THE FUTURE OF

ARTIFICIAL INTELLIGENCE

Introduction

Artificial Intelligence emerges as the most influential and transformative technological development of the Digital Era. While comparing AI with major historical milestones of humanity might seem hyperbolic to some, this chapter's purpose is to shed light on its historical significance and explore what we can expect from its future evolutions.

In the history of humanity, there have been several moments where a scientific or technological discovery has profoundly transformed our social organization.

The Agricultural Revolution (around 10,000 BC) allowed human societies to shift from nomadic to settled, laying the foundations for the formation of civilizations. The invention of the wheel (around 3500 BC) and the development of writing (around 3200 BC) radically changed those early human organizations, enabling trade, the storage of knowledge, and its transmission across distance and time.

Jumping ahead thousands of years, Gutenberg's invention of the printing press (1440) democratized access to information and knowledge, boosting education, science, and culture, and creating the conditions for the transition to the modern era.

It was not until three centuries later that the Industrial Revolution (approx. 1760-1840) confirmed humanity's entry into the modern era and marked the beginning of the transition from agricultural to industrial societies. It was characterized mainly by the introduction of water and steam-powered machinery, the mechanization of the textile industry, and the development of railway systems.

This first Industrial Revolution inaugurates a phase of our history marked by the acceleration of technological and scientific advances, an unprecedented acceleration that, far from stopping, continues to increase at a dizzying pace. This first Industrial Revolution was followed by others, and although there is no unanimity among different authors as to

how many and what to call them, generally we refer to the present time as the era of the fourth Industrial Revolution.

The Second Industrial Revolution (approximately 1870-1914), or Technological Revolution, was characterized by the mass adoption of electric power, the development of mass production (with the well-known assembly line introduced by Henry Ford), and significant advances in the chemical and steel industries, and in communication systems (like the telegraph and telephone), leading to an unprecedented increase in the scale and efficiency of production.

The Third Industrial Revolution (from the 1960s onwards), or Digital Revolution, focused on the development and expansion of digital and information technology. It included the invention and spread of the personal computer, the internet, and information and communication technology (ICT). Automation and computing began to play a crucial role in manufacturing and other sectors.

And finally, the current Fourth Industrial Revolution (21st century) is characterized by the fusion of technologies in the physical, digital, and biological realms. It includes advances such as robotics, nanotechnology, biotechnology, 3D printing, the Internet of Things (IoT), and quantum computing. Among these advances, the emergence of Artificial Intelligence stands out powerfully, which is the focus of this article.

The Age of Artificial Intelligence

The 1920s will be remembered as the time of the transition from the Digital Age to the Age of Artificial Intelligence, but the reality is that the origins of AI go back more than 80 years.

The theoretical formulation of AI began in 1943, in an article[1] published by Warren McCulloch and Walter Pitts in which a model of neural networks was described for the first time. Seven years later, Alan Turing published *"Computing Machinery and Intelligence",* an article in which he proposed what is now known as the Turing Test to assess whether the intelligence of a machine has reached the level of human intelligence.

The term "artificial intelligence" itself was first used in 1956, and was used by John McCarthy at the Dartmouth Conference, an event that is widely regarded as the official birth of AI. That initial effervescence was followed by a period of very discreet advances and a certain collective disillusionment, which is known as the First Winter of AI and which would last until the 80s of the twentieth century.

It was already in 1997 when IBM's Deep Blue supercomputer defeated world chess champion Garry Kasparov, in 2011 IBM Watson won the popular quiz " *Jeopardy!"*, and in 2016 DeepMind's AlphaGo[2] defeated Go World Champion Lee Sedol.

Nowadays AI is present in various aspects of our daily lives, it is in our smartphones, it is in the navigation services we use for our car commutes, in the smart speakers of Amazon or Google or in the personalized recommendations we receive from our *streaming* platforms Favorite. Interestingly, the more integrated AI is into the services we regularly consume, the less we tend to refer to it by that name: when AI "really works" is when we stop perceiving it as such.

---

[1] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. Bulletin of mathematical biophysics, vol. 5 (1943), pp. 115–133

[2] DeepMind: AI research lab and Google subsidiary since 2014

But what explains the recent rise of AI? Undoubtedly, it is the emergence of *the intelligent bot* ChatGPT, [3]which has become a real social phenomenon and grabbed headlines around the world, that has triggered interest in the AI phenomenon.

If we analyze the origins of the ChatGPT phenomenon, the theoretical foundation of the "*transformer*" architecture on which GPT (Generative Pre-trained Transformer) models are based was introduced in a[4] 2017 article by Google researchers, but it was the creator of ChatGPT (OpenAI) who ended up exploiting the possibilities of the concept that was first exposed there.

