

Artificial intelligence and Peace

1 January 2024

Message for the 57th World Day of Peace

1 January 2024

Pope Francis

At the beginning of the New Year, a time of grace which the Lord gives to each one of us, I would like to address God's People, the various nations, heads of state and government, the leaders of the different religions and civil society, and all the men and women of our time, in order to offer my fervent good wishes for peace.

1. The progress of science and technology as a path to peace

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). Human intelligence is an expression of the dignity with which we have been endowed by the Creator, who made us in his own image and likeness (cf. Gen 1:26), and enabled us to respond consciously and freely to his love. In a particular way, science and technology manifest this fundamentally relational quality of human intelligence; they are brilliant products of its creative potential.

In its Pastoral Constitution *Gaudium et Spes*, the Second Vatican Council restated this truth, declaring that "through its labours and its native endowments, humanity has ceaselessly sought to better its life". [1: No. 33.] When human beings, "with the aid of technology", endeavour to make "the earth a dwelling worthy of the whole human family", [2: *Ibid.*, 57.] they carry out God's plan and cooperate with his will to perfect creation and bring about peace among peoples. Progress in science and technology, insofar as it contributes to greater order in human society and greater fraternal communion and freedom, thus leads to the betterment of humanity and the transformation of the world.

We rightly rejoice and give thanks for the impressive achievements of science and technology, as a result of which countless ills that formerly plagued human life and caused great suffering have been remedied. At the same time, techno-scientific advances, by making it possible to exercise hitherto unprecedented control over reality, are placing in human hands a vast array of options, including some that may pose a risk to our survival and endanger our common home. [3: Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 104.]

The remarkable advances in new information technologies, particularly in the digital sphere, thus offer exciting opportunities and grave risks, with serious implications for the pursuit of justice and harmony among peoples. Any number of urgent questions need to be asked. What will be the consequences, in the medium and long term, of these new digital technologies? And what impact will they have on individual lives and on societies, on international stability and peace?

2. The future of artificial intelligence: between promise and risk

Progress in information technology and the development of digital technologies in recent decades have already begun to effect profound transformations in global society and its various dynamics. New digital tools are even now changing the face of communications, public administration, education, consumption, personal interactions and countless other aspects of our daily lives.

Moreover, from the digital footprints spread throughout the Internet, technologies employing a variety of algorithms can extract data that enable them to control mental and relational habits for commercial or political purposes, often without our knowledge, thus limiting our conscious exercise of freedom of choice. In a space like the Web, marked by information overload, they can structure the flow of data according to criteria of selection that are not always perceived by the user.

We need to remember that scientific research and technological innovations are not disembodied and "neutral", [4: Cf. *ibid.*, 114.] but subject to cultural influences. As fully human activities, the directions they take reflect choices conditioned by personal, social and cultural values in any given age. The same must be said of the results they produce: precisely as the fruit of specifically human ways of approaching the world around us, the latter always have an ethical dimension, closely linked to decisions made by those who design their experimentation and direct their production towards particular objectives.

This is also the case with forms of artificial intelligence. To date, there is no single definition of artificial intelligence in the world of science and technology. The term itself, which by now has entered into everyday parlance, embraces a variety of sciences, theories and techniques aimed at making machines reproduce or

imitate in their functioning the cognitive abilities of human beings. To speak in the plural of “forms of intelligence” can help to emphasize above all the unbridgeable gap between such systems, however amazing and powerful, and the human person: in the end, they are merely “fragmentary”, in the sense that they can only imitate or reproduce certain functions of human intelligence. The use of the plural likewise brings out the fact that these devices greatly differ among themselves and that they should always be regarded as “socio-technical systems”. For the impact of any artificial intelligence device – regardless of its underlying technology – depends not only on its technical design, but also on the aims and interests of its owners and developers, and on the situations in which it will be employed.

Artificial intelligence, then, ought to be understood as a galaxy of different realities. We cannot presume a priori that its development will make a beneficial contribution to the future of humanity and to peace among peoples. That positive outcome will only be achieved if we show ourselves capable of acting responsibly and respect such fundamental human values as “inclusion, transparency, security, equity, privacy and reliability”. [5: Address to Participants in the “Minerva Dialogues” (27 March 2023).]

