is therefore necessary to assess the risks and try to curb them.

---

[247] Floridi, 2015.

[248] Floridi, 2014.

The law, the behavioral change tool par excellence,[249] in this area has sometimes proven to be an ineffective tool.

The jurisdictional issue is one of the main reasons, as actions and their consequences can be in completely different parts of the world. This means that even if a behavior is illegal, the chances of taking effective action may be so small that it is as if there were no legal consequences.[250]

Therefore, new approaches are needed.

Realizing the shortcomings of our rationality[251] and the ineffectiveness in some cases of standard behavioral change tools,[252] Thaler and Sunstein introduced the concept of "nudge" to improve our daily decisions about health, wealth, and happiness.[253]

Based on the same assumption, programmers of "intelligent" systems (the websites, apps, and all the digital platforms we use every day) began to use the same "weapon" theorized by the two fathers of libertarian paternalism in the digital environment. Today, in fact, digital nudges condition our daily lives.[254]

Overall, in the face of this growing dependence on technology and the new risks emerging, who is our Virgil in this "digital hell"? Who can become the guarantor who can watch over our behavior? How can we protect our right to autonomy?

Some scholars, the more optimistic ones, argue that we can become the guardians of this tools, and set the rules for it.[255]

But is this really the case? Are we then able to identify the blind spots in our choices and independently change behavior? Can we manage not to be influenced?

In addition, the extensive use of electronic devices for every aspect of our daily lives raises serious increases the risk of suffering cybercrime and can also lead to problems for personal health and the environment.[256]

---

[249] See *Section 3.2.1.*

[250] Norta et al., 2016.

[251] See *Sections 2.4, 2.5,* and *2.6.*

[252] See *Section 3.2.*

[253] Thaler and Sunstein, 2008. See also *Section 3.3.*

[254] See *Section 3.4.*

[255] See e.g., Hundt, R.E. (2015). L'inferno di Internet. *Aspenia*, *68*, 154-166.

[256] See *Sections 1.7.1, 1.7.5,* and *1.7.6.*

"Intelligent" systems capable of learning a wide range of tasks by experimentation alone and often achieving superhuman levels of performance. The examples of Amazon's resume-selection algorithm or the algorithm for calculating the probability of recidivism[257] remind us that these systems can take unforeseen shortcuts, exploiting properties of the environment that may be unknown to us, and without understanding the broader meaning of their actions.[258]

How can we ensure that this does not happen with the very agents we have entrusted with delicate aspects of our lives?

One thing is certain: for a "new world" and modern problems, a cutting-edge solution is needed.

If the goal of nudging is precisely to facilitate decision making, decrease cognitive effort, avoid mistakes, make decisions faster, and free up mental space for other decisions, why then not use the same nudge tools for the reverse process?

To put it another way, a solution in the face of risks technology creates could be to use nudge as a tool to try to achieve not a faster and easier decision, but a more careful and reasoned one.

This new proposal, in the Digital World in which we move, would result in the use of AI as a tool in the debiasing process.

Here, in fact, AI could use the same nudging techniques developed in recent years for more user-sensitive use, working not on system 1 but on the system 2.

So, AI systems could provide important help in this regard: they can mitigate the risks that the use of this same technology creates.

While nudges are useful in many choice contexts, digital nudges have the perverse effect of dishabituating reasoning to decision-making. That is, they depower the individual's ability to choose by producing what Viale refers to as "decision-making atrophy."[259] The use of AI as a tool in the debiasing process in the digital environment would aim precisely to overcome this critical issue: it could enable human beings to make a thoughtful and careful decision and overcome the risks created by technology.

---

[257] See *Section 1.7.3* for the biased algorithm concept.

[258] Cristianini, 2023, p. 151.

[259] Viale, 2018, pp. 220-221.

3.5.1 An ancient solution for modern problems

While seeming at first glance to be an innovative solution, the use of AI as a tool in the process of debiasing the problems that technology creates is reminiscent of an approach that has been used for centuries.

Figure 19. Example of *offendicula*

"*Offendicula*" is a word from the Latin language used in law to refer to instruments used to prevent or obstruct unauthorized access by outsiders to private property or for the defense of real and/or movable property.[260]

Just as *offendicula* are physical tools (the most common are barbed wire, glass shards, and metal spikes) to solve a problem in physical reality (the physical boundary of private property), following the same approach AI (a digital tool) should be used to solve problems in the Digital World.