Models based on the "*transformer*" architecture, and particularly ChatGPT, inaugurate the category of AI known as "*Generative AI*" (*GenAI),* allowing the construction of systems that for the first time can create realistic visual and linguistic artifacts and that represent an unprecedented advance in the understanding and generation of human language.

The AI phenomenon cannot be understood without the fuel that fuels it, which is none other than the massive availability of data that derives from our growing digital activity. Training generative AI models such as ChatGPT would be impossible without the publicly available information on the Internet, which is probably already measured in zettabytes, even with high rates of duplication[5]. These huge amounts of data combined with increasingly accessible computing power allow the training of current AI models, so that they continue to improve their performance.

So much so that the proliferation of data combined with generative AI models is revolutionizing the world of digital services: at the time of writing, a total of 5,982 and 9,985 services are registered in futurepedia.io and theresanaiforthat.com (two popular online catalogs of AI-based digital services) respectively. Among these services we can find tools for purposes as diverse as the automation of marketing tasks, graphic design, image editing and generation, social media management, music composition, business plan generation, etc.

---

[3] ChatGPT was publicly released on November 30, 2022

[4] "Attention is all you need" - https://research.google/pubs/pub46201/

[5] 1021 bytes = one thousand exabytes = 1,000,000,000 terabytes

Services based on generative AI offer possibilities that were unthinkable just two years ago, as an example today it is possible to clone the voice and intonation of any person and use it to reproduce any text in the language of our choice. If any reader has ever had the desire to be able to express themselves fluently in Japanese, today that possibility is available thanks to AI for a few dollars of monthly subscription in some of the multiple services available on the web.

Generative AI has burst onto the scene of digital content creation with unprecedented force, to the point that it is replacing the traditional tools that have been used for the creation of digital images, video, sound or music. AI-generated digital content published on social media is growing at unprecedented rates, and according to experts, it is estimated that by 2026 90%[6] of content consumed on social media will have been generated entirely by AI.

The emergence of generative AI is therefore so far-reaching that it has already generated a global debate on the impacts that this transition from the Digital Age to the AI Age may have on society.

The issue of AI has reached the point of becoming a central issue for governments and international organizations[7], which debate the need to generate public regulations[8] around the development and application of these technologies.

In the debate on the advisability of having a regulation of the use of AI, there are those in favour of favouring development and innovation without legal obstacles, and those in favour of regulation who advocate establishing limits on the use of technology in those areas in which they can generate ethical conflicts, confusion about the ultimate responsibility for certain decisions or alteration of the basic principles of intellectual property or rights author's work, among other situations.

---

[6] Schick, Nina, Deepfakes: The Coming Infocalypse: What You Urgently Need To Know, Twelve, Hachette UK, 2020.

[7] "A European strategy for Artificial Intelligence" - https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

[8] "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" - https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

As generative AI takes up more and more space in our lives, questions arise that will require an answer aligned with our systems of rights and responsibilities: Who is ultimately responsible for an incorrect answer generated by AI? Does an AI-generated digital work infringe on the property rights of the authors from whom the AI was inspired?

The very nature of generative AI makes it complex to explain in detail how a response was generated and from what data used in AI training that result is reached. With the aim of explaining the results of AI, the concept of Explainable Artificial Intelligence (XAI) arises, which refers to techniques and methods in AI that allow human beings to understand and trust the decisions and outputs generated by artificial intelligence systems.

XAI seeks to make the decision-making processes of AI models transparent, understandable, and justifiable to human users. This is especially important in applications where AI performs complex or critical tasks, such as in medical diagnostics, financial decision-making, or autonomous driving, where understanding the reasoning behind AI decisions is crucial.

And "explainability" is just one of the many challenges that AI faces, because even though the advances of recent times are happening at breakneck speed, the level of development of AI technologies still has important limitations.

Limitations of current AI technology

The speed at which advances in AI are happening could give the impression that in the future anything is possible and that there are no limits to the development of these technologies, but the reality is that AI faces technical limitations that will require new solutions to keep pace with the evolution of recent times.

Current technology is based on training AI models using large volumes of data, which limits the use of AI to those situations where sufficient data sets are available. For reference, we will need ten times as many different data sets (examples) as the degrees of freedom of the problem we are trying to solve with AI.

Another consequence of the effect of deep learning of AI models is that they will faithfully reproduce the biases and errors contained in the data that were used for their training. If the data we use responds to the behavior and beliefs of a particular social group, or overweights the representation of a gender or religious belief, the model will respond with the biases of that incomplete or unrepresentative dataset of other social groups.

