



SALA STAMPA DELLA SANTA SEDE  
**BOLLETTINO**

HOLY SEE PRESS OFFICE BUREAU DE PRESSE DU SAINT-SIÈGE PRESSEAMT DES HEILIGEN STUHLS  
OFICINA DE PRENSA DE LA SANTA SEDE SALA DE IMPRENSA DA SANTA SÉ  
BIURO PRASOWE STOLICY APOSTOLSKIEJ دار الصحافة التابعة للكرسي الرسولي

N. 200225a

Tuesday 25.02.2020

**Press Conference for the presentation of the Workshop and Assembly of the Pontifical Council for Life on the theme “The ‘good’ algorithm? Artificial Intelligence: Ethics, Law, Health” (New Synod Hall in the Vatican, 26 to 27 February 2020)**

**Intervention by Archbishop Vincenzo Paglia**

**Intervention by Reverend Fr. Paolo Benanti, T.O.R.**

**Intervention by Professor Maria Chiara Carrozza**

At midday today, in the John Paul II Hall of the Holy See Press Office, a press conference was held to present the Workshop and Assembly of the Pontifical Academy for Life on the theme: “The ‘good’ algorithm? Artificial Intelligence: Ethics, Law, Health”, taking place in the New Synod Hall in the Vatican from 26 to 27 February 2020.

The speakers were Archbishop Vincenzo Paglia, president of the Pontifical Academy for Life; the Reverend Fr. Paolo Benanti, T.O.R., academician of the Pontifical Academy for Life; and Professor Maria Chiara Carrozza, professor of industrial bio-engineering, Sant’Anna School of Advanced Studies, Pisa.

The following are the interventions:

**Intervention by Archbishop Vincenzo Paglia**

We are in a time of epochal change, as Pope Francis likes to say. It is an unprecedented passage that is profoundly changing humanity and its future. For the first time in history man has the power to destroy himself: first by nuclear explosion, then the ecological one, and now with the technological one, an “explosion of intelligence”. Pope Francis, with the Letter *Humana Communitas*, invited the Pontifical Academy for Life to broaden its horizons, to revisit the very meaning of the term “human life”: it is not an abstract concept; life is the

reality of every single person and of the entire human family. Pope Francis asks the Academy to “develop reflection” on the “new technologies today defined as emerging and converging”, such as information and communication technologies, biotechnology, nanotechnology, and robotics. With the results obtained from physics, genetics and neuroscience, as well as the computational capacity of increasingly powerful machines, it is now possible to intervene profoundly in the human being. In fact, digital innovation touches all aspects of life, both personal and social; it affects our way of understanding not only the world, but also ourselves. Decisions, even the most important ones such as those in the medical, economic or social field, are today the result of human will and of a series of algorithmic contributions. Human life finds itself at the point of convergence between the properly human contribution and automatic calculation, so that it is increasingly complex to understand its object, to foresee its effects, to define its responsibilities.

The Academy has entered this field, without abandoning traditional areas such as birth (abortion, prenatal diagnosis...) end of life (euthanasia, assisted suicide, palliative care), and stem cell development. In 2017 the issue of the protection and promotion of human life in the technological era and therefore in the horizon of global bioethics was addressed. Last year attention turned to the ethical issues posed by robotics (so-called “roboethics”) and this year, in continuity, we will address the issue of ethics and artificial intelligence. Actually, we have also been solicited or, if you like, invited to deal with these issues by those directly concerned.

This is the horizon in which this General Assembly takes place, and in particular the event of 28 February, at the end of which a Call will be signed, which we will then present to Pope Francis. A strong moral ambition is needed to humanise technology rather than to technologise humanity.

The “Rome Call for AI Ethics” is not an official text of the Academy but rather a document of shared commitments, proposed by us, in which, in brief and synthetic form, some guidelines for an ethics of Artificial Intelligence are offered and some commitments are formulated, fundamentally linked to three chapters: ethics, law, and education. With this gesture, the Academy does not initiate exclusive industrial partnerships, nor does it sponsor anything, but it shares, without naivety, a part of the journey with those who have a serious desire to better understand how to promote the good of humanity and to take some steps in this direction, confirming their own practices and willing even to pay the costs that may result. The intention of the Call is to create a movement that expands and involves other actors: public institutions, NGOs, industries and groups to produce a direction in the development and use of AI-derived technologies. From this point of view we can say that the first signature of this call is not a point of arrival, but a beginning for a commitment that seems even more urgent and important than what has been done so far. The document will be available on Friday, from the moment of its signature, on the dedicated website [www.romecall.org](http://www.romecall.org).