Nor is it sufficient simply to presume a commitment on the part of those who design algorithms and digital technologies to act ethically and responsibly. There is a need to strengthen or, if necessary, to establish bodies charged with examining the ethical issues arising in this field and protecting the rights of those who employ forms of artificial intelligence or are affected by them. [6: Cf. *ibid.*]

The immense expansion of technology thus needs to be accompanied by an appropriate formation in responsibility for its future development. Freedom and peaceful coexistence are threatened whenever human beings yield to the temptation to selfishness, self-interest, the desire for profit and the thirst for power. We thus have a duty to broaden our gaze and to direct techno-scientific research towards the pursuit of peace and the common good, in the service of the integral development of individuals and communities. [7: Cf. Message to the Executive Chairman of the “World Economic Forum” meeting in Davos (12 January 2018).]

The inherent dignity of each human being and the fraternity that binds us together as members of the one human family must undergird the development of new technologies and serve as indisputable criteria for evaluating them before they are employed, so that digital progress can occur with due respect for justice and contribute to the cause of peace. Technological developments that do not lead to an improvement in the quality of life of all humanity, but on the contrary aggravate inequalities and conflicts, can never count as true progress. [8: Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 194; Address to Participants in the Seminar “The Common Good in the Digital Age” (27 September 2019).]

Artificial intelligence will become increasingly important. The challenges it poses are technical, but also anthropological, educational, social and political. It promises, for instance, liberation from drudgery, more efficient manufacturing, easier transport and more ready markets, as well as a revolution in processes of accumulating, organizing and confirming data. We need to be aware of the rapid transformations now taking place and to manage them in ways that safeguard fundamental human rights and respect the institutions and laws that promote integral human development. Artificial intelligence ought to serve our best human potential and our highest aspirations, not compete with them.

3. The technology of the future: machines that “learn” by themselves

In its multiple forms, artificial intelligence based on machine learning techniques, while still in its pioneering phases, is already introducing considerable changes to the fabric of societies and exerting a profound influence on cultures, societal behaviours and peacebuilding.

Developments such as machine learning or deep learning, raise questions that transcend the realms of technology and engineering, and have to do with the deeper understanding of the meaning of human life, the construction of knowledge, and the capacity of the mind to attain truth.

The ability of certain devices to produce syntactically and semantically coherent texts, for example, is no guarantee of their reliability. They are said to “hallucinate”, that is, to create statements that at first glance appear plausible but are unfounded or betray biases. This poses a serious problem when artificial intelligence is deployed in campaigns of disinformation that spread false news and lead to a growing distrust of the communications media. Privacy, data ownership and intellectual property are other areas where these technologies engender grave risks. To which we can add other negative consequences of the misuse of these technologies, such as discrimination, interference in elections, the rise of a surveillance society, digital exclusion and the exacerbation of an individualism increasingly disconnected from society. All these factors risk fueling conflicts and hindering peace.

4. The sense of limit in the technocratic paradigm

Our world is too vast, varied and complex ever to be fully known and categorized. The human mind can never exhaust its richness, even with the aid of the most advanced algorithms. Such algorithms do not offer guaranteed predictions of the future, but only statistical approximations. Not everything can be predicted, not everything can be calculated; in the end, “realities are greater than ideas”. [9: Apostolic Exhortation *Evangelii Gaudium* (24 November 2013), 233.] No matter how prodigious our calculating power may be, there will always be an inaccessible residue that evades any attempt at quantification.

In addition, the vast amount of data analyzed by artificial intelligences is in itself no guarantee of impartiality. When algorithms extrapolate information, they always run the risk of distortion, replicating the injustices and prejudices of the environments where they originate. The faster and more complex they become, the more difficult it proves to understand why they produced a particular result.

“Intelligent” machines may perform the tasks assigned to them with ever greater efficiency, but the purpose and the meaning of their operations will continue to be determined or enabled by human beings possessed of their own universe of values. There is a risk that the criteria behind certain decisions will become less clear, responsibility for those decisions concealed, and producers enabled to evade their obligation to act for the benefit of the community. In some sense, this is favoured by the technocratic system, which allies the economy with technology and privileges the criterion of efficiency, tending to ignore anything unrelated to its immediate interests. [10: Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 54.]