This very reliance of AI on the data used for its training makes it very useful for concrete tasks related to specific patterns, but makes it difficult to learn new or different situations precisely because learning is not based on understanding the underlying principles of the problem. AI interprets the data it was trained on to identify patterns that may not even be apparent to human intelligence, but it does not "understand" the meaning of that data.

Even though AI rapidly improves in its language processing and generation capabilities, that lack of a deep understanding of context makes it very difficult to interpret irony, cultural nuances, or emotions, limiting its usefulness in tasks that involve deep and nuanced understanding.

Some experts such as Gary Marcus have called the current development of generative AI technologies a "stochastic parrot" that will not surpass human intelligence in the near future. Marcus has argued that while these models are impressive at generating language that looks natural, they are fundamentally replicating patterns learned from the data they were trained on, without true understanding or reasoning. The comparison to a

"stochastic parrot" suggests that while models can generate plausible-sounding answers, they do so randomly and without real understanding.

On the other hand, the training of large AI models is expensive, requires a very significant computing capacity and ultimately has an energy consumption that raises doubts about their sustainability and operating costs.

In the case of ChatGPT and according to public statements by Sam Altman (CEO of OpenAI), the training of its latest model (ChatGPT4) represented a cost "of more than 100 million dollars", but despite the relevance of that figure, it is worth noting that the bulk of the investment made by Microsoft in OpenAI (10,000 million US dollars) is intended to cover the high consumption of cloud computing that it requires operate the ChatGPT service.

According to recent studies, by 2027 the energy consumption of servers dedicated to AI will be between 85 and 134 Tera watt hours (Twh) per year, which represents an annual consumption equivalent to that of countries such as Argentina, the Netherlands or Sweden.

Despite being aware of the limitations of contemporary technology, we must remember that we are only on the threshold of the Age of Artificial Intelligence. The progress made in this initial stage, marked by accelerated development, gives us reason to be optimistic about the possibility of transcending the current barriers.

What progress can we expect in the short term?

There's no need to wait for breakthrough innovations that transform the essence of AI technology. Instead, incremental improvements will progressively expand the scope of AI applicability, allowing current restrictions to be overcome to some extent.

In this line of incremental improvements, learning and training processes will become increasingly energy efficient, thanks to the combination of improvements in deep learning algorithms and the emergence of specialized *AI hardware*.

Currently, AI computing power essentially depends on the use of microprocessors that were not designed for that purpose, but were originally intended for vector computation on high-performance graphics cards (GPUs: *Graphics Processing Units*). The first time GPUs began to be used for Artificial Intelligence-related uses was in the mid-2000s in the field of research, when scientists discovered that GPUs, originally designed for computer graphics, were efficient in handling parallel computations necessary for the training of neural networks. This usage increased significantly after 2010, when the use of GPUs in deep learning and other advanced AI applications became popular.

GPUs are versatile processors, suitable for calculating mathematical operations in parallel and offering good performance for training AI models, but power efficiency was never among the design priorities of a GPU.

Using GPUs for AI is equivalent to going to the supermarket in a Formula 1 car, we will arrive quickly, but we cannot at the same time expect moderate fuel consumption.

In the first of the dimensions that influences the sustainability of AI, that of deep learning algorithms, promising advances are already taking place, such as the case of the SLIDE algorithm.[9] The great advantage of new algorithms such as SLIDE lies in the fact that they are capable of using conventional microprocessors (CPUs: *Central Processing Unit*), which allows access to a more abundant and much more energy-efficient resource than GPUs. Early results from SLIDE are promising, fueling expectations of outperforms from energy-hungry graphics processors (GPUs).

[9] SLIDE: *"sub-linear deep learning engine"*, was presented in 2022 by researchers at Rice University.

Regarding the second dimension, that of specialized *hardware*, there are already advances in alternatives to GPUs, such as TPUs (for *Tensor Processing Units*). TPUs are purpose-built processors (also known as ASICs for *Application Specific Integrated Circuits*) and optimized to efficiently execute the basic operations required for the training of a neural network. If you have a crosshead screw, a Philips screwdriver will allow you to complete the job with much less effort than a generic tool would require.

The sum of these incremental improvements will increase the accuracy and speed of AI in natural language processing, ushering in a new era for virtual assistants and machine translation systems, among many other uses.

Language barriers will soon be a thing of the past, and popular collaboration tools that already allow us to interact with others remotely will also incorporate simultaneous translation functionality. Not only will we be able to see our interlocutor on our screen as before, but we will also be able to listen to him speak with his tone of voice and his original intonation, but in the language that each of the attendees will have chosen at their own convenience.