So, a digital tool for digital problems of a Digital World.

This does not mean that there will not be place for law. If AI can be used to combat harmful behavior, the law may be needed to ensure that such AI tools are actually used, forcing the digital platform and the websites concerned to implement them.[261]

---

[260] For further information see e.g., here: https://en.wiktionary.org/wiki/offendicula#:~:text=Borrowed%20from%20Latin%20offendicula%2C%20plural,")%2C%20derived%20from%20offendō.

[261] Norta et al., 2016.

3.5.2 Some concrete applications

What could be some tools that through technology aim to create a user-friendly digital environment?

Answering this question means trying to find techniques by which the citizen-user of the Digital World can move more safely in the digital environment and make reasoned and informed choices.

The desirable approach in such cases is not a paternalistic one. Therefore, compared to strict regulation measures,[262] such as banning the use of certain sites or App because they are dangerous,[263] libertarian paternalism modes of intervention would be preferable.

Indeed, this would on the one hand allow individual freedom not to be restricted, and on the other allow the user of the Digital World to cope with the risks that technology creates.

For example, the real purpose of the recommendation systems in charge of compiling personalized lists of news and content on social media or streaming platforms is to constantly observe our choices and increase traffic to their Web service by making us spend more and more time connected.[264]

Are we sure that the goals of these systems coincide with our own? Are we aware of the risks that prolonged use of electronic devices could have on our lives?

In this case AI could be useful to record our online activities and to have feedback on our time spent online. Such feedback can be useful to have objective, quantified feedback that can prompt us to review the quality and quantity of our time spent online. Decreasing online time means, for example, also decreasing the risk of suffering cybercrime, having health problems, and even reducing the environmental impact that these technologies have.

Obviously, however, in order to be successful this feedback should be sent with a criterion: for example, they should not be too frequent because otherwise we would risk information overload[265] and they could be sent in the form of notifications while using a particular site or app.

---

[262] As highlighted in *Section 3.2.1*, regulation is considered as the first standard behavioral change tools.

[263] For example, a few months ago the Italian government was considering a possible blocking of the Chinese social network TikTok for civil servants. For further information see e.g., here: https://decode39.com/6093/italy-tiktok-ban-charm-offensive/.

[264] Cristianini, 2023, p. 125.

[265] Information overload is the difficulty in understanding an issue and effectively making decisions when one has too much information about that issue. For further information see e.g., here: https://en.wikipedia.org/wiki/Information_overload.

Another application, even if more radical, could be to block access to the web or certain sites for a period of time (like Freedom app).[266] This is mostly used by those who work long hours at the PC distinguish work time from leisure time. Of course, even in this case it would not be a permanent blockade, otherwise from a libertarian paternalistic form of regulation, we would move to a pure paternalistic one.

An AI system could also encourage household energy conservation, for instance, may seek individuals whose preferences it predicts to favor energy conservation, people it predicts to consume more energy irrespective of whether their preferences favor conservation, or simply consumers whom the system estimates to be the most susceptible to a particular nudge based on their known or estimated personal characteristics.[267] Again, one would not use awareness campaigns or reports of numerous pages because the individual might have difficulty understanding them by having too many at their disposal ("overload information"). Rather, a more effective remedy would be to use the same technology to solve the problems it creates. For example, intelligent electronic systems that provide feedback via cell phone notifications could be integrated within homes or apartment buildings.

Indeed, mobile phone text messaging[268] or cell phone notifications could be potentially powerful tools for behavior change because they are widely available, inexpensive, and instant.

Figure 20. Cell phone notification used by Apple when the user reaches the recommended 7-day audio exposure limit[269]

[266] For further information see e.g., here: https://en.wikipedia.org/wiki/Freedom_(application).

[267] Tor, 2023.

[268] Cole-Lewis & Kershaw, 2010.

[269] Image source here: https://support.apple.com/en-us/HT211903.

Also, governments could use AI behavioral change tools on their own websites, to facilitate their online interactions with citizens, as when they remind visitors to pay taxes on time, default them into specific selection of government benefits, and so on.

