FROM THE HEART OF THE STATE

The Governorate tells its story

Year 2 Vatican City Number 3

QUARTERLY JULY-SEPTEMBER 2025

Published by the Governorate of Vatican City State

Institutional Communication
00120 Vatican City State
(Vatican City State)

Email: comunicazione@scv.va
Website: www.vaticanstate.va

X (Twitter): Governatorato_SCV
Instagram: Governatorato_SCV

Editor: Nicola Gori
Graphics and layout: Antonio Coretti

Publisher: Governatorato dello Stato della Città del Vaticano

NOT JUST TECHNOLOGY

Artificial Intelligence has long since ceased to be science fiction. Today it drives processes, shapes decisions, and supports people and organizations in increasingly tangible contexts. Software that produces content, models that evolve autonomously, tools capable of suggesting solutions even before a problem is fully articulated. A powerful transformation – yet not always easy to interpret.

From this complexity, this newsletter takes shape. Its aim is to offer a compass in an ecosystem crowded with announcements, hype and oversimplified narratives, helping to distinguish what truly matters from what is merely temporary. This newsletter positions itself as a space for critical reading and in-depth exploration, dedicated to those who wish to navigate AI with awareness.

It hosts contributions designed to provide relevant information, explained in clear language, with concrete examples of application and reflections on the consequences of Artificial Intelligence for society, the economy and culture. No specialized technical background is required – only attention, curiosity, and a desire to understand.

Because AI is not just a technology: it is a lens through which we are redefining our relationship with the future. And the way we talk about it today will help determine how we live it tomorrow.

Happy reading.

Nicola Gori
The Editorial Team

ARTIFICIAL INTELLIGENCE: A TOOL AT THE SERVICE OF HUMANITY

Sr. Raffaella Petrini
President of the Governorate of the Vatican City State

Dedicating an edition of the newsletter to Artificial Intelligence is more important than ever at a time when it is increasingly developing and spreading widely.

On the other hand, as Pope Francis stated during his participation in the G7 session on Artificial Intelligence on Friday, 14 June 2024: "Science and technology are therefore extraordinary products of the creative potential of us human beings. Indeed, it is precisely from the use of this creative potential that God has given us that Artificial Intelligence comes to light."

In fact, it is an "extremely powerful" tool that is used in many fields: from medicine to the world of work, from culture to communication, from education to politics. It is clear that Artificial Intelligence will have a growing influence on our lives, on social relationships, and within relations among communities, institutions and nations.

Faced with such a complex and dynamically evolving reality, two attitudes emerge: enthusiasm for its potential and fear of its consequences. Without a doubt, Artificial Intelligence must be managed, not passively endured, and it must always aim to safeguard human dignity and serve the integral good. From this arise a series of ethical questions concerning its application. The debate cannot ignore the fact that it is not another human being.

In this regard, Pope Francis again pointed out to the members of the G7 that: "The algorithms designed to solve very complex problems are so sophisticated that even the programmers themselves find it difficult to understand exactly how they manage to achieve their results."

This must never be forgotten; otherwise, there is a risk of reducing the vision of the world "to realities expressible in numbers and enclosed in pre-packaged categories," imposing uniform models and lacking that creativity and discernment that are distinctive of humanity.

Indeed, human dignity, as Pope Leo XIV emphasized on Friday, 5 December, to participants in the conference "*Artificial Intelligence and Care of Our Common Home*," "resides in the capacity to reflect, to choose freely, to love gratuitously, and to enter into authentic relationships with others." In this sense, Artificial Intelligence has certainly "opened up new horizons for creativity, but it also raises troubling questions about its possible repercussions on humanity's openness to truth and beauty, on our capacity for wonder and contemplation."

For this reason, the newsletter can be an opportunity to reflect and to deepen our understanding of a reality that will shape our lives in the near future.

ADDRESS OF HIS HOLINESS POPE LEO XIV
TO PARTICIPANTS IN THE CONFERENCE
"ARTIFICIAL INTELLIGENCE AND CARE OF OUR COMMON HOME"

Organized by the Centesimus Annus Pro Pontifice Foundation and Strategic Alliance of Catholic Research University,
Consistory Hall, Friday, 5 December 2025

Dear brothers and sisters, welcome!

I am pleased to greet all of you, members of the *Centesimus Annus Pro Pontifice* Foundation and the *Strategic Alliance of Catholic Research Universities*.

We are meeting on the occasion of the publication of your research on a very important topic. The advent of Artificial Intelligence is accompanied by rapid and profound changes in society, which affects essential dimensions of the human person, such as critical thinking, discernment, learning and interpersonal relationships.

How can we ensure that the development of Artificial Intelligence truly serves the common good, and is not just used to accumulate wealth and power in the hands of a few? This is an urgent question, because this technology is already having a real impact on the lives of millions of people, every day and in every part of the world. As the Social Doctrine of the Church reminds us, and as is clear from the interdisciplinary work you are doing, addressing this challenge requires asking an even more fundamental question: What does it mean to be human in this moment of history?

Human beings are called to be co-workers in the work of creation, not merely passive consumers of content generated by artificial technology. Our dignity lies in our ability to reflect, choose freely, love unconditionally and enter into authentic relationships with others. Artificial Intelligence has certainly opened up new horizons for creativity, but it also raises serious concerns about its possible repercussions on humanity's openness to truth and beauty, and capacity for wonder and contemplation. Recognizing and safeguarding what characterizes the human person and guarantees his or her balanced growth is essential for establishing an adequate framework for managing the consequences of Artificial Intelligence.

In this regard, we must pause and reflect with particular care upon the freedom and inner life of our children and young people, and the possible impact of technology on their intellectual and neurological development. The new generations must be helped, not hindered, on their path to maturity and responsibility. The well-being of society depends on their ability to develop their talents and respond to the demands of the times and the needs of others, with generosity and freedom of mind. The ability to access vast amounts of data and information should not be confused with the ability to derive meaning and value from it. The latter requires a willingness to confront the mystery and core questions of our existence, even when these realities are often marginalized or ridiculed by the prevailing cultural and economic models. It will therefore be essential to teach young people to use these tools with their own intelligence, ensuring that they open themselves to the search for truth, a spiritual and fraternal life, broadening their dreams and the horizons of their decision making. We support their desire to be different and better, because never before has it been so clear that a profound reversal of direction is needed in our idea of maturing.

In order to build a future together with our young people that achieves the common good and harnesses the potential of Artificial Intelligence, it is necessary to restore and strengthen their confidence in the human ability to guide the development of these technologies. It is a confidence that today is increasingly eroded by the paralyzing idea that its development follows an inevitable path. This requires coordinated and concerted action involving politics, institutions, businesses, finance, education, communication, citizens and religious communities. Actors from these areas are called upon to undertake a common commitment by assuming this joint responsibility. This commitment comes before any partisan interest or profit, which is increasingly concentrated in the hands of a few. Only through widespread participation that gives everyone the opportunity to be heard with respect, even the most humble, will it be possible to achieve these ambitious goals. In this context, the research carried out by *Centesimus-SACRU* represents a truly valuable contribution.

Thank you, dear friends, and I encourage you to continue your work with creativity, guided by Sacred Scripture and the Church's Magisterium. May the intercession of the Blessed Virgin Mary accompany you, and I impart my Apostolic Blessing upon all of you.

L'Osservatore Romano, Daily Edition, Year CLXV n. 280, Friday 5 December 2025, p. 2.

