Issues for Future: Coping Ethically with AI in Europe

Prof. Markus Krienke (Lugano–Roma–Milano–Messina)

One of the main topics of European Union's policies is to sustain Artificial Intelligence's development under deontological and juridical rules. The new President of the European Commission, after presenting the Green New Deal on December 11th, 2019, specified only two months later that the goal of climate neutrality by 2050 is not achievable without AI (Ursula von der Leyen, *Shaping Europe's Digital Future*, 19 February 2020). She strengthened positively the technical and scientific presuppositions for the evolution of that sector in the EU which produces, for example, 25% of all industrial robots. With the perspective to increase cloud infrastructures and job employment, but also on the basis of the "Ethics guidelines for a trustworthy IA" (8 April 2019), the European Commission presents a twofold, technical and ethical project under the principles of "Respect for human autonomy", "Prevention of harm", "Fairness" and "Explicability". At the same time, precisely on February 28th, 2020, Pontifical Academy for Life, Microsoft, IBM, FAO and the Italian Government firmed in the Vatican the "Rome Call for AI Ethics" in order to orientate the rapid development at "the good of humanity and of the environment" by the criterions of "Transparency", "Inclusion", "Responsibility", "Impartiality", "Reliability" and "Security and privacy". The interest of Pope Francis and the Vatican institutions for this topic is more than noteworthy: already in January 2018 he emphasized at the World Economic Forum in Davos that «artificial intelligence, robotics and other technological innovations must be so employed that they contribute to the service of humanity and to the protection of our common home». And one year later the Pope addressed to the Pontifical Academy of Life, which after its workshops in 2019 and 2020 released the "Rome Call for AI Ethics", his preoccupation that «the risk of man being "technologized", rather than technology humanized, is already real». The principles of both European and Vatican documents try to find the right balance between promises and limits, or between development and rules; and for better individuating this balance there are at least three anthropological questions to resolve, i.e. (1) to understand the objective future impact of this technology on our individual lives and social institutions, (2) to get aware of the qualitative limits of AI, and (3) to define the exact difference between "intelligent" and "morally acting" robots or cyborgs, on the one hand, and human beings on the other. This twofold reflection follows the fundamental value-orientation of our European constitutional idea that every social reality has to be measured in its relationship to the absoluteness of human dignity.

(1) The exponential – or probably more than exponential – growth of the AI's technological development is concretely evident in two aspects: on the one hand, the time of doubling the data produced by humanity is only about one year, and in 10 years, when the number of devices connected all over the world will be 150 billion, it will be reduced to 12 hours. On the other hand, an 2015's iPhone "6s" is 120 million times faster in data processing than the

computer that took the Apollo 11 to the Moon in 1969, while that differential grew another 240 million times between the iPhone "6s" and the "X" version only two years later. An average Ghanaian teenager today has more information on his smartphone than the US president disposed at 20 years ago. *Big Data* and information technology do not only produce systems which imitate human intelligence, intuitiveness and creativity, but also his acting, morality and emotions. Up to 50% of our nowadays professions could be replaced by robots, which maybe will drive our cars, cure our ancient people, doing surgeries, revolution our educative systems and war strategies. Someone therefore speaks of a "fourth industrial revolution" or a "second Great Transformation" which will redefine society and lives as well as the industrial revolution did two centuries ago.

(2) At the same time, we get aware of the specific limits of this development which brings intelligent systems to always more perfection in singular issues, without producing the "holy grail of AI research" which would be the "General Artificial Intelligence" (GAI), i.e. the complete substitution of human intelligence by AI. The qualitative difference is best exemplified by the "Winograd Schema Challenge" which shows the semantic limits of AI which elaborates *data* but cannot intentionally refer to structures of *sense* as the simple referring of pronouns: given the phrase "I can't cut that tree down with that axe; it is too small", IA isn't able to answer to the question "What is too small, the tree or the axe?" Even if AI is in its results *similar* to human intelligence (i.e. it can produce poems, art or write newspaper articles), it is never other than elaborating *data*, so it could be defined "intelligence without reason". And if it acts as "artificial moral agent" (AMA), it only simulates human actions, without the ability to give reasons for its acting or to assume moral responsibility. Let's admit that a doctor (robot) accidentally kills a patient, because he lacked the information that instead of medicine gave him poison, and expresses regret for the death of the patient. Since it is a robot, there is no way to morally judge this situation or to attribute responsibility to it. The robot simply "lacked" information, and "regret" is only the reaction it has learned to show in such cases. Of course, it would be better for the robot too if the patient had not died because he would receive positive feedback, but it is not possible for him to make a moral ("conscientious") judgement about his actions. In analogy to "intelligence without reason", we could call this type of acting "agency without morality".