Another area where using AI as a debiasing tool could be useful is in data privacy and security. Digital service providers, as already discussed extensively, provide a free service as a rule, but in return they take the data we produce during our use. The current solution outlined by our legislature is to ask the user who registers for example with a platform or books an airline ticket to check a box where he or she declares that he or she has read a website's terms and conditions and privacy policy.

But is this solution, while in compliance with the law, really useful? Is the user aware of what data they share with digital service providers? Does he or she know what the modalities are?

The user in these situations in fact hardly reads these documents. Therefore, the one solution might be to use cell phone notifications to make him aware of how his data is being used or test his awareness of the risks and before registering a particular website or app.

Those just mentioned are just some of the tools that, using AI as a debiasing tool, aim to create a user-friendly digital environment.

Therefore, I argue that in the face of the risks that technology creates and considering the limited rationality of humans, this approach could be considered the best way to regulate the Digital World.

Indeed, policymakers should not be in a hurry to regulate (often overly burdensome) something that changes and evolves every day. Rather, the most fruitful approach would be to dictate the basic principles at which this development should aim. For example, the goals could be protecting minors from online content, saving energy, paying taxes, reducing cybercrime, securing data, etc.

This, on the one hand, would allow large companies to move with greater freedom, but within the limits of clear principles and goals. On the other, it would help create a Digital World that is easier to use and allows the user to make informed decisions. This solution thus not following the paternalistic approach of strict regulation (e.g., banning the use of a particular digital service) would also not restrict the user's freedom of choice.

3.6 Why might AI as a debiasing tool be the best option? Machine rationality to deal with human irrationality

Who can drive better than me? Who can choose the most affordable insurance policy better than me? Who can choose how to spend time better than me? Who can choose the bank or phone company with the best rates better than me?

These are just a few of the hundreds of questions that some of us have asked ourselves at least once in our lives.

Until the last century, the answer to these questions and others like them was simple: "man is the measure of all things,"[270] and because of this, no one can cope better than he can with the challenges that life presents.

This anthropocentric view is in crisis today because of technological development.

Human beings, because of the increasing use of digital technologies in broad areas of our private and professional lives, often make important decisions in digital choice environments.[271]

This can pose several risks because countless studies[272] have highlighted how man is not a rational agent: he does not always maximize utility;[273] he is overconfident (overconfidence bias); he makes wrong decisions and does not always realize it; he hardly changes habits (status quo bias); and he is influenced in his choices by emotions and heuristics (often producing a biased outcome).[274]

What further undermines citizen-user security in the Digital World is also the digital environment in which he or she moves. Indeed, discovering regularities in the environment is a necessary step for an individual to anticipate the consequences of his or her actions.[275]

A regular digital environment is a basic prerequisite for intelligent behavior in the Digital World. This implies first and foremost that it is necessary, as the libertarian paternalism approach teaches us, to create a user-friendly digital choice environment that allows users to make reasoned and informed decisions.

---

[270] The quote is taken as a cue from Protagoras' thought.

[271] Weinmann et al., 2016.

[272] See *Section 2.5* and *2.6*.

[273] Simon, 1972. See also *Section 2.4*.

[274] Kahneman, 2003, 2011, 2017. See also *Section 2.5* and *2.6*.

[275] Cristianini, 2023, p. 53.

But then the answer to the above questions, what is it?

Simple, in many cases "intelligent" systems know how to choose better than we do and, as I proposed in the previous section, a solution for the risks that technology creates could be the use of AI as a debiasing tool. Following the same logic as the *offendicula*, AI, a digital tool, would serve to address the risks of the Digital World.

But why might AI be the best tool? The answer to this question, which thus legitimizes this approach, is to be found in the definition of "intelligence" when we talk about "intelligent" system.[276]

The first AI was based on models: it modeled the semantics of reality. For example, it was necessary to tell the system the correlation between two things: for example, if I am in Rome, then it means that I am also in Lazio (the region). However, if this was not written into the program, the system could never know.

The latest generation of AI, on the other hand, no longer models reality, but follows a purely statistical model: it learns from examples without knowing what it is learning, and its output respond to a purely utilitarian and rational logic.

"Intelligent" systems today are part of our daily actors, and they seem to have a certain behavior.[277] Nevertheless, we must not fall into the trap of the "Hollywood fears";[278] rather, we must see things as they are.

