The Pope's G7 Speech on Artificial Intelligence 14 June 2024

'An exciting and fearsome tool'

Esteemed ladies and gentlemen,

I address you today, the leaders of the Intergovernmental Forum of the G7, concerning the effects of artificial intelligence on the future of humanity.

"Sacred Scripture attests that God bestowed his Spirit upon human beings so that they might have 'skill and understanding and knowledge in every craft' (*Ex* 35:31)".[1] Science and technology are therefore brilliant products of the creative potential of human beings.[2]

Indeed, artificial intelligence arises precisely from the use of this God-given creative potential.

As we know, artificial intelligence is an extremely powerful tool, employed in many kinds of human activity: from medicine to the world of work; from culture to the field of communications; from education to politics. It is now safe to assume that its use will increasingly influence the way we live, our social relationships and even the way we conceive of our identity as human beings.[3]

The question of artificial intelligence, however, is often perceived as ambiguous: on the one hand, it generates excitement for the possibilities it offers, while on the other it gives rise to fear for the consequences it foreshadows. In this regard, we could say that all of us, albeit to varying degrees, experience two emotions: we are enthusiastic when we imagine the advances that can result from artificial intelligence but, at the same time, we are fearful when we acknowledge the dangers inherent in its use.[4]

After all, we cannot doubt that the advent of artificial intelligence represents a true cognitive-industrial revolution, which will contribute to the creation of a new social system characterised by complex epochal transformations. For example, artificial intelligence could enable a democratization of access to knowledge, the exponential advancement of scientific research and the possibility of giving demanding and arduous work to machines. Yet at the same time, it could bring with it a greater injustice between advanced and developing nations or between dominant and oppressed social classes, raising the dangerous possibility that a "throwaway culture" be preferred to a "culture of encounter".

The significance of these complex transformations is clearly linked to the rapid technological development of artificial intelligence itself.

It is precisely this powerful technological progress that makes artificial intelligence at the same time *an exciting and fearsome tool*, and demands a reflection that is up to the challenge it presents.

In this regard, perhaps we could start from the observation that artificial intelligence is above all else *a tool*. And it goes without saying that the benefits or harm it will bring will depend on its use.

This is surely the case, for it has been this way with every tool fashioned by human beings since the dawn of time.

Our ability to fashion tools, in a quantity and complexity that is unparalleled among living things, speaks of a *techno-human condition*: human beings have always maintained a relationship with the environment mediated by the tools they gradually produced. It is not possible to separate the history of men and women and of civilization from the history of these tools. Some have wanted to read into this a kind of shortcoming, a deficit, within human beings, as if, because of this deficiency, they were forced to create technology.[5] A careful and objective view actually shows us the opposite. We experience a state of "outwardness" with respect to our biological being: we are beings inclined toward what lies outside-of-us, indeed we are radically open to the beyond. Our openness to others and to God originates from this reality, as does the creative potential of our intelligence with regard to culture and beauty. Ultimately,

our technical capacity also stems from this fact. Technology, then, is a sign of our orientation towards the future.

The use of our tools, however, is not always directed solely to the good. Even if human beings feel within themselves a call to the beyond, and to knowledge as an instrument of good for the service of our brothers and sisters and our *common home* (cf. *Gaudium et Spes*, 16), this does not always happen. Due to its radical freedom, humanity has not infrequently corrupted the purposes of its being, turning into an enemy of itself and of the planet.[6] The same fate may befall technological tools. Only if their true purpose of serving humanity is ensured, will such tools reveal not only the unique grandeur and dignity of men and women, but also the command they have received to "till and keep" (cf. *Gen* 2:15) the planet and all its inhabitants. To speak of technology is to speak of what it means to be human and thus of our singular status as beings who possess both freedom and responsibility. This means speaking about ethics.

In fact, when our ancestors sharpened flint stones to make knives, they used them both to cut hides for clothing and to kill each other. The same could be said of other more advanced technologies, such as the energy produced by the fusion of atoms, as occurs within the Sun, which could be used to produce clean, renewable energy or to reduce our planet to a pile of ashes.