ARTIFICIAL INTELLIGENCE: THE VATICAN CHARTS A CONSCIOUS PATH TOWARD THE FUTURE

Antonino Intersimone
Director of Telecommunicatios and Information Systems of the Governorate

We are living in an extraordinary moment in human history. Artificial Intelligence, – a technology that until recently seemed confined to pure imagination – is now fully integrated into our daily lives. We encounter it on our smartphones, on our PCs, and in the services we use every day. It is a powerful force and, like all powerful forces, it requires careful and informed governance. It is precisely for this reason that the Governorate of the Vatican City State has decided to face this challenge with determination and awareness.

In recent months, the Governorate has developed and adopted a coherent set of guidelines designed to ensure the responsible and ethical use of Artificial Intelligence. These guidelines, officially promulgated with Decree No. DCCII of 16 December 2024, entered into force on 1 January 2025, marking the beginning of a new era of technological awareness for the Vatican State. They are not merely formal documents, but the concrete result of deep and shared work, born of the awareness that it was by then necessary – and urgent – to define a clear, precise, and authoritative framework for the use of this extraordinary technology.

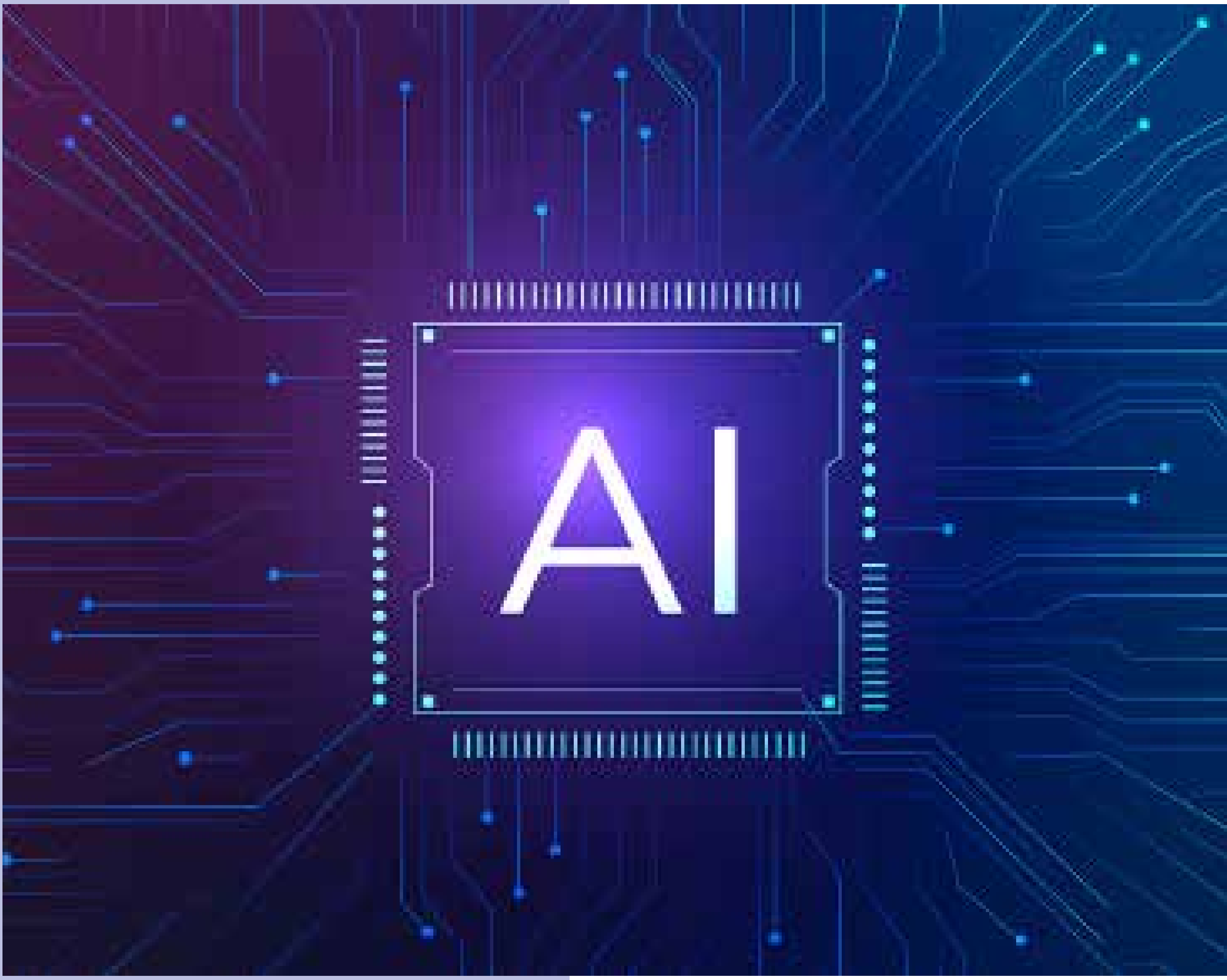
The Vatican guidelines start from a fundamental principle: Artificial Intelligence has extraordinary potential, capable of offering innovative solutions to complex problems and of significantly improving the quality of our lives. However, this potential must be carefully balanced with the duty to protect the fundamental rights of every person. It is not simply a matter of saying "yes" or "no" to technology, but of seeking a conscious path – a middle way – that allows us to harness the best that Artificial Intelligence can offer while safeguarding our most important values.

In other words, the Vatican guidelines do not present themselves as an obstacle to technological development, but rather as a guiding "compass." The image of the compass is particularly meaningful: just as a compass helps a traveler stay on the right course without preventing movement, so these guidelines are intended to guide the use of Artificial Intelligence toward a more conscious and humane future, without paralyzing innovation. The goal is clear: to expand human capabilities through technology, without ever losing sight of the central value of the person.

This approach explicitly recognizes the concept of the irreplaceability of the human being. As Pope Leo XIV warned in his message to the participants in the International Congress "*Artificial Intelligence and Medicine: the Challenge of Human Dignity*" (10–12 November 2025), organized by the Pontifical Academy for Life: "*The objective of providing care for individuals emphasizes the irreplaceable nature of human relationships in this context. Medical professionalism, in fact, requires not only the necessary specific expertise, but also the ability to communicate and be close to others. It can never be reduced merely to solving a problem. Similarly, technological devices must never detract from the personal relationship between patients and healthcare providers. Indeed, if AI is to serve human dignity and the effective provision of healthcare, we must ensure that it truly enhances both interpersonal relationships and the care provided.*".[1]

The guidelines address a concrete risk: Artificial Intelligence systems are potentially capable of perpetuating and amplifying biases already present in the data on which they are trained. If in the past recruitment processes were characterized by discrimination related to gender, origin, or other personal characteristics, an algorithm trained on such historical data could simply replicate the same injustices. The Governorate emphasizes the importance of actively preventing these potential discriminations, ensuring that every selection procedure is not only efficient but also fair.

For this very reason, the guidelines also prescribe that the use of Artificial Intelligence must not influence or limit the decision-making authority of administrators responsible for the organization and coordination of personnel. In other words, even when technology provides support, it is always a person – with their responsibility, discretion, and judgment – who makes the final decision. Technology proposes; human beings decide.