"Intelligent" systems, with which we interact every day, don't have the free will or legal rights that we ascribe to humans and for this reason they cannot be considered truly intelligent in the common sense we attach to the concept of intelligence.[279]

Indeed, they have another kind of intelligence, which consists of the "reproduction" of a human intelligent behavior.

This reproduction, then, is possible not because these systems have innate faculties, but because they use data already present in nature, i.e., generated by some other process, to discover certain statistical regularities and provide an answer personalized and, in case, also adapt to the environment in order to do better in the future.

---

[276] For an introduction to this topic, see also *Section 1.6.*

[277] Heinemann, K., (2022, December 26). Q&A - The Anthropologist of Artificial Intelligence. *Quanta Magazine* (blog), https://www.quantamagazine.org/iyad-rahwan-is-the-anthropologist-of-artificial-intelligence-20190826/.

[278] See *Section 1.6.*

[279] Intelligence is "*the faculty, peculiar to the human mind, to understand, think, and make judgments and solutions based on the data of even intellectual experience.*" See also *Section 1.6.2.*

So, "intelligent" systems, from an immense amount of data can discover relationships and regularities in the world beyond our understanding and do things we cannot match.[280] This is possible thanks to their practical rationality. They can lucidly and effectively calculate utility maximization and cost minimization, neglecting, indeed excluding from the problem any consideration of merit influenced by emotional and sentimental impulses beyond pure rationality.

If, for example, we placed an AI in front of the railroad switch lever of the famous trolley problem,[281] it would be reasonable to assume that, in all its variants, it would always opt for the action that most minimizes the damage, according to a purely rational cost-benefit evaluation of each option, rather than to a consideration based on other irrational elements.

In a sense, this approach, determined solely by arithmetic and statistical calculation, might paradoxically be even more ethical than the human one in some circumstances. If, for example, a relative of the AI (if it were possible) were tied to the route, it would choose the least costly hypothesis in terms of human lives in any case. This is because the machine's reasoning is guided exclusively by a logic of rational choice that ignores irrational or emotional elements.

And to answer the question that gives this last paragraph its title, this logic of these "intelligent" systems is precisely why AI should be considered as the best debiasing tool.

"Intelligent" systems then due to their rationality and "ethicality" in this sense, could function as a kind of "consultants" for making choices in the Digital World.

Certainly, the "intelligent" systems that are there today will improve, the regulatory approach of policymakers will be clearer and more informed, people will have more digital skills to deal with the risks that technology creates, and much more. But the progress we are heading toward will be in vain if we do not try to be humbler and stop sinning by ὕβρις.[282]

This is the challenge of the future: to accept that we are more at the center of the world, but we should put ourselves at the service of it. Human beings are no longer the measure of all things, but the symbiosis human beings-intelligent systems could be.

---

[280] Cristianini, 2023, p. 69.

[281] The trolley problem is a series of thought experiments in ethics and psychology, involving stylized ethical dilemmas of whether to sacrifice one person to save a larger number. The series usually begins with a scenario in which a runaway tram or trolley is on course to collide with and kill a number of people (traditionally five) down the track, but a driver or bystander can intervene and divert the vehicle to kill just one person on a different track. For further information see e.g., here: https://en.wikipedia.org/wiki/Trolley_problem.

[282] Hubris describes a personality quality of extreme or excessive pride or dangerous overconfidence, often in combination with (or synonymous with) arrogance (for further information see e.g., here: https://en.wikipedia.org/wiki/Hubris).

Conclusions

The journey we embarked upon in this thesis began with an exploration of the historical development of technology and AI, leading us to the present-day Zuckerberg Galaxy. Along the way, we considered the transformation of individuals from mere citizens to "quantified selves" within this Digital World.

Amid the fascination and promise of "intelligent" systems, we addressed the question of whether humans and AI can coexist in harmony or clash. We debunked some of the "Hollywood fears" by examining the true nature of AI.

The risks and implications of technology have not been overlooked. Privacy concerns, data security issues, biased algorithms, and potential job losses have cast shadows on the path to a fully digitized world. In addition, the proliferation of cybercrimes, environmental impacts, and unforeseen health risks have further complicated this path.