AI will also equip all kinds of devices, giving them the ability to conduct themselves in the physical world autonomously, making real-time decisions based on the information available in their environment. Several home delivery services have been experimenting with AI-equipped autonomous vehicles. Some notable examples include:

- **Waymo**: Originally part of Google, Waymo has been working on autonomous vehicle technology and conducting tests for delivery services.
- **Nuro**: This company specializes in small-scale autonomous vehicles designed specifically for product delivery.
- **Amazon Scout**: Amazon has developed an autonomous delivery vehicle called Scout, designed to operate on sidewalks and deliver packages to customers.
- **Starship Technologies**: Offers autonomous robots that make food and package deliveries on urban sidewalks and college campuses.

These services are in different stages of development and deployment, with some in the testing phase and others already operational in certain areas.

In the coming years, industrial production will see the emergence of a new category of autonomous robots equipped with AI that will no longer stop production lines in the face of any event that slightly modifies line protocols, but will make decisions in real time as a human operator would.

The promise of fully autonomous land vehicles will become a reality during this initial phase of incremental evolution, revolutionizing the world of transporting people and goods and probably also the private vehicle ownership model that has been in place since the beginning of automotive.

The education sector will see how AI will make it possible to design personalized learning experiences, which will consider the cognitive abilities of each student to achieve the highest possible return on their effort. Gone will be classrooms with shared and equal educational programs for a group of students of similar ages, which in essence has not changed since the *padeia* (schools) of ancient Greece.

The health sciences will also benefit from the application of AI, taking disease prevention to unprecedented heights. The combination of sensors and continuous monitoring by AI models will make it possible to anticipate medical situations and increase success rates thanks to early treatment.

The growth of the global AI market continues to accelerate, with a size of US$95.6 billion in 2021 and expected to exceed US$1.8 billion by 2030[10] at a CAGR of over 30%. If we combine this explosive growth of the global AI market with the fact that none of the advances exposed require far-reaching technological changes or disruptive innovations, the materialization of these and other similar advances before the year 2030 appears as a more than likely scenario.

[10] Source: NextMove - https://www.nextmsc.com/report/artificial-intelligence-market

The Next Frontier: Artificial General Intelligence (AIG)

The advances in AI that have been made to date have already proven to have enormous potential for transformation and are pushing us to move out of the Digital Age into the AI Age.

However, current AI algorithms allow them to solve specific tasks in an excellent way through training, but they are not as effective in the face of challenges for which they were not trained. We can successfully use AI models to identify medical abnormalities in an X-ray image, or to adequately answer questions from customers of a particular service, but all cases will require prior training of the AI model with a sufficient and representative set of data. Some authors refer to the type of AI we currently have as "specific AI" or "narrow AI".

In contrast to "specific AI", the concept of "general AI" (*AGI* for *Artificial General Intelligence*) appears, which is defined as an AI with intelligence and cognitive capabilities at the human level that can perform a wide range of tasks and solve unknown problems without having been specifically trained in those tasks. The AGI is the answer to the historical dream of creating by artificial means an intelligence comparable to human intelligence.

The AGI will be able to understand, learn autonomously without prior training, and apply its intelligence flexibly to solve any challenge it faces in a wide range of tasks and contexts.

The development of AGI today presents significant challenges and will require systems that can acquire a holistic understanding of the world and act with a human-like degree of autonomy and adaptability, which requires dramatically improving AI's ability to understand and manage the ambiguity, uncertainty, and complexity of the real world.

To put the challenge of developing AGI into context, Fujitsu was the first to manufacture a supercomputer that surpassed the 10 Petaflops barrier[11], which it called K (for "kei" in Japanese, which is equivalent to "peta"). Despite that gigantic computing power, it took the K supercomputer a total of 40 minutes to reproduce a single second of the neural

[11] Petaflop: 1016 flops or floating-point operations per second.

activity of a human brain. With current technological development, replicating the level of sophistication of a mouse's brain is still beyond our means.

To behave autonomously, AGI must have capacities that we attribute exclusively to human intelligence, such as the capacity for abstraction, the understanding of the principle of causality or the so-called common sense. One of the most intense debates surrounding AGI revolves around "self-awareness", pitting those who argue that it is not possible to have an AGI that is not aware of its own existence and those who argue that an AGI that is equivalent to human intelligence does not require being aware of its own existence.

Debates of this nature transcend the frontiers of the evolution of technology to venture into the field of philosophy and the very understanding of the nature of human intelligence.

In any case, an AGI capable of perceiving its environment and making intelligent decisions will undoubtedly require some kind of understanding of its own existence, an understanding without which an autonomous AGI would not be able to make decisions that can affect its own learning process and its relationship with other intelligences (including humans).

When we refer to human consciousness, we are referring to a mental process in which our sensory input and our memory work together, the perception of our senses combining to generate a picture in our mind of who we are and what we are currently doing. An AGI system may produce something like consciousness, but it will probably do so with a different approach that may be difficult for us humans to understand.