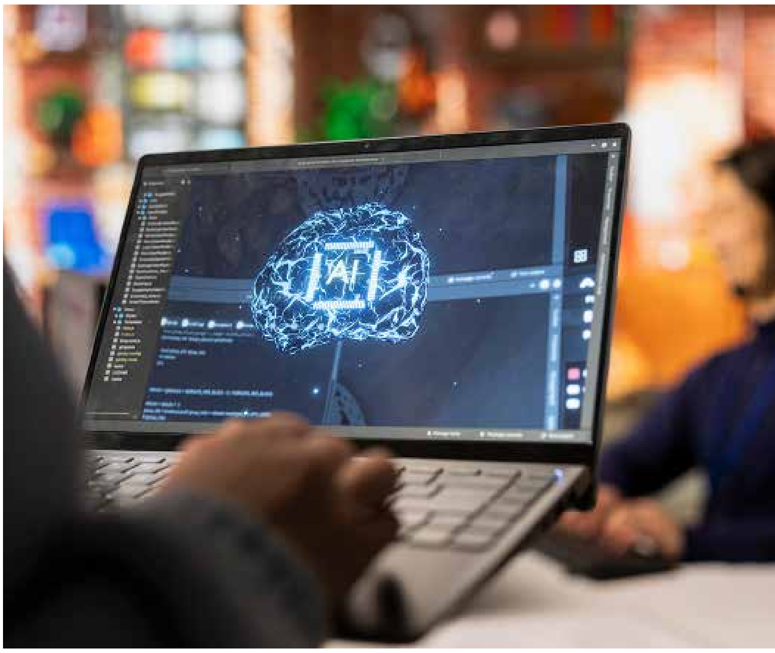


The guidelines honestly acknowledge a fact that is often underestimated: despite its extraordinary capabilities, Artificial Intelligence can cause concrete harm. These are not theoretical fears, but real and documented risks. An imperfect AI model can cause discrimination, violate fundamental rights, and unjustly harm the individuals concerned. For this reason, the guidelines devote special attention to the need to protect every individual from these potential harms. This protection operates on several levels. On the one hand, there is a requirement for transparency and awareness: people must know when Artificial Intelligence is influencing decisions that concern them. On the other hand, there is recognition that certain areas of human life – justice, fundamental rights, dignity – can never be fully entrusted to algorithms. Finally, there is a commitment to continue monitoring, updating, and improving these guidelines as technology evolves and new challenges emerge.

The guidelines represent a foundational document for the ethical approach that underlies them. This approach explicitly recognizes a concept that can sometimes be forgotten amid technological enthusiasm: the irreplaceability of the human being. No algorithm, however sophisticated, can fully replace human wisdom, empathy, or the ability to understand context and the deeper meaning of a situation.

For this reason, they stand as a crucial ethical and regulatory reference for the responsible integration of Artificial Intelligence. They represent an important first step toward a future in which technology truly serves humanity, rather than the other way around. It is a lesson that could – and should – be carefully considered by other institutions, governments, and organizations around the world.

With technology evolving so rapidly, the possible scenarios are too varied, and the challenges yet to be discovered are probably innumerable. In such a context, the guidelines do not claim to have all the answers; rather, they define and establish fundamental principles. They establish that justice must remain human. They establish that transparency is a right, not a privilege. They establish that human dignity is non-negotiable, even in the face of the most fascinating technologies.



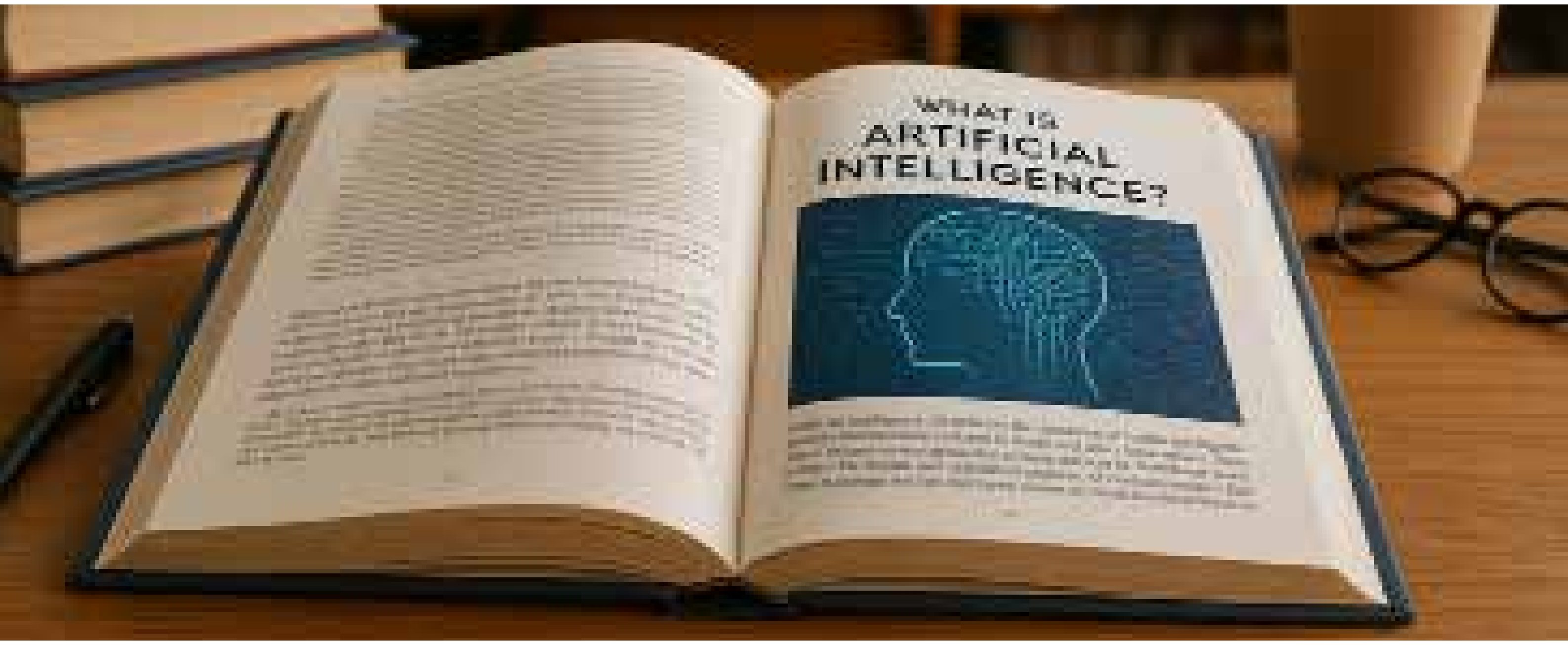
The Vatican decree represents an important model for the responsible regulation of Artificial Intelligence at a global level. It does not aspire to be a perfect instrument. It was developed and implemented with a careful, reflective attitude and – perhaps most importantly – with a firm conviction that technology must serve humanity, and not vice versa. This is perfectly in line with what was emphasized in the message to the *AI for Good Global Summit* of 10 July 2025, in which the Holy Father had already called for Artificial Intelligence to be placed at the service of all humanity, recalling the need to promote the *tranquillitas ordinis*: *"Ultimately, we must never lose sight of the common goal of contributing to that "tranquillitas ordinis – the tranquility of order", as Saint Augustine called it (De Civitate Dei) and fostering a more humane order of social relations, and peaceful and just societies in the service of integral human development and the good of the human family."*[2]

[1] https://press.vatican.va/content/salastampa/it/bollettino/pubblico/2025/11/10/0854/01523.html

[2] https://www.vaticanstate.va/it/novita/2317-messaggio-di-sua-santita-leone-xiv-all-ai-for-good-global-summit-2025.html

WHAT IS ARTIFICIAL INTELLIGENCE? A SIMPLE EXPLANATION TO TRULY UNDERSTAND IT

Davide Giordano
Member of the Commission for Artificial Intelligence of the Governorate



Over the past few years, the term "*Artificial Intelligence*" has suddenly burst into the lives of all of us. We hear it on television, in the news, on social media, and of course even at dinner with friends. But what is Artificial Intelligence really, for most of us? From a technical and technological point of view, it is certainly something extremely sophisticated and impenetrable, accessible only to a few. For everyone else, the story might be slightly different.