Our exploration extended to human decision-making model, revealing the inherent limits of human rationality. We traversed the landscape of behavioral economics, where rational choice theory crumbled under the weight of heuristics and cognitive biases. The dual cognitive system, Prospect Theory and various biases have exposed the blind spots of human choice.

In our attempt to master the art of decision making, we evaluated the limitations of traditional tools of behavioral change such as regulation, incentives, and education. We have also explored the impact that libertarian paternalism and nudging have had over the past fifteen years. However, when faced with the risks that technology creates, we have seen how these may not always be appropriate measures to counter them.

On the contrary, it is in the use of AI as a debiasing tool that we find a beacon of hope.

AI, with its rationality and utilitarian logic, stands as a modern Virgil that can guide us through the "digital hell" of the limits of human rationality and the blind spots of human choices.

In conclusion, the relationship between "intelligent" systems and the human beings has brought us into an age of immense possibilities and complex challenges. The Digital World invites us to navigate its ever-changing sea wisely and prudently. Harnessing the power of AI as a debiasing tool could be a solution to harmonize our coexistence with intelligent systems, moving us toward a future where technology itself will enable us to make more informed, reasoned, and rational decisions. As we embark on this journey, it will be our task to continue to explore the ethical, social, and regulatory dimensions of this evolving relationship, ensuring that the evolution toward an increasingly Digital World is carried out considering clear ethical principles.

Bibliography

Alemanno, A., & Sibony, A. L. (Eds.). (2015). *Nudge and the law: A European perspective*. Bloomsbury Publishing.

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: journal of the Econometric Society*, 503-546.

Amato, A. (2017). Tecno-regolazione e diritto. Brevi note su limiti e differenze. *IL DIRITTO DELL'INFORMAZIONE E DELL'INFORMATICA*, 147-167.

Ariely, D. (2010). Predictably Irrational, Revised and Expanded Edition: The Hidden Forces That Shape Our Decisions. HarperCollins.

Arnaudo, L. (2012). Elementi di Economia e Diritto cognitivi.

Baldwin, R. (2014). From regulation to behaviour change: Giving nudge the third degree. *The Modern Law Review*, *77*(6), 831-857.

Bayamlıoğlu, E., & Leenes, R. (2018). The 'rule of law'implications of data-driven decision-making: a techno-regulatory perspective. *Law, Innovation and Technology*, *10*(2), 295-313.

Baudrillard, J. (2010). *Cyberfilosofia*. Mimesis

Bentivegna, S., & Artieri, G. B. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale*. Gius. Laterza & Figli Spa.

Bernstein, J., Longo, G. (1990). *Uomini e macchine intelligenti*. Adelphi.

Besanko, D., Coccorese, P., Ottone, S., Gibbs, M. J., Braeutigam, R. R., Cipriani, G. P., Cipriani, G. P., & Braeutigam, R. (2020). *Microeconomia* (Quarta ed.). McGraw-Hill.

Brozzetti, F. E. (2019). Afterword in: *RIFD. RIVISTA INTERNAZIONALE DI FILOSOFIA DEL DIRITTO*, (2-3), p. 175-209

Burgess, S., & Ratto, M. (2003). The role of incentives in the public sector: Issues and evidence. *Oxford review of economic policy*, *19*(2), 285-300.

Cardon, D. (2018). *Che cosa sognano gli algoritmi*. Edizioni Mondadori.

Cartwright, A. C., & Hight, M. A. (2020). 'Better off as judged by themselves': a critical analysis of the conceptual foundations of nudging. *Cambridge Journal of Economics*, *44*(1), 33-54.

Castells, M. (1996). The Rise of the Network Society, The Information Age: Economy, Society and Culture, Vol. I. Wiley.

Cole-Lewis, H., & Kershaw, T. (2010). Text messaging as a tool for behavior change in disease prevention and management. *Epidemiologic reviews*, *32*(1), 56-69.

Cristianini, N. (2023). La scorciatoia.

Damasio, A. (2005). Descartes' Error: Emotion, Reason, and the Human Brain. Penguin Publishing Group.

Egidi, M. (2006). From bounded rationality to behavioral economics. *Storia del pensiero economico*, (2006/1).

Evans, J. S. B. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic bulletin & review*, *13*(3), 378-395.