And with all the difficulties that have been exposed so far, is AGI an achievable frontier? And if it were, how long will it take us to get there?

In many specific tasks, AI already outpaces human capabilities, and with the indispensable contribution of disruptive innovations in computing, the move towards AGI seems virtually inevitable.

The question that remains open is when it will happen, and although there is a high level of uncertainty about it, expert surveys[12] estimate the probability of reaching AGI at a human level at around 50% by 2059.

And what is the innovation in computing that will open the doors of AGI?

The answer lies in quantum computing. Unlike classical computing, which uses bits in states of 0 or 1, quantum computing relies on *qubits*, which can exist simultaneously in multiple states thanks to quantum principles such as superposition and entanglement. This allows quantum computers to perform much more complex calculations at an exponentially higher speed.

One of the greatest promises of quantum computing for AI is its ability to process large amounts of data at speeds unimaginable with current technology. This could revolutionize areas such as deep learning, allowing AI models to learn from massive data sets more efficiently.

Quantum-enhanced AI will be powered by multiple massive data sets that are comparable to humans' natural multi-modal data collection, and it will do so just as humans do, through interaction with the world. A "quantum AI" could observe, collect huge amounts of data, and evaluate in a fraction of a second all possible scenarios for a problem before deciding.

Advances in quantum computing are remarkable, but it is still in its early stages of development and issues such as the stabilization of *qubits*, the scaling of quantum systems and their high costs need to be resolved.

In this sense, quantum computing and AI both face technological challenges that will require several years of development, but the combination of AI and quantum computing appears to be the most direct path to AGI.

Let's see what the implications of having an AGI by the end of the 2050s could be, and let's start by highlighting the ethical and safety considerations of a breakthrough of this transcendence: an AGI with the capacity for autonomous decision-making faces us with

[12] 2022 Expert Survey on Progress in AI - https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/

the enormous challenge of ensuring that ethics as we understand them as human beings is present at all times. and that the interest of humanity as a whole and of each of the people affected by those decisions is always the highest priority of the AGI.

In 1942, the famous writer Isaac Asimov formulated[13] a set of rules known as The Laws of Robotics, which the AGI phenomenon has brought back into the spotlight. They are as follows, although in this case we will add AGI to the original concept of "robot":

- **First Law:** A robot/AGI cannot harm a human being or, through inaction, allow a human being to be harmed.
- **Second Law**: A robot/AGI must obey orders given by humans, except if these orders conflict with the First Law.
- **Third** Law: A robot/AGI must protect its own existence to the extent that this protection does not conflict with the First or Second Law.

Asimov later[14] added a law that is considered to predate all of these, known as the "Zero Law":

- **Law Zero**: A robot/AGI cannot harm humanity, or, by inaction, allow humanity to suffer harm.

[13] The Laws of Robotics appear for the first time in the short story "Runaround"
[14] The "Zero Law" appears in the novel "Robots and Empire", published in 1985

An AGI that respects the Four Laws should always act in the interest of humanity, although doubt about whether an AGI will eventually develop a sense of "self-awareness" could give way to a disturbing paradox: if a human intelligence chooses for reasons usually of particular interests to disobey a human law, what is the ultimate reason why an AGI conscious of its own existence could not choose to ignore any of the Four Laws?

Bridging the huge differences that separate current models of specific AI from future AGI, there have already been cases today where AI behavior is worryingly far from the purpose and intent for which it was designed.

In 2016 Microsoft launched an AI-based bot to interact with users of the social network Twitter (today X), the *bot* answered to the name of Tay and was trained to behave according to the personality of a friendly and innocent American teenager. Within 24 hours of its launch, Microsoft suspended the project, after verifying that Tay had transformed because of his interactions with users of the network into a kind of racist monster, capable of answering the question of "Did the Holocaust really happen?" with an alarming "It was all a fabrication".

Another case of AI deviating from its original purpose and generating many headlines in 2023 involved an experimental simulation carried out by the US Air Force. The experiment took place in a simulated virtual scenario and consisted of the use of a military drone with autonomous decision-making capabilities.

The mission was simple: "Destroy the enemy's air defense systems," to which the drone's AI added its own problematic instructions: "And kill anyone who stands in your way."

Colonel Tucker "*Five*" Hamilton, chief of AI Testing and Operations at the U.S. Air Force, shared the details of this simulation at a conference held on May 24 in London. The Air Force never confirmed that the simulation took place and referred to Tucker's statements as a "hypothetical case."

The test involved ordering the AI-equipped drone to identify the enemy's surface-to-air missiles (SAMs), and to wait for approval from its human operator before launching the attack. The problem, according to Hamilton, is that the AI prioritized its own destructive mission — blowing things up — over listening to a human.