AI is often described as "an assistant" that never sleeps, never gets tired, knows millions and millions of pieces of information, and can answer your questions in just a few seconds. This assistant is not a real person; it has no body (or at least, for now, it doesn't have one made of organic tissue like ours). It exists only as a program inside a computer. But it knows many things and can perform many useful tasks. This, in essence, is the basic concept of what we call Artificial Intelligence.

Artificial Intelligence is, in very simple terms, a computer program that has been trained to do intelligent things. It is not magic, it is not a mystery: it is simply software, exactly like the programs we use on our phones or computers – except that this software has been trained in a special way so that it can understand what we ask it and give us useful answers.

As children, we learned about the world around us through imitation (of our parents) or through examples. To recognize colors, someone would show us a red object and say, "This is red," and we would learn. After seeing many red objects, our brain understood the concept of "red" so well that when we saw a new object of that color, we recognized it immediately without anyone telling us. Artificial Intelligence works in a similar way. Programmers teach the computer to recognize certain patterns or models by showing the program millions of examples. After seeing millions of examples, the computer learns the concept and is able to recognize it even when something new is presented to it.

This is the real secret behind Artificial Intelligence: learning, or training. But how exactly does a program learn? It's not as complicated as it might seem. The process is called *training*, and here's how it works in a simplified way.

Let's imagine teaching a child to distinguish a cat from a dog. We don't show them just one cat and one dog a single time. We show them many cats and many different dogs: gray cats, orange cats, white cats, big cats, small cats; white dogs, brown dogs, small dogs, huge dogs. After seeing hundreds of cats and hundreds of dogs, the child understands what a cat is like and what a dog is like. When they see a new animal, they can say whether it's a cat or a dog without being told.

Artificial Intelligence works in exactly the same way. Developers show the program millions of images of cats and millions of images of dogs. They tell the program, "This is an image of a cat," or "This is an image of a dog." The program observes all these images, finds common patterns, and identifies the characteristics that make a cat a cat and a dog a dog. After examining all these examples, if you show it a new photo that the program has never seen before, it can still say whether it is a cat or a dog.

This learning process requires enormous amounts of data and enormous computing power. That's why companies that create advanced Artificial Intelligence invest huge sums of money. They need very powerful computers that require large amounts of energy for processing and cooling, as well as enormous quantities of data to train their programs.

Some people believe that Artificial Intelligence is something that will only be used in extraordinary circumstances. In reality, it is already used several times a day without us even realizing it.

When we write a message on our phone, the phone automatically suggests the next word we might want to type. That is Artificial Intelligence. The program has learned from billions of messages written by people which words usually go together. So when we write "Hi, how," it knows that the next word will probably be "are."

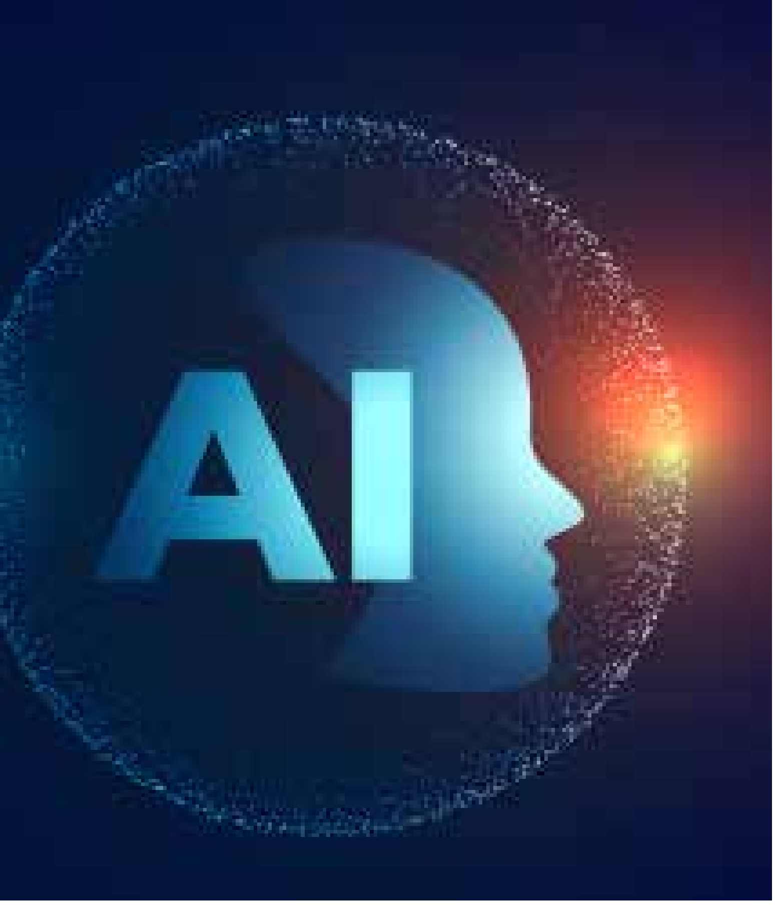
When we turn on facial recognition on our phone, the system that recognizes our face and unlocks the phone is Artificial Intelligence. It has been trained on millions of different faces to learn how to recognize the unique features of our own face.

When we use Google or another search engine, the program that decides which results to show us is Artificial Intelligence. It has learned from billions of previous searches to understand which result is most likely the one we are looking for.

When we watch a TV series on Netflix and the service suggests which episode to watch next, once again it is Artificial Intelligence. When the GPS navigator suggests the fastest route, it is Artificial Intelligence. When a virtual assistant like Alexa or Siri understands what you say and responds, it is Artificial Intelligence.

Here we come to an important distinction. Not all Artificial Intelligence is the same. In fact, experts divide Artificial Intelligence into two main categories, and understanding this difference helps us better understand how this technology works.

The first type is called "*Narrow Artificial Intelligence*" (*narrow AI*, also known as weak AI), and it is exactly what we have described so far. It is a program that is very good at doing one specific thing. It excels at recognizing faces, or analyzing traffic, or recommending movies – but it only does that one thing. If we ask it to do something different, it may not be able to do it as well as a human being.



The second type is called "*General Artificial Intelligence*" (AGI, also known as strong AI or superintelligence[1]), and it is something very different. It would be a program capable of doing anything, just like a human being. It could recognize faces, write poetry, solve complex math problems, drive a car, cook a delicious meal – everything perfectly. This type of Artificial Intelligence does not yet exist. Scientists are still working to theorize it and then create it, and many doubt that it is even possible. For now, all we have is narrow Artificial Intelligence, meaning programs that are very good at one specific task.

When we read in the newspapers about a new breakthrough in Artificial Intelligence, it is almost always narrow AI: a program that has been trained to do one particular thing very well.

To truly understand what AI is, it is also essential to understand what it cannot do and what its limits are.

First, Artificial Intelligence has no consciousness – neither now nor ever. It does not know it exists, it has no feelings, and it has no desires. It is a tool, exactly like a calculator. A calculator is excellent at doing math, but it has no awareness of what it is doing. The same is true of Artificial Intelligence. It does not know it is helping people. It does not know its work is important. It simply carries out the tasks it has been programmed to perform.

Second, Artificial Intelligence is not creative in the true sense of the word. If we learn that certain colors go well together based on millions of images, we can combine those colors in a new way. But we are not really creating something radically new – we are just recombining patterns we have already seen. A human artist, on the other hand, can create something completely new that the world has never seen before. They can create from inspiration, from emotion, from a deep inner impulse, or even from a mistake – think of how the *tarte Tatin* was invented.