The Academy feels called to advance the specific impact that these technologies have on the medical and health care world and on the care and protection of life. Human activity in these areas appears to be increasingly broken down into multiple elements that are not easily traceable to the control or will of individual subjects. This new way in which personal action takes place in a structured context particularly challenges the medical and health care professions which have as their object values as fundamental as those related to human corporeity and life. Technological innovation challenges us as an Academy and as a Church; the PAV thus begins to take a position and take part in a historical and social context engaged in profound and continuous transformation.

### **Intervention by Reverend Fr. Paolo Benanti, T.O.R.**

Today we are facing a fourth industrial revolution linked to the pervasive spread of a new form of technology: artificial intelligence or AI. Like electricity and electronics, AI is not used to do one specific thing; rather, it is destined to change the way we do everything.

How is this possible? In recent years, increasingly powerful computers have generated a huge amount of computing power, available at ever lower prices. At the same time we have started to accumulate a quantity of data that continues to grow at a dizzying pace: ninety per cent of the data ever generated in the entire history of

mankind has been created over the last two years. These two factors have made some families of algorithms that give rise to the complex world of AI - a world that scientists have been thinking about, at least theoretically, since the 1960s.

What will change all this? The first and second industrial revolution (with coal and steam, and with electricity and oil, respectively) have provided us with forms of energy that offer an alternative to muscles; the third produced automatic machines, disrupting the concept of assembly line and worker; what is about to happen risks automating not force, not labour, but our cognition.

AI systems are capable of adapting and adjusting to the changing conditions in which they operate, simulating what a person would do. In other words, today, the machine can often substitute man in making decisions and choices. If the other industrial revolutions were about blue collar workers, what is happening now is mainly about white collars. AI will not lead to the apocalypse, but may lead to the end of the middle class.

Today machine learning algorithms and other forms of AI are able to make medical diagnoses with a percentage of accuracy that in some cases exceeds that of an average doctor (at least in some disciplines or with some pathologies); they can predict who will be able to repay a loan much more accurately than a bank manager; according to some developers, they can understand better than us if there is an affective affinity with the person in front of us. AIs are becoming increasingly predictive.

However, when faced with such accuracy, they do not have as much explanatory capacity: the most efficient algorithms are the ones we understand the least, compared to which we are less able to say why the machine indicates this result.

At this level a great question arises. When the machine surrogates man in making decisions, what kind of certainties should we have in order to let the machine choose who should be treated and how? On the basis of what should we allow a machine to designate which of us is worthy of trust and which is not? And what about love, that unique quest that has moved generations of women and men before us?

If we can turn human problems into statistics, graphs and equations with a computer, we create the illusion that these problems can be solved with computers. That is not so.

The use of computers and information technology in technological development in fact highlights a linguistic challenge that occurs on the borderline between man and machine. In the process of reciprocal interrogation between man and machine, projections and exchanges arise, hitherto unthought of: the machine becomes human while man becomes stained.

What does it mean, then, to humanise technology and not to "machinise" man?

When he makes choices, the human being knows a deep and radical qualification of his actions: good and evil. Man discovers with his freedom a sense of responsibility that our western tradition has called ethics. The characteristically human ethic makes us unique and is based on values. The machine also chooses on the basis of values - but they are the numerical values of data.

If we want the machine to support man and the common good, without ever taking the place of the human being, then the algorithms must include ethical values and not just numerical values.

If we want the machine to support man and the common good, without ever taking the place of the human being, then the algorithms must include ethical values and not just numerical ones.

In essence, we need to be able to indicate ethical values through the numerical values that feed the algorithm.

Ethics needs to contaminate computing. We need an *algorithm-ethics*, that is, we need a way to make good and bad evaluations computable. Only in this way we can create machines that can become tools for the humanisation of the world. We need to codify ethical principles and norms in a language that can be understood and used by machines. For AI to be a revolution leading to real development, it is time to think of an *algorithm-ethics*.

On 28 February the Rome Call will mark an important step in this direction: two of the major manufacturers of AI, IBM and Microsoft, along with the Pontifical Academy for Life, will sign this call for some ethical principles to be present in the AI products they develop, sell and implement. The Call, an open structure, is intended to be the beginning of a movement that brings together men of good will to cooperate so that ethical choices, legal paradigms and appropriate educational actions make civil society capable of facing this new era.

### **Intervention by Professor Maria Chiara Carrozza**

Artificial Intelligence (AI) is one of the enabling technologies that characterised the fourth industrial revolution, but its influence will go far beyond the world of the production of goods and services, as it will have a disruptive social and cultural impact through the pervasiveness with which it will enter our future, changing our relationship with society.