This should lead us to reflect on something frequently overlooked in our current technocratic and efficiency-oriented mentality, as it is decisive for personal and social development: the “sense of limit”. Human beings are, by definition, mortal; by proposing to overcome every limit through technology, in an obsessive desire to control everything, we risk losing control over ourselves; in the quest for an absolute freedom, we risk falling into the spiral of a “technological dictatorship”. Recognizing and accepting our limits as creatures is an indispensable condition for reaching, or better, welcoming fulfilment as a gift. In the ideological context of a technocratic paradigm inspired by a Promethean presumption of self-sufficiency, inequalities could grow out of proportion, knowledge and wealth accumulate in the hands of a few, and grave risks ensue for democratic societies and peaceful coexistence. [11: Cf. Meeting with Participants in the Plenary Assembly of the Pontifical Academy for Life (28 February 2020).]

5. Burning issues for ethics

In the future, the reliability of an applicant for a mortgage, the suitability of an individual for a job, the possibility of recidivism on the part of a convicted person, or the right to receive political asylum or social assistance could be determined by artificial intelligence systems. The lack of different levels of mediation that these systems introduce is particularly exposed to forms of bias and discrimination: systemic errors can easily multiply, producing not only injustices in individual cases but also, due to the domino effect, real forms of social inequality.

At times too, forms of artificial intelligence seem capable of influencing individuals' decisions by operating through pre-determined options associated with stimuli and dissuasions, or by operating through a system of regulating people's choices based on information design. These forms of manipulation or social control require careful attention and oversight, and imply a clear legal responsibility on the part of their producers, their deployers, and government authorities.

Reliance on automatic processes that categorize individuals, for instance, by the pervasive use of surveillance or the adoption of social credit systems, could likewise have profound repercussions on the social fabric by establishing a ranking among citizens. These artificial processes of categorization could lead also to power conflicts, since they concern not only virtual users but real people. Fundamental respect for human dignity demands that we refuse to allow the uniqueness of the person to be identified with a set of data. Algorithms must not be allowed to determine how we understand human rights, to set aside the essential human values of compassion, mercy and forgiveness, or to eliminate the possibility of an individual changing and leaving his or her past behind.

Nor can we fail to consider, in this context, the impact of new technologies on the workplace. Jobs that were once the sole domain of human labour are rapidly being taken over by industrial applications of artificial intelligence. Here too, there is the substantial risk of disproportionate benefit for the few at the price of the impoverishment of many. Respect for the dignity of labourers and the importance of employment for the economic well-being of individuals, families, and societies, for job security and just wages, ought to be a high

priority for the international community as these forms of technology penetrate more deeply into our workplaces.

6. Shall we turn swords into ploughshares?

In these days, as we look at the world around us, there can be no escaping serious ethical questions related to the armaments sector. The ability to conduct military operations through remote control systems has led to a lessened perception of the devastation caused by those weapon systems and the burden of responsibility for their use, resulting in an even more cold and detached approach to the immense tragedy of war. Research on emerging technologies in the area of so-called Lethal Autonomous Weapon Systems, including the weaponization of artificial intelligence, is a cause for grave ethical concern. Autonomous weapon systems can never be morally responsible subjects. The unique human capacity for moral judgment and ethical decision-making is more than a complex collection of algorithms, and that capacity cannot be reduced to programming a machine, which as “intelligent” as it may be, remains a machine. For this reason, it is imperative to ensure adequate, meaningful and consistent human oversight of weapon systems.

Nor can we ignore the possibility of sophisticated weapons ending up in the wrong hands, facilitating, for instance, terrorist attacks or interventions aimed at destabilizing the institutions of legitimate systems of government. In a word, the world has no need of new technologies that contribute to the unjust development of commerce and the weapons trade and consequently end up promoting the folly of war. By so doing, not only intelligence but the human heart itself would risk becoming ever more “artificial”. The most advanced technological applications should not be employed to facilitate the violent resolution of conflicts, but rather to pave the way for peace.