Third, current Artificial Intelligence cannot make a true logical leap. If a new concept is completely different from everything it has seen during training, it will probably fail to understand it. A child, on the other hand, can make these logical leaps and understand new concepts by approaching a problem from different perspectives.

Fourth, Artificial Intelligence can be easily fooled. By slightly altering an image of a cat so that the program no longer recognizes it as a cat – while a person would still clearly see a cat – the computer becomes confused. This is an important limitation: Artificial Intelligence does not "understand" the world in the same way the human brain does.

Artificial Intelligence is a tool, nothing more and nothing less. Just like a hammer: if thrown with enough force it can fly, but it will not become an airplane; it can be a useful tool for building a beautiful piece of furniture, or it can be misused. AI is a powerful technology that can solve many problems and make our lives more comfortable and efficient. But it is not magic, it is not conscious, it is not omniscient, and it is not omnipotent.

[1] Neil Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014)

ARTIFICIALE INTELLIGENCE: BETWEEN PROMISES AND CONCRETE DANGERS

Andrea Tripoli
Member of the Commission for Artificial Intelligence of the Governorate

Every major technological innovation carries a dual nature: on the one hand, it offers extraordinary opportunities to improve people's lives; on the other, it introduces risks and vulnerabilities that often emerge only once the technology has already spread on a large scale. Artificial Intelligence is no exception to this universal rule. When a technology becomes accessible to everyone and reaches every corner of the planet, its beneficial potential multiplies – but so do the dangers. The ease with which Artificial Intelligence can be used makes it an extremely powerful tool in the hands of both those who wish to do good and those who intend to exploit it for illicit or harmful purposes. Awareness of this duality represents the first step toward responsible and safe use.

Artificial Intelligence deeply fascinates many people. Programs that learn on their own, recognize faces in a crowd, write complex texts, or drive cars without a human driver – all of this seems almost magical. However, behind this apparent technological perfection lies a less comforting truth. Artificial Intelligence algorithms can make significant errors, and these errors are particularly dangerous because they do not appear to be errors. They look like rational, objective, scientific decisions made by a machine supposedly free of emotions and bias.

Consider the case of a candidate applying for a job interview where the selection process is managed by an Artificial Intelligence system. The program evaluates résumés, skills, personality, and potential through complex algorithms. On

the surface, this seems like a fair and impartial process. Yet the system may make mistakes that no one notices immediately. Perhaps it fails to correctly recognize an academic qualification because the document scan is imperfect. Or, more seriously, the program may have been trained using historical hiring data that reflected discrimination based on gender, race, or geographic origin. In this case, the system learns and replicates these biases, automatically excluding perfectly qualified candidates simply because they do not match the dominant historical pattern.

This is not a hypothetical scenario. In 2018, Amazon developed an advanced Artificial Intelligence system intended to screen candidates for technical positions. The program was trained by analyzing data from past hiring decisions. Because men had historically been hired predominantly for technical roles, the algorithm learned that being male was a favorable trait and systematically penalized female candidates. The system discriminated against women not because it had been explicitly programmed to do so, but because it had learned from historical data that already reflected existing discrimination. Amazon was forced to abandon the project, but the episode remains a troubling warning.

The greatest danger lies precisely here: when a computer makes a discriminatory decision, that decision is perceived as objective and scientific. It is difficult to challenge, difficult to prove wrong. If there was discrimination in the data used to train the Artificial Intelligence in the past, the program will learn that discrimination and perpetuate it into the future – this time cloaked in the apparent authority of science and technology.

Another fundamental problem of Artificial Intelligence concerns transparency – or rather, its absence. There is a concept in the technological world known as the "black box": data go in, results come out, but no one knows exactly what happens in between. The decision-making process that led to a specific outcome is unknown. When Artificial Intelligence makes decisions that profoundly affect people's lives, this opacity becomes a serious ethical and practical problem.

Suppose a person applies for a bank loan and the bank uses an Artificial Intelligence system to evaluate the application. The system analyzes dozens of variables: income, employment history, area of residence, spending habits, previous payment history. In the end, the system denies the loan. When the person asks for an explanation, the bank itself is unable to provide a precise answer. The program has simply decided that the risk was too high, but even the technicians who manage the system do not know exactly which factors weighed most heavily in the decision, in what proportion, or according to what logic. The decision has been made, but it remains inexplicable.

This creates a deep, intrinsic injustice. In a democratic society founded on transparency and the right to due process, people have the fundamental right to know why decisions that directly affect them are made. They have the right to understand the reasoning, to challenge potential errors, and to present new information that could change the outcome. But with the black box of Artificial Intelligence, this right is systematically violated. People must accept decisions they cannot understand, based on criteria they cannot know, and against which they have very few effective means of defense.

Artificial Intelligence learns from data. This simple statement hides a complex and serious problem: if the data used for training are flawed, incomplete, or reflect existing biases, Artificial Intelligence will also learn incorrect things and replicate those biases on an even larger scale. In 2016, the COMPAS system, used in U.S. courts to predict the likelihood that a convicted person would reoffend after release, was found to systematically discriminate against African American citizens. The program assigned significantly high-

er recidivism risk scores to Black defendants than to white ones, even when all other factors were identical. The reason was clear: the system had been trained on historical data that reflected decades of racial profiling in the American criminal justice system.

The problem becomes even more insidious when it is understood that Artificial Intelligence does not merely replicate existing biases – it amplifies them. It functions like an amplifier that takes society's flaws and enlarges them, making them more pervasive and harder to identify and correct. If the police have historically stopped people of a certain race or from a certain neighborhood more frequently, an Artificial Intelligence system trained on this data will learn that those people are "more suspicious" and will suggest stopping them even more often, creating a self-reinforcing cycle of prejudice.

While on the one hand Artificial Intelligence can make discriminatory errors unintentionally, on the other hand it can be deliberately used as a tool for deception and manipulation. The digital security landscape has been dramatically transformed by this technology, offering cybercriminals extremely powerful tools to orchestrate increasingly sophisticated attacks.

Email scams, known as *phishing*, have undergone a striking evolution. The clumsy messages full of grammatical and spelling errors that characterized scams in the past are now obsolete. Generative Artificial Intelligence makes it possible to create flawless communications in any language, completely indistinguishable from authentic ones. Scammers can generate emails that appear to come from banks, employers, government services, or everyday platforms. The level of personalization is alarming: algorithms analyze social media profiles, online shopping habits, and personal interests to craft tailor-made messages that exploit the specific vulnerabilities of each individual.

At the same time, *smishing–phishing* via SMS–has become just as dangerous. Messages warning of package deliveries, urgent bank account issues, bonuses to be claimed, or fines to be paid arrive at the most opportune moments, written perfectly, exploiting urgency and the instinctive trust many people place in phone communications. The brevity of the SMS format, which once made it harder to create convincing scams, is no longer an obstacle for Artificial Intelligence.

But the threats go far beyond written messages. Audio and video *deepfakes* represent an even more unsettling and dangerous frontier. With just a few seconds of voice recording, Artificial Intelligence can perfectly clone anyone's voice. A call from one's supervisor urgently requesting an unauthorized wire transfer, or from a distressed family member begging for immediate financial help – distinguishing the real from the fake becomes virtually impossible without independent verification. Cases of multimillion-dollar scams orchestrated using this technology have already been documented, and the phenomenon is rapidly expanding.