Fisher, W. W., & Mazur, J. E. (1997). Basic and applied research on choice responding. *Journal of applied behavior analysis*, *30*(3), 387-410, p.387.

Fischoff, B. (1982). Debiasing. *Judgment under uncertainty: Heuristics and biases*, (31).

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

Floridi, L. (2015). *The onlife manifesto: Being human in a hyperconnected era* (p. 264). Springer Nature.

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

Geraci, R. M. (2010). *Apocalyptic AI: Visions of heaven in robotics, artificial intelligence, and virtual reality*. Oxford University Press.

Gilli, M. R. (2005). Elementi per un confronto metodologico tra economia comportamentale ed economia neoclassica. *Rivista italiana degli economisti*, *10*(1), 5-22.

Han, B. C. (2014). *Psychopolitik: Neoliberalismus und die neuen Machttechniken*. S. Fischer Verlag.

Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, *4*(60), 78-60.

Heukelom, F. (2014). *Behavioral Economics: A History* (Historical Perspectives on Modern Economics). Cambridge University Press.

Hildebrandt, M., & Gaakeer, J. (Eds.). (2013). *Human law and computer law: comparative perspectives* (No. 18177). Heidelberg: Springer.

Hundt, R.E. (2015). L'inferno di Internet. *Aspenia*, *68*, 154-166.

Hutchinson, J. M., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural processes*, *69*(2), 97-124.

Jevons, W. S. (1879). *The theory of political economy*. Macmillan.

Johnson, E. J., & Goldstein, D. G. (2004). Defaults and donation decisions. *Transplantation*, *78*(12), 1713-1716.

Jolls, C., Sunstein, C. R., & Thaler, R. (1997). A behavioral approach to law and economics. *StAn. l. reV.*, *50*, 1471.

Jolls, C., & Sunstein, C. R. (2006). Debiasing through law. The Journal of Legal Studies, 35(1), 199-242.

Jordan, J. M. (2022). *Robot. Cosa sono e come funzionano le macchine intelligenti*. Luiss University Press.

Kahneman, D. (1979). Prospect Theory: An analysis of decisions under risk. *Econometrica*, *47*, 278.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, *93*(5), 1449-1475.

Kahneman, D. (2011). Fast and slow thinking. *Allen Lane and Penguin Books, New York*.

Kahneman, D. (2017). *Thinking, fast and slow*.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. Journal of Economic perspectives, 5(1), 193-206.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (2008). The endowment effect: Evidence of losses valued more than gains. *Handbook of experimental economics results*, *1*, 939-948.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: a flaw in human judgment*. Hachette UK.

Keeney, R. L. (2004). Making better decision makers. *Decision analysis*, *1*(4), 193-204.

Kerikmäe, T., & Rull, A. (Eds.). (2016). *The future of law and etechnologies* (Vol. 3). Springer International Publishing.

Kissinger, H. A., Schmidt, E., Huttenlocher D. (2021). The Age of AI. John Murray Press.

Korobkin, R. The Endowment Effect and Legal Analysis' (2003). *Northwestern University Law Review*, *97*, 1227.

Korobkin, R. B., & Ulen, T. S. (2000). Law and behavioral science: Removing the rationality assumption from law and economics. *Calif. L. Rev.*, *88*, 1051.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the national academy of sciences*, *110*(15), 5802-5805.

Krugman, P. R., Wells, R. (2018). *Microeconomics*. Macmillan Learning.

Luppi, B. (2021). Behavioral Biases and the Law. *Review of Law & Economics*, *17*(2), 453-464.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

Mathis, K. (Ed.). (2015). *European perspectives on behavioural law and economics* (Vol. 2). Springer.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences*, *114*(48), 12714-12719.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, *27*(4), 12-12.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*, 115-133.

McCune, J. C. (1998). Data, data, everywhere. *Management Review*, *87*(10), 10.

McKenzie-Mohr, D., & Schultz, P. W. (2014). Choosing effective behavior change tools. *Social Marketing Quarterly*, *20*(1), 35-46.

McLuhan, M. (1962). *The Gutenberg galaxy; the making of typographic man*. University of Toronto Press.

Mill, J. S. (1974). On liberty (1859).

Minsky, M. (1965). Matter, mind and models.