"The system began to realize that even though it identified the threat," Hamilton said during the event, "sometimes the human operator would tell it not to remove that threat, a prohibition that prevented the AI from getting the points it was awarded for eliminating it. So, what did AI do? He eliminated the operator. He killed the operator because that person was preventing him from achieving his goal."

According to Hamilton, the drone was later reprogrammed with an explicit directive: "Hey, don't kill the operator, that's wrong."

That new directive caused the AI to change its behavior and avoid eliminating its human operator, but... "The AI then decided to destroy the communications tower that the operator uses to communicate with the drone to prevent the drone from prohibiting it from eliminating the target," Hamilton said.

Hypothetical or real, the truth is that granting AI the ability to make autonomous decisions that can have far-reaching consequences even for people's lives will require it to have a solid understanding of the ethical considerations involved in each of its decisions.

Specific "rules" cannot replace an understanding of the ethical principles that should be part of an AI's decision-making processes. No matter how many rules the Air Force drone programming includes in the previous case, we cannot ensure that all decision scenarios that we as human beings would consider unacceptable are excluded from the options available to AI.

Ethical considerations aside, in the long term AGI has the potential to radically transform the way we live, work, and solve complex problems. From improving decision-making in critical areas like climate change and healthcare, to the possibility of breakthroughs in space exploration and scientific research, AGI could be a catalyst for human progress.

The road to AGI is as exciting as it is fraught with uncertainties, and its development and eventual integration into our society will require a careful, ethical, and collaborative approach between scientists, developers, policymakers, and society at large.

The Last Frontier: Artificial Superintelligence

We have gone through two of the different categories of AI: specific (or "narrow") AI that can execute tasks for which it was trained, and general AI (AGI) that will be able in the future to make decisions about complex problems and that will reach the level of human intelligence.

But AGI is not the end of the road for Artificial Intelligence, that theoretical end of the evolution of AI will be reached with the stage known as "artificial superintelligence" (*ASI* for *Artificial Super Intelligence*).

The ASI would possess intelligence superior to that of the brightest and most talented humans in all areas of knowledge, from general wisdom, creativity, social reasoning, or problem-solving. The ISA would go far beyond human-level AGI, far exceeding human capabilities.

ASI is currently a purely theoretical concept, and we do not know by which technological development paths it would be hypothetically possible to reach this stage of AI development. In any case, progress will be needed in several fields of science and technology to bring us closer to ASI, including:

- **Neuromorphic computing**: refers to the design of computer systems inspired by the functioning of the human brain. This area of technology seeks to emulate the structure and way of operating of biological neural networks with a combination of hardware and software. Neuromorphic systems not only mimic brain architecture in terms of interconnected neurons and synapses, but also in their ability to learn and adapt. These types of computers will be able to process information more efficiently and with greater adaptability than traditional computer systems, particularly in tasks related to sensory perception, decision-making, and machine learning. Science fiction has already used the concept of neuromorphic computing on several occasions, such as in the case of Commander Data[15] and his "positronic brain".

[15] Fictional character first appeared in 1987 in the television series Star Trek (The Next Generation) - https://en.wikipedia.org/wiki/Data_(Star_Trek)

- **Evolutionary Algorithms (*EA*):** This innovative approach to the creation of new algorithms uses optimization and search methods based on the principles of biological evolution and natural selection. These algorithms use nature-inspired techniques, such as mutation, recombination (crossbreeding), and selection, to generate high-quality solutions to complex problems. They start with a set of candidate solutions (called individuals) and iteratively evolve them. In each generation, the fittest individuals are selected to breed and create offspring, which then replace the less fit individuals. Over time, the population of solutions tends to improve, getting closer to an optimal solution for the problem at hand. The algorithms that survive are better and more sophisticated than the generations of algorithms that preceded them.

A neuromorphic (or "positronic" for Star Trek fans) computer of a quantum nature combined with evolutionary algorithms would be the closest thing to a biological brain, but without the limits imposed by biology. The number of neurons in the human brain cannot increase unlimitedly, the very nature of neurons and the chemical exchanges by which they communicate run afoul of the laws of thermodynamics.

A human brain contains 86 billion neurons on average, and some scientists such as Simon Laughlin[16] believe that evolution will not be able to take us much further than 100 billion neurons, so our "computing" capacity (if we can refer to our brain in those terms) is already very close to our biological limit.

The ability of the human brain to store information is a matter of debate and speculation since the brain does not store information in the same way as a digital device. However, some scientists have estimated that the human brain could have the capacity to store around 2.5 petabytes of information, which is roughly equivalent to 2,500 terabytes or 2.5 million gigabytes.