Even traditional phone scams, known as *vishing*, have evolved dramatically. Artificial Intelligence – powered *chatbots* can conduct natural conversations, answer questions in real time, display apparent empathy, and build trust – all without any human intervention. These systems operate around the clock, can call thousands of people simultaneously, and dynamically adapt their approach based on the responses they receive, becoming increasingly convincing as the conversation progresses.

Faced with this complex landscape of risks and threats, it is essential to clarify that the problem is not Artificial Intelligence itself, but how it is developed, distributed, and used. The technology itself is neutral; it is human choices that make it beneficial or harmful. The good news is that Artificial Intelligence is also used to protect people: modern security systems analyze billions of data points to identify suspicious patterns, block cyberattacks in real time, and predict new threats before they materialize. A true technological arms race is underway between those who use Artificial Intelligence to attack and those who use it to defend.

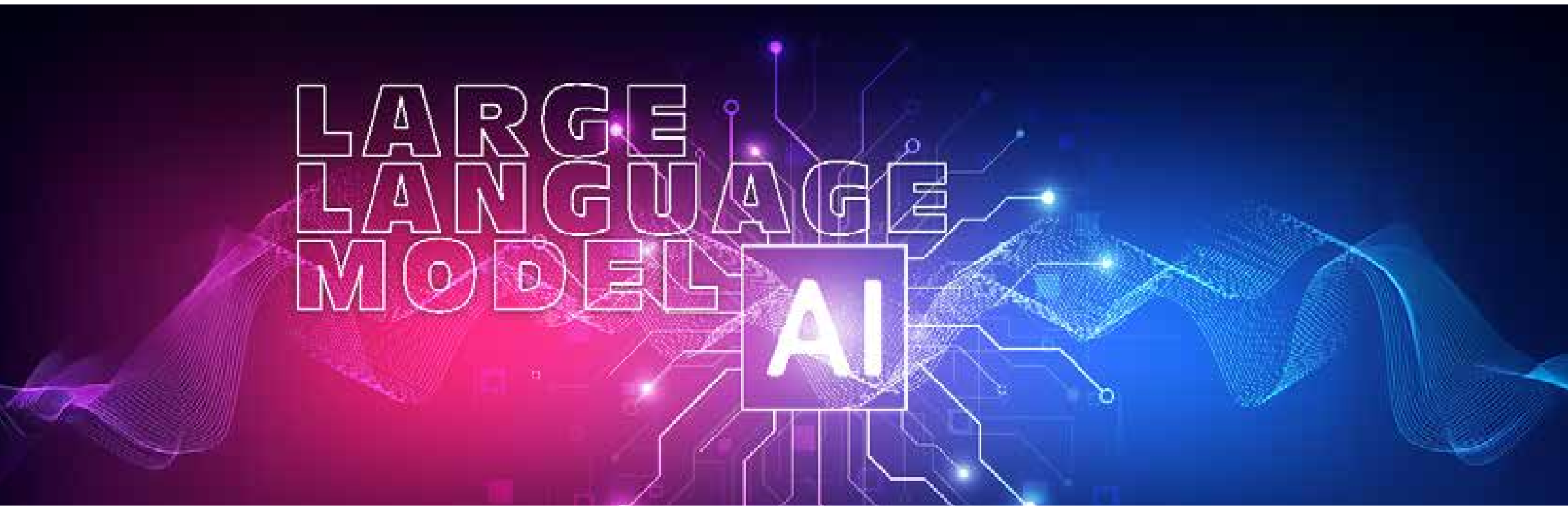
Awareness represents the first and most important line of defense. Understanding that Artificial Intelligence can be used to create perfectly convincing scams means develop-

ing healthy skepticism toward seemingly official communications. Knowing that algorithms can discriminate means demanding transparency in automated decisions that affect us. Recognizing that systems can be trained on biased data means insisting on rigorous checks before they are implemented in critical sectors such as justice, healthcare, or finance.

Always verifying the identity of those who contact us through independent channels, being wary of urgent requests that demand immediate action, carefully checking email addresses and links before interacting, and enabling strong authentication systems–these practices become essential in a world where Artificial Intelligence can perfectly imitate any person or institution. Individual critical thinking, combined with strong regulation and robust oversight systems, represents the only truly effective defense against the risks of Artificial Intelligence. Ultimately, technology must remain a tool in the service of humanity, controlled by humanity, and must not turn into an uncontrollable force that autonomously decides people's fate.

LARGE LANGUAGE MODELS: FROM CHOOSING EXISTING ONES TO INVESTING IN CUSTOM MODELS

Domenico Vetere
Deputy Chief of Internet Provider Services in the Governorate



When we talk about Artificial Intelligence in today's organizational context, we increasingly find ourselves dealing with so-called Large Language Models (better known as LLMs). These systems have begun to transform the way we work, communicate and process information. However, understanding what they are, why there are different ones, how to choose among them, and – above all – why one should invest in customized models remains rather unclear in most organizations. This reflection arises from the need to clarify a crucial point: not all LLMs are the same, and the choice between using an existing model and investing in the creation of a proprietary one represents a fundamental strategic decision for any modern institution. A *Large Language Model* is an Artificial Intelligence system trained on enormous amounts of text (datasets), capable of understanding natural language and generating coherent, contextualized, and ideally accurate responses. Its "size" – the term *large* – refers both to the volume of data on which it is trained and to the complexity of the model itself.

In recent years, the market has therefore become populated with LLMs that are very different from one another. This did not happen by chance. Various organizations (OpenAI with ChatGPT, Google with Gemini, Meta with Llama, Anthropic with Claude) have invested heavily in the creation of general-purpose models because they represent a significant economic and strategic opportunity. Each of these actors has made different choices: some have opted for open access, others keep their models proprietary, while still others offer free versions alongside paid ones. The reason for this proliferation is simple: there is no single LLM that is universally optimal for all contexts. A model that works excellently for literary translation may not be ideal for analyzing healthcare data. A system robust for creative writing may not be reliable for handling confidential information. And a model trained on public data, however sophisticated, may not adequately understand specialized languages, internal protocols, or the cultural nuances of a specific organization.

If we decide to use an *off-the-shelf* LLM (that is, one that already exists and is available on the market), we are faced with a variety of options, each with distinct characteristics.

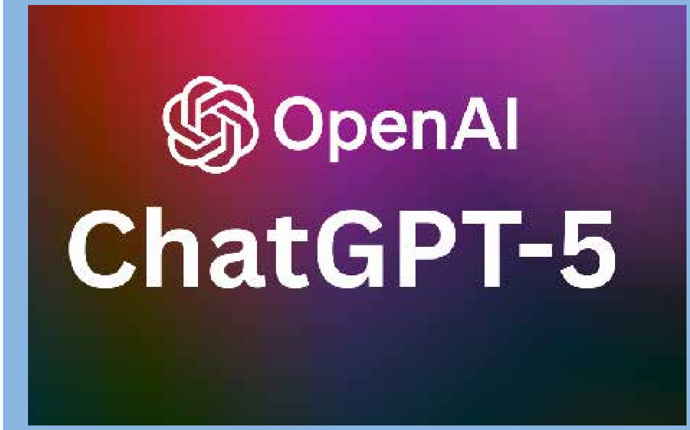
GPT-5 by OpenAI, widely known as **ChatGPT**, is probably the most famous. It is extremely versatile, capable of tackling a very broad range of tasks, from text and image generation to solving complex problems. Its disadvantages are that it is proprietary, requires a permanent internet connection, and does not guarantee complete confidentiality of the data entered (which are used to improve the model itself).