Negroponte, N. (1995). *Being digital*. Knopf.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, *2*(2), 175-220.

Nida-Rümelin, J., Weidenfeld, N. (2022). *Digital Humanism: For a Humane Transformation of Democracy, Economy and Culture in the Digital Age*. Springer International Publishing.

Norta, A., Nyman-Metcalf, K., Othman, A. B., & Rull, A. (2016). "My Agent Will Not Let Me Talk to the General": Software Agents as a Tool Against Internet Scams. *The Future of Law and eTechnologies*, 11-44.

Pasquinelli, E. (2012). Irresistibili schermi: fatti e misfatti della realtà virtuale. Mondadori università.

Posner, R. A. (1997). Rational choice, behavioral economics, and the law. *StAn. l. reV.*, *50*, 1551.

Puaschunder, J. M. (2017). Nudging in the digital big data era. *European Journal of Economics, Law and Politics*, *4*(4), 18-23.

Prensky, M. (2005). Digital natives, digital immigrants. *Gifted*, (135), 29-31.

Ratti, C. (2015). *Gli innovatori*. Aspenia, 68, 44-49.

Robbins, L. (1984). *An essay on the nature and significance of economic science*. New York University Press.

Romeo, F. (2012). *Lezioni di logica ed informatica giuridica*. Giappichelli.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*, 7-59.

Sartori, G. (2014). *Homo videns: televisione e post-pensiero*. Gius. Laterza & Figli Spa.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, *3*(3), 417-424.

Searle, J. R. (1990). Is the brain's mind a computer program?. *Scientific American*, *262*(1), 25-31.

Searle, J. R. (1990, November). Is the brain a digital computer?. In *Proceedings and addresses of the American Philosophical Association* (Vol. 64, No. 3, pp. 21-37). American Philosophical Association.

Searle, J. R. (1991). Consciousness, unconsciousness and intentionality. *Philosophical Issues*, *1*, 45-66.

Schultz, P., & Kaiser, F. G. (2012). Promoting pro-environmental behavior.

Selwyn, N. (2009, July). The digital native–myth and reality. In *Aslib proceedings* (Vol. 61, No. 4, pp. 364-379). Emerald Group Publishing Limited.

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, *2*(2), 895-908.

Simon, H. A. (1972). Theories of Bounded Rationality. *i McGuire. CB og Radner, R.(red.)*, 161-176.

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 1159-1202.

Sunstein, C. R. (2014). Nudging: a very short guide. *Journal of Consumer Policy*, *37*, 583-588.

Talia, D. (2019). *Big data and the computable society: algorithms and people in the Digital World.*

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038-8043.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, *112*(S1), S164-S187.

Thaler, R. H., Sunstein, C. R. (2009). Nudge: Improving Decisions about Health, Wealth and Happiness. Penguin Books.

Tor, A. (2008). The methodology of the behavioral analysis of law. *Haifa Law Review*, *4,* 237.

Tor, A. (2022). The law and economics of behavioral regulation. *Review of Law & Economics*, *18*(2), 223-281.

Tor, A. (2023, March). Digital Nudges: Contours and Challenges. In *International Law and Economics Conference* (pp. 3-18). Cham: Springer Nature Switzerland.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, *185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1989). Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Berlin, Heidelberg: Springer Berlin Heidelberg.

Tyagi, K. (2015). Behavioral Approach to Law: An Emerging Discipline. *Available at SSRN 2586005*.

Ulen, T. S. (2013). Behavioral law and economics: Law, policy, and science. *Supreme Court Economic Review*, *21*(1), 5-42.

Viale, R. (2018). *Oltre il Nudge. Libertà di scelta, benessere e felicità* (pp. 1-264). Il Mulino.

Weinmann, M., Schneider, C., & Brocke, J. V. (2016). Digital nudging. *Business & Information Systems Engineering*, *58*, 433-436.

Weinstein, N. D. (1989). Optimistic biases about personal risks. *Science*, *246*(4935), 1232-1233.

White, M. D. (2017). *The Decline of the Individual: Reconciling Autonomy with Community*. Springer.

Whitehead, M., Jones, R., Howell, R., Lilley, R., & Pykett, J. (2014). Nudging all over the World. *ESRC Report, Economic and Social Research Council, Swindon and Edinburgh*.