(3) This analysis on the possible impact of AI in our future lives and societies, and the awareness of its necessarily limits, brings us to the reflection about the exact anthropological difference between AI and human beings. While Ray Kurzweil doesn't see any and expects the moment of "singularity" – when AI substitutes and surpasses completely human intelligence – others like John Searle define the qualitative difference in the intentionality of human intelligence: In order to answer to the question "Could a Machine Think" he imagines a person in a room which a very clear handbook of rules that instructs him how to answer to question a Chinese asks him from outside of the room, but without understanding the significance neither of the questions nor of the answers. This "Chinese room" example answers to the initial question clearly "No" and explains that human understanding means – in difference to AI – "intentionality", i.e. referring to a *real sense*. Or let's imagine a neuronal scientist, Mary, who knows all the information about what it means to see colors, but she lives in a house without any colors. Until she leaves the house, she'll never know what it's like to see colors. From this example we can learn that "intentionality" of the human intelligence is given with the bodily existence of intelligence: therefore our very human experience is never reducible to the set of data available to an AI.

Konrad-Adenauer-Stiftung e.V.

AUSLANDSBÜRO ITALIEN

May 2020

In the case of moral decisions, AI explores *data* and finds the "optimum" or "ideal" solution according to a pre-defined "program". In this sense, AI aspires to "perfect decisions", but the moral character of human choices and decisions does not lie in "perfection" but in the ability to weigh the different reasons for the one or the other option, trying to realize "good" and avoid "bad" achievements. In other words, while the imperative of AI is optimization, that of human beings is "responsibility". For example, it would be useless to ask a robot to justify a certain choice or Google Translate because it used one word instead of another.

In this sense we can summarize the distinction between human and artificial intelligence by the definition "if a machine can do it, then it's no longer (human) intelligence", adding the analogous affirmation that "if the computer can resolve it, then it's no longer a moral problem". On the basis of this anthropological and ethical distinction, the "Ethics guidelines for a trustworthy IA" (2019) and the "White Paper on Artificial Intelligence" (2020) can be read as a concrete contribution for a new "Digital Humanism" in an European politics oriented to our future society. Only at that condition, "Artificial Intelligence is a huge opportunity in Europe, for Europe", as von der Leyen stated on February 19th. Meaning technological evolution, but also implying ethical and anthropological reflection, she then emphasized: "We do have a lot, but we have to unleash this potential that is out there".

Author

Markus Krienke (1978): Full Professor of Modern Philosophy and Social Ethics at the Faculty of Theology of Lugano and Director of the Rosmini Chair. He also teaches Philosophical Anthropology at the Pontifical Lateran University in Rome and Social Doctrine of the Church at the Theological Faculty in Milan. He is visiting professor in Political Science at the University of Messina and is part of the master's degree in Ethics of Peace, Rights and Protection of the Person in productive educational and economic contexts at the University of Bari. As a member of the European Academy of Sciences and Arts and the Italian Scientific Committee of the Konrad Adenauer Foundation, he organized in Lugano in 2016 a public conference on interreligious dialogue with participants from 14 countries, and in 2018 the exhibition on the European project of Adenauer, De Gasperi and Schuman and the relevance of relations between Switzerland and Europe. He collaborates with Corriere del Ticino and Swiss Radio and Television. In his studies he deals with Antonio Rosmini's thought, metaphysics in modern times, themes of political philosophy and human rights, the ethical issues of economics and lately the challenges arising from Artificial Intelligence, trans- and post-humanism.