Therefore, an understanding of the phenomenon linked to the diffusion of AI becomes fundamental so that we can govern the change connected to it and guide it towards the common good, in a scenario of geopolitical and institutional reference deeply intertwined with the development and “possession” of the different technologies.

One of the areas of reference with the greatest impact for the development of AI is undoubtedly medicine, starting from the so-called “digital transformation” centred on the use of the available data through opportune infrastructures. The success of the use of AI in the many fields in which it can be developed certainly depends on the ability to effectively select the data that will feed the algorithm underlying the functioning of AI mechanisms. Data can be generated from different sources, in particular: from humans, machines, organisations or a combination of these actors. The possibilities of obtaining data are increasing exponentially thanks to the use of various technologies in the world of “Internet Of Things”.

In fact, there are numerous possibilities for using AI algorithms in medicine: from translational experimentation and research to personalised medicine, from diagnostics to the physician-patient relationship, from tele-assistance and tele-rehabilitation to robotic surgery, from virtual coaching to predictive medicine, from support to functional enhancement of the patient through robotics and sensors that can be worn or implanted in the human body. Within the individual fields of application, AI can assume different roles that may vary according to the type of diagnosis to be made, the hospital or territorial nature of the treatment, and the acute or chronic character of the pathology. Moreover, depending on the field of application, AI mechanisms can give more or less reliable results, and must therefore undergo different forms of validation to guarantee the patient’s rights.

The strength of collecting information in databases is that, because they contain a large amount of information, it is possible to search for relationships between the various datasets to find correlations. This type of analysis has an important predictive value - its use in medicine can make it possible to predict future conditions, situations and events, with statistical forecasts at the level of the population or even the individual (“personalised medicine”).

Data represent the flow of fuel through which AI can enable increasingly effective analysis and decisions also in the clinical setting, but it is the task of institutions to ensure that the processes are in any case based on methodologies based on scientific evidence and respect for ethical principles.

A series of key ethical principles can therefore be identified as a sort of framework for the application of AI in

medicine:

- Medicine is a human prerogative; AI is a tool that can support the professional with greater or lesser intensity depending on the field involved. A reflection on the doctor-patient relationship is necessary because AI cannot represent an element in a possible reduction of the responsibility of the doctor.
  - A significant human intervention is however necessary also because the applications of AI to medicine, despite what may sometimes be perceived, are not neutral. AI planning presupposes discretionary choices that can have very important consequences also in terms of health and that should be made on the basis of an interdisciplinary comparison between different competences.
  - It is important that AI/IV is designed in a way that responds to well-defined ethical principles that define the contours of the relationship between patient and technology and between patient and doctor, inspired by a path already traced by the law on "informed consent".
  - The need for specific training in AI is also fundamental, and necessarily interdisciplinary and continuous, accompanying the professional throughout his or her career and able to constantly follow global changes.
  - The application of AI in the medical field has also had repercussions in terms of market choices, promoting different realities of companies that have brought innovation in the healthcare field. For example, in the European context, so-called "digital therapies" are emerging, showing how it is possible to make good use of AI by administering integrated therapies that can change life and improve conditions even in the presence of serious chronic diseases, guiding patients towards optimised behaviour according to their physiological and cognitive conditions. Access to the therapies will be possible through appropriate platforms that will be presented as applications on smartphones.
  - In this scenario, AI becomes a development tool that must guarantee an ethical use of data in an economy that protects the individual in the community, in relation to the turnover that is created around the data and the platforms that exploit the algorithms, and therefore the ethics of AI must be based on respect for the fundamental rights of the patient. It is an evolving picture with enormous potential but also characterised by great risks of iniquity and exploitation. It is therefore necessary to respond to the challenge of establishing an ethics of AI inspired by the universal principle of equality. The availability of technologies for the few does not create value; on the contrary it generates disvalue and inequality. The commitment is therefore to make these technologies available to everyone, regardless of their geographical origin or economic conditions. For the public administration and in general for all stakeholders, the principles of equality and equity should be the compass for the development of AI in medicine.
  - I would like to conclude my speech by echoing the words recently expressed by the President of the European Commission, Ursula von der Leyen, at the presentation of the EU digital strategy: "Today we are presenting our ambition to shape Europe's digital future. The strategy covers everything from cyber security to critical infrastructure, from digital education to skills, from democracy to media. I want digital Europe to reflect the best of Europe: open, fair, diverse, democratic and self-confident". And I too, like the President (I quote from her Twitter profile) "I am a technology optimist. I believe in technology as a force for good", as long as it is regulated and equipped with an essential apparatus of ethical values.
-