Artificial intelligence, however, is a still more complex tool. I would almost say that we are dealing with a tool *sui generis*. While the use of a simple tool (like a knife) is under the control of the person who uses it and its use for the good depends only on that person, artificial intelligence, on the other hand, can autonomously adapt to the task assigned to it and, if designed this way, can make choices independent of the person in order to achieve the intended goal.[7]

It should always be remembered that a machine can, in some ways and by these new methods, produce algorithmic choices. The machine makes a technical choice among several possibilities based either on well-defined criteria or on statistical inferences. Human beings, however, not only choose, but in their hearts are capable of deciding. A decision is what we might call a more strategic element of a choice and demands a practical evaluation. At times, frequently amid the difficult task of governing, we are called upon to make decisions that have consequences for many people. In this regard, human reflection has always spoken of wisdom, the *phronesis* of Greek philosophy and, at least in part, the wisdom of Sacred Scripture. Faced with the marvels of machines, which seem to know how to choose independently, we should be very clear that decision-making, even when we are confronted with its sometimes dramatic and urgent aspects, must always be left to the human person. We would condemn humanity to a future without hope if we took away people's ability to make decisions about themselves and their lives, by dooming them to depend on the choices of machines. We need to ensure and safeguard a space for proper human control over the choices made by artificial intelligence programs: human dignity itself depends on it.

Precisely in this regard, allow me to insist: in light of the tragedy that is armed conflict, it is urgent to reconsider the development and use of devices like the so-called "lethal autonomous weapons" and ultimately ban their use. This starts from an effective and concrete commitment to introduce ever greater and proper human control. No machine should ever choose to take the life of a human being.

It must be added, moreover, that the good use, at least of advanced forms of artificial intelligence, will not be fully under the control of either the users or the programmers who defined their original purposes at the time they were designed. This is all the more true because it is highly likely that, in the not-too-distant future, artificial intelligence programs will be able to communicate directly with each other to improve their performance. And if, in the past, men and women who fashioned simple tools saw their lives shaped by them – the knife enabled them to survive the cold but also to develop the art of warfare – now that human beings have fashioned complex tools they will see their lives shaped by them all the more.[8]

The basic mechanism of artificial intelligence

I would like now briefly to address the complexity of artificial intelligence. Essentially, artificial intelligence is a tool designed for problem solving. It works by means of a logical chaining of algebraic operations, carried out on categories of data. These are then compared in order to discover correlations,

thereby improving their statistical value. This takes place thanks to a process of self-learning, based on the search for further data and the self-modification of its calculation processes.

Artificial intelligence is designed in this way in order to solve specific problems. Yet, for those who use it, there is often an irresistible temptation to draw general, or even anthropological, deductions from the specific solutions it offers.

An important example of this is the use of programs designed to help judges in deciding whether to grant home-confinement to inmates serving a prison sentence. In this case, artificial intelligence is asked to predict the likelihood of a prisoner committing the same crime(s) again. It does so based on predetermined categories (type of offence, behaviour in prison, psychological assessment, and others), thus allowing artificial intelligence to have access to categories of data relating to the prisoner's private life (ethnic origin, educational attainment, credit rating, and others). The use of such a methodology – which sometimes risks *de facto* delegating to a machine the last word concerning a person's future – may implicitly incorporate prejudices inherent in the categories of data used by artificial intelligence.

Being classified as part of a certain ethnic group, or simply having committed a minor offence years earlier (for example, not having paid a parking fine) will actually influence the decision as to whether or not to grant home-confinement. In reality, however, human beings are always developing, and are capable of surprising us by their actions. This is something that a machine cannot take into account.

It should also be noted that the use of applications similar to the one I have just mentioned will be used ever more frequently due to the fact that artificial intelligence programs will be increasingly equipped with the capacity to interact directly (*chatbots*) with human beings, holding conversations and establishing close relationships with them. These interactions may end up being, more often than not, pleasant and reassuring, since these artificial intelligence programs will be designed to learn to respond, in a personalised way, to the physical and psychological needs of human beings.

It is a frequent and serious mistake to forget that artificial intelligence is not another human being, and that it cannot propose general principles. This error stems either from the profound need of human beings to find a stable form of companionship, or from a subconscious assumption, namely the assumption that observations obtained by means of a calculating mechanism are endowed with the qualities of unquestionable certainty and unquestionable universality.