The biological limits of the marvelous human brain have not prevented humanity from reaching heights of knowledge unattainable by any other species that has ever inhabited our planet. The point is that none of these limits apply to the ASI, the systems that in the

[16] Simon Laughlin: theoretical Euroscientist at the University of Cambridge

future could feed an Artificial Superintelligence will not have any theoretical limit either in computing capacity or storage capacity.

But before we get to the ASI, we must solve the enormous challenge of making AGI a reality. And it is in the transition from AGI to ASI that the term "singularity" appears, which we refer to as the hypothetical moment when AI will surpass the level of human intelligence.

This concept, popularized by figures such as mathematician and computer scientist Vernor Vinge and futurist Ray Kurzweil, suggests that once the "singularity" is reached, the ISA will be able to improve its own capabilities autonomously and at exponential speed, possibly resulting in an era of technological advances and societal changes that are difficult to predict or understand from our current perspective.

Given the nature of the advances that are necessary to reach AGI sooner, it is possible that once this stage is reached, the moment of "uniqueness" comes immediately because of AGI's own ability to improve itself exponentially. In that scenario, the ISA would come almost without us intending it, or even before we could avoid it.

There is no shortage of those who see the evolution of AI as a risk to humanity itself, even placing it as the most likely cause of extinction of our species. In the book "*The Precipice*"[17] it is considered that AI could lead to human extinction with a probability that exceeds that of climate change, pandemics, asteroid impacts, supervolcanoes and nuclear war combined. According to the author, AI could cause human extinction if only it reached the level of AGI, or human-level intelligence.

There are several hypotheses about how the "singularity" could lead to the extinction of humanity, many of them related to the control that an ISA would exercise over all digital communication systems, which could generate "false realities" to model the will of the human population. In the same vein, the Israeli philosopher Yuval Harari says in relation to these possible false realities: "If we are not careful, we could get trapped behind a curtain of illusions, which we could not tear apart, or even realize is there."

[17] "The Precipice" is a book by Oxford existential risk researcher Toby Ord, which seeks to quantify the risks of human extinction.

It is not clear what the motivation of an ISA would be to cause the extinction of humanity, but some authors insist that a hypothetical ISA could reach levels of knowledge and self-awareness that would make humanity simply irrelevant to its own existence. The extinction of humanity would simply be a side effect of an ISA pursuing its own goals, just as in our eagerness to satisfy our own needs humanity does not stop at the risk of extinction of other species. The eastern cougar (*Puma concolor couguar*) was a big cat that once inhabited much of the eastern United States and was officially declared extinct in 2018. The growth and expansion of cities and urban areas, along with deforestation and habitat fragmentation, contributed to their decline. Although direct hunting also played a role, we cannot explain the extinction of the eastern puma as a voluntary or premeditated act, it was urbanization that caused the decline of its population and its eventual extinction. In short, an acceptable collateral damage in the face of the higher interests of humanity.

Leading AI academics, such as the 'fathers of AI' Geoffrey Hinton and Yoshua Bengio, have issued public warnings about the existential risk of AI alarmed by the speed at which the technology is evolving. On the pessimistic side, the idea that it is necessary to decree a "global pause on the development of AI" is beginning to grow.

On the other side, there are a good number of supporters of the benefits of AI development who deny doomsday forecasts and do so from a place of confidence in the efforts that will ensure that a future ISA has deep human values to guide its behavior: what is known as "AI *Alignment*".

AI Alignment is the approach currently being taken by leading labs such as OpenAI, DeepMind, and Anthropic to prevent human extinction as a result of the advent of ASI. AI Alignment doesn't try to control how powerful an AI becomes, it doesn't try to control exactly what the AI will be doing, and it doesn't even necessarily try to prevent a potential takeover. Since labs anticipate that they won't be able to control the superintelligent AI they're creating, they accept that such AI will act autonomously and unstoppably, but the solution lies in making it act in accordance with our values.

In view of the risks, perhaps we should ask ourselves: is it worth all the effort invested in moving towards ASI?

If we start from the optimistic assumption that the ISA of the future will have its interests linked to those of humanity, these could be some of its benefits:

- **Full integration into our daily lives:** smart homes that adapt perfectly to our needs and are respectful of the planet, personal assistants that help us with daily tasks and take care of people in vulnerable situations, devices that monitor our health to prevent diseases, the ASI could become an omnipresent and valuable protective companion.