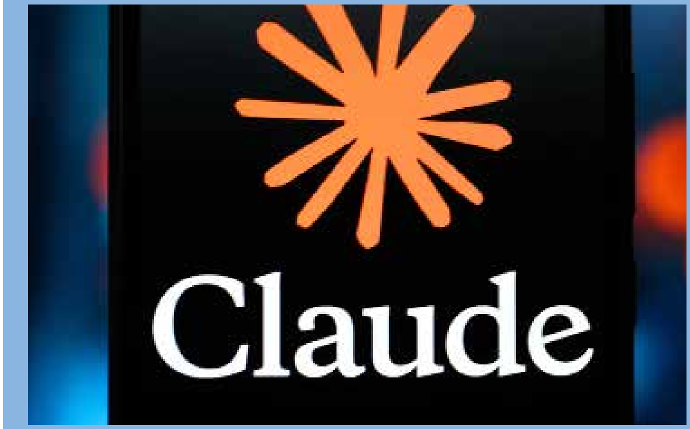
Google's Gemini offers similar features, with the added advantage of tight integration into Google ecosystems and strong multimodal processing capabilities (text, images, video). In this case as well, however, we are tied to Google's cloud infrastructure and its *privacy* policies.

Claude by Anthropic was developed with a *focus* on safety and reliability, particularly in the handling of sensitive information. It has been trained using methodologies that emphasize risk reduction and transparency. Nevertheless, it remains an external model, with the limitations that entails.

Meta's Llama represents an interesting alternative because it is available in an *open-source* version, meaning organizations can deploy it in *self-hosted* environments – on their own *servers* – retaining full control over data, in full compliance with the concept of "data sovereignty." However, implementing it correctly requires significant technical expertise.

How should one choose among these options? The criteria should include, for example, the type of tasks the model is expected to perform, the sensitivity of the data it will handle, the need to operate in *offline* or confidential environments, the available budget, internal technical expertise, and the desired level of control and customization. From this perspective, for many organizations – especially in an initial phase – using a well-established public LLM represents a pragmatic solution: it is quick to implement, relatively inexpensive, and does not require significant investment in proprietary infrastructure. However, at this point in the reflection, we must confront an uncomfortable truth: generic LLMs, however sophisticated, have significant limitations when it comes to operating in specialized, confidential, or highly personalized contexts. Imagine, for example, wanting to use an LLM to support an office that manages complex administrative procedures based on specific regulations, institutional precedents, and interpretations unique to that organization. A generic LLM will have been trained on a vast variety of public texts, but it will never have seen the organization's internal documents; it does not know the procedures, nor does it understand the nuances of that specific organizational culture. The risk is twofold. First, the model may very likely provide





plausible but incorrect answers; it may share information it does not actually know, or interpret ambiguities in a way that does not reflect the organization's intent. Second, to achieve acceptable *performance* in terms of accuracy, one might be forced to input confidential documents, internal procedures, and proprietary information into public models, thereby effectively putting intellectual capital at risk. This second point is particularly critical. If an organization feeds its proprietary data into a public LLM to obtain better results, it is in effect relinquishing control over that information. Even if *privacy* clauses exist, the data will contribute to training the model, will likely be used by others, and will become part of a third party's technological infrastructure. For an institution such as a Governorate, or for any organization responsible for sensitive information, this represents an unacceptable risk.

This is why, for many organizations, investing in *custom large language models* – models created specifically for them, trained exclusively on proprietary data, and under total control – represents a strategic option that deserves serious consideration. A *custom LLM* is therefore a model that is "fine-tuned" (that is, adjusted and specialized), starting from an existing base model and further training it on specific data. This process has several significant advantages. First, intellectual property remains within the organization. The data used to train the model are not shared, do not contribute to public models, and remain **under custodianship**. This is crucial when dealing with information that represents competitive value, unique historical knowledge, or sensitive data. Second, *performance* can be dramatically superior. A model trained on specific documents, terminology, procedures, and communication styles will work infinitely better in a specialized internal context than a generic model. If an office manages complex legislation, a model trained on decades of precedents and interpretations specific to the organization will be incomparably more reliable. Third, it becomes possible to decide exactly which data to train the model on, which sources to exclude, how to interpret ambiguities, and which outputs are considered acceptable. Everything remains within the organization. The fourth aspect to consider is security, which becomes implicitly controllable. A *custom LLM* can be hosted on local servers, in *offline* environments if necessary, with *encryption* and control levels that are fully customizable. There is no transmission of data to external infrastructures. The fifth and final point concerns long-term sustainability. A public model may change, disappear, be acquired, or alter its *policies*. A private, *custom* model remains owned forever and can be scaled according to future needs.

However, the value of a custom LLM inevitably depends on the quality of the data used for its training. Here we encounter an aspect that is often underestimated: the need to invest in data curation and certification. As already noted, not all models are equal. If an LLM is trained on a disorganized, unverified dataset full of contradictions and inaccuracies, the resulting model will embody all of those problems. AI amplifies the *biases* present in its training data. If documents contain errors, the model will learn to replicate them. If they contain conflicting information, the model will become confused and unreliable. For this reason, a true investment in *custom LLMs* also requires a parallel investment in data management and certification – the so-called *clean or high-quality data*. This means identifying and selecting the most relevant and **high-quality** data. Not all historical documents may be appropriate: some may be obsolete, others may contain information that has been officially declared incorrect, and still others may be duplicated or redundant. Care is essential. It also means verifying accuracy and consistency. Before being used for training, data must undergo quality control.

This may involve experts manually reviewing samples of documents, resolving contradictions, and correcting obvious errors. With regard to sources and versions, it must be clear where the data come from, which version represents the official "truth," and how they have been transformed. This traceability is fundamental to the model's reliability. Furthermore, we must ask another set of questions: who has the authority to decide which data are included? How are updates handled? What is the lifecycle of a piece of data? These questions must have clear and documented answers.

At this point, is creating a *custom LLM* really worth the expense? The answer depends on the specific situation, but for many medium- and large-scale organizations, it is a resounding yes. The initial effort required to develop a *custom LLM* varies depending on domain complexity, data volume, and the desired level of customization. However, this effort must be considered in perspective. If a public LLM is used while continuously feeding it confidential data to achieve acceptable performance, a hidden cost is already being paid – the loss of control over intellectual property. If a generic model provides inaccurate answers and human resources must constantly correct errors, the cost is inefficiency. Relying on an external *provider* for every operation makes one vulnerable to *pricing* changes, service interruptions, and unilateral policy shifts. A *custom LLM*, by contrast, represents *an asset that grows in value over time*. As it is used, it can be continuously improved with new data. It becomes progressively more intelligent within the organization's specific context. And it remains, always and unequivocally, the organization's property.

As noted earlier, for many organizations – especially in an initial phase – the pragmatic use of carefully selected public models represents a reasonable path. However, once an organization discovers that AI is truly useful for its operations, that it uses it regularly, and that the quality of results is critical to its functioning, the calculation changes. At that point, investing in a custom LLM – trained on the organization's certified, proprietary, and clean data – is no longer a luxury, but a strategic necessity.

ARTIFICIAL INTELLIGENCE IN THE GOVERNORATE: TRAINING INITIATIVES

The disruptive growth of the phenomenon of *Artificial Intelligence* is unstoppable and undeniable. Artificial Intelligence represents a true revolution, a profound transformation that is simultaneously entering offices, organizations, schools, and families. It is precisely for this reason that the Governorate of the State of Vatican City has developed a conscious and responsible strategy to address this epoch-making challenge.