This assumption, however, is far-fetched, as can be seen by an examination of the inherent limitations of computation itself. Artificial intelligence uses algebraic operations that are carried out in a logical sequence (for example, if the value of X is greater than that of Y, multiply X by Y; otherwise divide X by Y). This method of calculation – the so-called "algorithm" – is neither objective nor neutral.[9] Moreover, since it is based on algebra, it can only examine realities formalised in numerical terms.[10]

Nor should it be forgotten that algorithms designed to solve highly complex problems are so sophisticated that it is difficult for programmers themselves to understand exactly how they arrive at their results. This tendency towards sophistication is likely to accelerate considerably with the introduction of quantum computers that will operate not with binary circuits (semiconductors or microchips) but according to the highly complex laws of quantum physics. Indeed, the continuous introduction of increasingly high-performance microchips has already become one of the reasons for the dominant use of artificial intelligence by those few nations equipped in this regard.

Whether sophisticated or not, the quality of the answers that artificial intelligence programs provide ultimately depends on the data they use and how they are structured.

Finally, I would like to indicate one last area in which the complexity of the mechanism of so-called Generative Artificial Intelligence clearly emerges. Today, no one doubts that there are magnificent tools available for accessing knowledge, which even allow for self-learning and self-tutoring in a myriad of fields. Many of us have been impressed by the easily available online applications for composing a text or producing an image on any theme or subject. Students are especially attracted to this, but make disproportionate use of it when they have to prepare papers.

Students are often much better prepared for, and more familiar with, using artificial intelligence than their teachers. Yet they forget that, strictly speaking, so-called generative artificial intelligence is not really "generative". Instead, it searches big data for information and puts it together in the style required of it. It does not develop new analyses or concepts, but repeats those that it finds, giving them an appealing form. Then, the more it finds a repeated notion or hypothesis, the more it considers it legitimate and valid. Rather than being "generative", then, it is instead "reinforcing" in the sense that it rearranges existing content, helping to consolidate it, often without checking whether it contains errors or preconceptions.

In this way, it not only runs the risk of legitimising fake news and strengthening a dominant culture's advantage, but, in short, it also undermines the educational process itself. Education should provide students with the possibility of authentic reflection, yet it runs the risk of being reduced to a repetition of notions, which will increasingly be evaluated as unobjectionable, simply because of their constant repetition.[11]

Putting the dignity of the human person back at the centre, in light of a shared ethical proposal

A more general observation should now be added to what we have already said. The season of technological innovation in which we are currently living is accompanied by a particular and unprecedented social situation in which it is increasingly difficult to find agreement on the major issues concerning social life. Even in communities characterised by a certain cultural continuity, heated debates and arguments often arise, making it difficult to produce shared reflections and political solutions aimed at seeking what is good and just. Thus aside from the complexity of legitimate points of view found within the human family, there is also a factor emerging that seems to characterise the above-mentioned social situation, namely, a loss, or at least an eclipse, of the sense of what is human and an apparent reduction in the significance of the concept of human dignity.[12] Indeed, we seem to be losing the value and profound meaning of one of the fundamental concepts of the West: that of the human person. Thus, at a time when artificial intelligence programs are examining human beings and their actions, it is precisely the *ethos* concerning the understanding of the value and dignity of the human person that is most at risk in the implementation and development of these systems. Indeed, we must remember that no innovation is neutral. Technology is born for a purpose and, in its impact on human society, always represents a form of order in social relations and an arrangement of power, thus enabling certain people to perform specific actions while preventing others from performing different ones. In a more or less explicit way, this constitutive power dimension of technology always includes the worldview of those who invented and developed it.

This likewise applies to artificial intelligence programs. In order for them to be instruments for building up the good and a better tomorrow, they must always be aimed at the good of every human being. They must have an ethical "inspiration".

Moreover, an ethical decision is one that takes into account not only an action's outcomes but also the values at stake and the duties that derive from those values. That is why I welcomed both the 2020 signing in Rome of the *Rome Call for AI Ethics*,[13] and its support for that type of ethical moderation of algorithms and artificial intelligence programs that I call "algor-ethics".[14] In a pluralistic and global context, where we see different sensitivities and multiple hierarchies in the scales of values, it might seem difficult to find a single hierarchy of values. Yet, in ethical analysis, we can also make use of other types of tools: if we struggle to define a single set of global values, we can, however, find shared principles with which to address and resolve dilemmas or conflicts regarding how to live.