On a more positive note, if artificial intelligence were used to promote integral human development, it could introduce important innovations in agriculture, education and culture, an improved level of life for entire nations and peoples, and the growth of human fraternity and social friendship. In the end, the way we use it to include the least of our brothers and sisters, the vulnerable and those most in need, will be the true measure of our humanity.

An authentically humane outlook and the desire for a better future for our world surely indicates the need for a cross-disciplinary dialogue aimed at an ethical development of algorithms – an algor-ethics – in which values will shape the directions taken by new technologies. [12: Cf. *ibid.*] Ethical considerations should also be taken into account from the very beginning of research, and continue through the phases of experimentation, design, production, distribution and marketing. This is the approach of ethics by design, and it is one in which educational institutions and decision-makers have an essential role to play.

7. Challenges for education

The development of a technology that respects and serves human dignity has clear ramifications for our educational institutions and the world of culture. By multiplying the possibilities of communication, digital technologies have allowed us to encounter one another in new ways. Yet there remains a need for sustained reflection on the kinds of relationships to which they are steering us. Our young people are growing up in cultural environments pervaded by technology, and this cannot but challenge our methods of teaching, education and training.

Education in the use of forms of artificial intelligence should aim above all at promoting critical thinking. Users of all ages, but especially the young, need to develop a discerning approach to the use of data and content collected on the web or produced by artificial intelligence systems. Schools, universities and scientific societies are challenged to help students and professionals to grasp the social and ethical aspects of the development and uses of technology.

Training in the use of new means of communication should also take account not only of disinformation, “fake news”, but also the disturbing recrudescence of “certain ancestral fears... that have been able to hide and spread behind new technologies”. [13: Encyclical Letter *Fratelli Tutti* (3 October 2020), 27.] Sadly, we once more find ourselves having to combat “the temptation to build a culture of walls, to raise walls... in order to prevent an encounter with other cultures and other peoples”, [14: *Ibid.*] and the development of a peaceful and fraternal coexistence.

8. Challenges for the development of international law

The global scale of artificial intelligence makes it clear that, alongside the responsibility of sovereign states to

regulate its use internally, international organizations can play a decisive role in reaching multilateral agreements and coordinating their application and enforcement. [15: Cf. *ibid*, 170-175.] In this regard, I urge the global community of nations to work together in order to adopt a binding international treaty that regulates the development and use of artificial intelligence in its many forms. The goal of regulation, naturally, should not only be the prevention of harmful practices but also the encouragement of best practices, by stimulating new and creative approaches and encouraging individual or group initiatives. [16: Cf. Encyclical Letter *Laudato Si'* (24 May 2015), 177.]

In the quest for normative models that can provide ethical guidance to developers of digital technologies, it is indispensable to identify the human values that should undergird the efforts of societies to formulate, adopt and enforce much-needed regulatory frameworks. The work of drafting ethical guidelines for producing forms of artificial intelligence can hardly prescind from the consideration of deeper issues regarding the meaning of human existence, the protection of fundamental human rights and the pursuit of justice and peace. This process of ethical and juridical discernment can prove a precious opportunity for shared reflection on the role that technology should play in our individual and communal lives, and how its use can contribute to the creation of a more equitable and humane world. For this reason, in debates about the regulation of artificial intelligence, the voices of all stakeholders should be taken into account, including the poor, the powerless and others who often go unheard in global decision-making processes.

I hope that the foregoing reflection will encourage efforts to ensure that progress in developing forms of artificial intelligence will ultimately serve the cause of human fraternity and peace. It is not the responsibility of a few but of the entire human family. For peace is the fruit of relationships that recognize and welcome others in their inalienable dignity, and of cooperation and commitment in seeking the integral development of all individuals and peoples.

It is my prayer at the start of the New Year that the rapid development of forms of artificial intelligence will not increase cases of inequality and injustice all too present in today's world, but will help put an end to wars and conflicts, and alleviate many forms of suffering that afflict our human family. May Christian believers, followers of various religions and men and women of good will work together in harmony to embrace the opportunities and confront the challenges posed by the digital revolution and thus hand on to future generations a world of greater solidarity, justice and peace.

From the Vatican, 8 December 2023