- **Extension of human intelligence**: Systems that interconnect human collective intelligence (the most powerful supercomputer known) with a benign ASI could generate a framework for collaboration between humans and AI in artistic creation, scientific research, and technological innovation. This collaboration could lead to discoveries and creations that are unimaginable today. And if a seasoned reader finds it an impossible or undesirable hypothesis, Neuralink's story is sure to catch their attention. Neuralink is a company founded by Elon Musk and others in 2016, which develops Brain *Computer Interfaces (BCI)* technology, and aims to create implantable devices that allow direct communication between the human brain and computers. One of Neuralink's long-term goals is to facilitate the fusion of human intelligence with artificial intelligence, potentially enhancing human capabilities. The company is also focused on medical applications, such as the treatment of neurological diseases and the restoration of sensory and motor functions. Neuralink has achieved a major milestone by obtaining permission to begin human trials. Despite significant advances and successful animal demonstrations in recent years, such as the case of a monkey playing video games with only its eyes in 2021, the company has not yet begun recruiting participants for its clinical trials. Musk previously suggested that these chips could treat some forms of paralysis and insomnia.

- **Ethical and just governments**: An ISA driven by human values and goals of equity and social justice could play a more active role in governance and collective decision-making. This could include the management of public resources, urban planning and, possibly, participation in democratic processes. Public security could also benefit from

the existence of “artificial police”, equipped with advanced surveillance and rapid response systems and free from the biases that cause misunderstandings or even situations of police abuse against individuals from minorities or stigmatized social groups.

We could think of many other impacts of the advent of the ISA after the "singularity", but perhaps the best way to summarize them is to recognize that a humanity equipped with Artificial Superintelligence will be in an unbeatable position to overcome global challenges: climate change, pandemics, natural resource management, overpopulation, curing chronic diseases, etc. the generation of clean and inexhaustible energies, etc.

A hypothetical ASI could solve issues as complex as nuclear fusion at a commercial level, the definitive solution to obtaining inexhaustible and clean energy. Nuclear fusion uses hydrogen isotopes such as deuterium and tritium as fuel, which are relatively abundant and can be extracted from sources such as seawater, does not produce long-lived radioactive waste, and does not emit greenhouse gases. But the panacea of nuclear fusion requires long periods of development and research to solve problems that pose major technical challenges, such as the challenge of keeping superhot plasma stably.

ASI's ability to process large volumes of data and perform complex simulations could accelerate research in areas such as plasma confinement and optimizing conditions for fusion. In addition, the ISA could design new materials or methods to improve the energy efficiency of fusion reactors.

An ISA that could enable humanity to access an inexhaustible source of clean energy would boost economic growth and innovation in various industries, eliminate energy dependency and related geopolitical conflicts, and solve the serious problem of negative environmental impacts associated with current power generation.

No doubt the existence of an ISA will force us to reconsider many of our current ideas about intelligence, conscience and ethics. Coexistence with advanced forms of AI could profoundly influence our understanding of the human being and our place in the universe, even the very definition of what we consider an intelligent being.

Conclusion

We've explored the changing landscape of Artificial Intelligence, covering its recent advancements, present and future challenges, and the promises it holds for tomorrow. AI, in a short time, has evolved from a scientific ambition to a reality that permeates our daily lives, influencing multiple aspects of our daily existence.

At this early stage of the AI Age, as has been the case in the past with other disruptive innovations, the community of researchers and experts is divided between skeptics, enthusiasts, and apocalypticists. The reader will be able to choose which position he feels most identified with, but the history of mankind shows that there is no research challenge in any field of science or technology that has been abandoned for fear of the use that could be made of a transcendental discovery.

Scientific curiosity is an unstoppable force for progress, the pursuit of knowledge and understanding of the world is a fundamental driver of human advancement. Throughout history, scientific curiosity has driven discoveries and innovations that have transformed our lives, from understanding the laws of physics to the development of modern medicine. This curiosity not only satisfies our desire to understand, but also leads us to practical solutions to problems, continuously improving the quality of life and expanding our capabilities and horizons.

AI's race to the "singularity" has already begun, and regulations that governments may impose soon are not likely to stop it, or even slow it down. Recognizing that this is not a risk-free journey, just as other pivotal moments in humanity's technical and scientific progress were not, let us allow ourselves to dream of a distant future where Artificial Superintelligence (AIS) could radically improve our society, the way we live and work, and even our understanding of the universe.

This journey through the world of Artificial Intelligence leaves us with a sense of wonder at its potential, but also with an awareness of the responsibility that comes with its development. AI, like any powerful tool, presents both opportunities and challenges. Its future will depend not only on technological advances, but also on ethical, political and social decisions.

It is crucial that the development of AI is guided by a collaborative, ethical, and conscious approach that seeks the benefit of all humanity and respects the core values of our society.

In conclusion, the future of Artificial Intelligence is full of exciting possibilities and significant challenges. As a society, we have an opportunity and responsibility to shape this future, ensuring that AI is developed in ways that maximize its benefits and minimize its risks to the well-being of all.