This is why the *Rome Call* was born: with the term "algor-ethics", a series of principles are condensed into a global and pluralistic platform that is capable of finding support from cultures, religions, international organizations and major corporations, which are key players in this development.

The politics that is needed

We cannot, therefore, conceal the concrete risk, inherent in its fundamental design, that artificial intelligence might limit our worldview to realities expressible in numbers and enclosed in predetermined categories, thereby excluding the contribution of other forms of truth and imposing uniform

anthropological, socio-economic and cultural models. The technological paradigm embodied in artificial intelligence runs the risk, then, of becoming a far more dangerous paradigm, which I have already identified as the "technocratic paradigm".[15] We cannot allow a tool as powerful and indispensable as artificial intelligence to reinforce such a paradigm, but rather, we must make artificial intelligence a bulwark against its expansion.

This is precisely where political action is urgently needed. The Encyclical *Fratelli Tutti* reminds us that "for many people today, politics is a distasteful word, often due to the mistakes, corruption and inefficiency of some politicians. There are also attempts to discredit politics, to replace it with economics or to twist it to one ideology or another. Yet can our world function without politics? Can there be an effective process of growth towards universal fraternity and social peace without a sound political life?".[16]

Our answer to these questions is: No! Politics is necessary! I want to reiterate in this moment that "in the face of many petty forms of politics focused on immediate interests [...] 'true statecraft is manifest when, in difficult times, we uphold high principles and think of the long-term common good. Political powers do not find it easy to assume this duty in the work of nation-building' (*Laudato Si'*, 178), much less in forging a common project for the human family, now and in the future".[17]

Esteemed ladies and gentlemen!

My reflection on the effects of artificial intelligence on humanity leads us to consider the importance of "healthy politics" so that we can look to our future with hope and confidence. I have written previously that "global society is suffering from grave structural deficiencies that cannot be resolved by piecemeal solutions or quick fixes. Much needs to change, through fundamental reform and major renewal. Only a healthy politics, involving the most diverse sectors and skills, is capable of overseeing this process. An economy that is an integral part of a political, social, cultural and popular programme directed to the common good could pave the way for 'different possibilities which do not involve stifling human creativity and its ideals of progress, but rather directing that energy along new channels' (*Laudato Si'*, 191)".[18]

This is precisely the situation with artificial intelligence. It is up to everyone to make good use of it but the onus is on politics to create the conditions for such good use to be possible and fruitful.

Thank you.

---

[1] *Message for the 57th World Day of Peace*, 1 January 2024, 1.

[2] Cf. *ibid.*

[3] Cf. *ibid.*, 2.

[4] This ambivalence was already noted by Pope Saint Paul VI in his *Address to the Personnel of the "Centro Automazione Analisi Linguistica" of the Aloysianum*, 19 June 1964.

[5] Cf. A. GEHLEN, *L'uomo. La sua natura e il suo posto nel mondo*, Milan 1983, 43.

[6] Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 102-114.

[7] *Message for the 57th World Day of Peace*, 1 January 2024, 3.

[8] The insights of Marshall McLuhan and John M. Culkin are especially relevant to the consequences of the use of artificial intelligence.

[9] Cf. *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[10] Cf. *Message for the 57th World Day of Peace*, 1 January 2024, 4.

[11] Cf. *ibid.*, 3, 7.

[12] Cf. Dicastery for the Doctrine of the Faith, Declaration *Dignitas Infinita* on Human Dignity (2 April 2024).

[13] Cf. *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[14] Cf. *Address to Participants in the Congress on Child Dignity in the Digital World,* 14 November 2019; *Address to Participants in the Plenary Assembly of the Pontifical Academy for Life*, 28 February 2020.

[15] For a more extensive explanation, see the Encyclical Letter *Laudato Si'* on Care for Our Common Home (24 May 2015).

[16] Encyclical Letter, *Fratelli Tutti* on Fraternity and Social Friendship (3 October 2020), 176.

[17] *Ibid*, 178.

[18] *Ibid*, 179.

[01030-EN.01] [Original text: Italian]