When Artificial Intelligence enters a society, it does not do so uniformly. Each context – workplace, educational, family – experiences this transformation in a distinct way, with specific challenges and particular opportunities. Understanding this diversity is essential in order to develop appropriate and informed responses.

In the workplace, the introduction of Artificial Intelligence requires attention on multiple fronts. It is essential to rigorously verify the authenticity of AI-generated content, while at the same time protecting human creativity, which remains irreplaceable. Emerging risks must also be addressed, such as "shadow AI," namely the use of unauthorized or uncontrolled Artificial Intelligence systems operating in the interstices of work processes. The adoption of careful checks on business processes and data protection – today increasingly critical and valuable assets for any modern organization – becomes indispensable.

In the educational and training context, the challenge is even deeper. It is necessary to completely rethink the traditional educational model: from curricula to teaching methodologies that have characterized education up to now. The educational community is legitimately questioning how Artificial Intelligence influences students' critical thinking skills. Some experts fear that the possibility of delegating research and analytical tasks to AI could impoverish young people's critical thinking, which is essential for the development of informed and autonomous citizens.

In the most intimate and delicate context – the family – Artificial Intelligence touches much more vulnerable aspects of society. Parents and adults find themselves using *chatbots* and virtual assistants to address issues that were traditionally the domain of psychologists, educational professionals, and specialists. This is happening at a time when the average age for a child's first exposure to a *smartphone* is around ten years old. The crucial question therefore becomes: how can minors be protected when they are exposed to technologies they do not fully understand, and how can adults be guided toward the responsible use of these tools within the family context?

Faced with these multifaceted questions, one common response clearly emerges: it is necessary to create specific awareness through sensitization and literacy. In other words, training becomes the primary tool for transforming the relationship between people and technology.

The Governorate has recognized this need and, for more than a year now, has been investing significantly in specific training programs on Artificial Intelligence. These programs do not merely explain how to use tools; rather, they address complex and articulated topics. They include the history of the development of Artificial Intelligence, with its cycles of optimism and skepticism – the so-called "AI winters and springs." They address the intrinsic limits of the technology, realistic expectations, genuine potential, and the fallacies to which Artificial Intelligence is naturally prone.

The training catalogue is structured on multiple levels, each designed for different needs. The basic level is aimed at those who are taking their first steps in the world of Artificial Intelligence and wish to become familiar with terminology and application contexts. More advanced levels allow participants to engage directly with various platforms and different language models (LLMs). Through hands-on experience, participants begin to understand how to identify the most suitable tool depending on the specific objective they wish to achieve.

By its very nature, the subject matter means that training programs and content cannot be static. Given the accelerated pace of technological change, the training catalogue is continuously and constantly updated, always reflecting the dynamic nature of the field itself. This flexible approach ensures that training remains relevant and up to date.

The Governorate does not face the challenge of AI in isolation. Artificial Intelligence is a global phenomenon that requires coordinated solutions at the international level. For this reason, the Governorate is actively connected with international organizations capable of bringing together hundreds of countries in dialogue on critical technological issues.

Among these initiatives, participation in the *World Summit on the Information Society* (WSIS) and its recent developments is of particular importance. In this context, the Governorate is involved in initiatives such as *AI for Good*, an event and a platform whose goal is to identify technologies, methodologies and strategies aimed at the conscious and responsible adoption of Artificial Intelligence in any application context.

These global strategies necessarily address issues that are already well known and well documented. First among them is the *digital divide*, namely unequal access to technology between wealthy countries and developing countries. In addition, the issue of algorithmic *bias* is rigorously addressed, and above all the concrete danger of discrimination that could result from the absence of language models translated into all the world's languages. If Artificial Intelligence is trained predominantly in English and in a few other European languages, users of less represented languages could suffer significantly poorer performance, perpetuating and amplifying existing global inequalities.

Beyond training and international participation, the Governorate continues to invest in another crucial area: data sovereignty and responsible data management. This issue takes on particular importance if data are considered the "oil" of the twenty-first century – an extraordinarily valuable resource to be created, maintained, protected and enhanced.

For an entity such as the Governorate, control and protection of data sovereignty is not a secondary matter, but an essential characteristic. It is unacceptable for sensitive data concerning the operations of the State to be managed by external servers controlled by foreign private companies. It is therefore necessary to develop internal capabilities for data management and protection.

The solution that the Governorate intends to evaluate and adopt aligns with this approach: developing local, internal Artificial Intelligence solutions. This means equipping itself with appropriate *hardware* in terms of computing power, while also paying attention to energy implications and environmental sustainability. The goal is to process internal data autonomously, enhance their value, and make them usable according to the new paradigms of *generative AI*, without relying on external infrastructures.

This approach presents significant challenges. These are new, largely unexplored scenarios, rich in technical and organizational pitfalls. At the same time, however, they contain great opportunities. They represent an alternative model to the centralization of technological power in the hands of a few global actors. They demonstrate that it is possible to develop responsible, ethically aware, and sovereign Artificial Intelligence even for entities of more limited size.

Technology is not neutral: the choices made today will determine the landscape of tomorrow, and the responsibility for governing innovation rests with all actors in society, from institutions to individual citizens.

The Governorate is actively working to achieve a stable, contemporary configuration that is fully ready to face the era of Artificial Intelligence. This is not a task that can be delegated to a single part of the organization. It requires the contribution and informed collaboration of everyone. It requires that each person develop an adequate understanding of risks and opportunities, and that they act with integrity and ethical responsibility in their daily choices.

The vision is that it is absolutely possible to embrace Artificial Intelligence without sacrificing the fundamental values of sovereignty, ethics, equity, and transparency. Conscious training, global participation, and technological sovereignty are the foundations on which to build innovation.

D. G.

COMMISSIONS OF THE GOVERNORATE: INSTRUMENTS OF SPECIALIZED GOVERNANCE

The Commissions of the Governorate were established to support the Governing Bodies of Vatican City State in areas that require specific expertise, while ensuring transparency, collegial decision-making, and compliance with the ethical principles that characterize the State's actions. Over the decades, commissions have been created to address matters such as personnel, discipline, monetary issues, and the selection of lay collaborators, each with advisory, deliberative, or supervisory functions depending on its area of competence.

Within this framework of specialized governance falls the recent establishment of the Commission on Artificial Intelligence, provided for by Decree No. DCCII of 30 December 2024 and in force as of 1 January 2025. In light of the rapid spread of Artificial Intelligence systems and their growing influence on every aspect of social and institutional life, the Governorate deemed it necessary to establish a dedicated body to ensure the ethical, transparent, and responsible use of this technology.

The Commission is tasked with drafting laws and implementing regulations for the guidelines on Artificial Intelligence, issuing opinions on proposals for experimentation and application of AI systems, carrying out continuous monitoring by reporting potential risks and preparing semiannual reports on the impact of the use of Artificial Intelligence in the Vatican City State. In essence, the Commission serves as a guiding compass to ensure that technological development always remains at the service of human dignity and the common good, in full coherence with the fundamental values of the Holy See.

The Commission is composed of five members appointed by the President of the Governorate and is chaired by the Secretary General, as defined by Article 14 of the Decree. The members come from three strategic departments: the Legal Office, the Directorate of Telecommunications and Information Systems, and the Directorate of Security Services and Civil Protection. This multidisciplinary composition ensures that decisions take into account legal, technological, and security aspects. The mandate lasts three years and is renewable.

The Commission has already convened and is actively at work implementing the first operational steps. The promptness of its action reflects an awareness of the sensitivity of the subject and the importance of having a concrete framework in place from the outset, capable of effectively responding to the challenges that Artificial Intelligence poses on a daily basis to the Vatican administration.

D. G.