



The RenAIssance logo in front of St. Peter's Square, Rome.

Pontifical Academy for Life



THE “GOOD” ALGORITHM?

ARTIFICIAL INTELLIGENCE
ETHICS, LAW, HEALTH

PROCEEDINGS OF THE XXVI GENERAL
ASSEMBLY OF MEMBERS

Vatican City, February 26-28, 2020

Edited by
VINCENZO PAGLIA and RENZO PEGORARO

Rome 2021

All rights reserved

© Pontifical Academy for Life

ISBN 979-12-80365-00-2

Contents

Introduction	11
Address of His Holiness Pope Francis to Participants in the XXVI General Assembly of the Pontifical Academy for Life	19
Opening address	
Vincenzo Paglia	23
First session. Artificial Intelligence and Ethics	
Adriano Pessina	
<i>Is a Good Algorithm also an Ethically Good Algorithm?</i>	27
Paolo Benanti	
<i>Algor-ethics: Artificial Intelligence calls for an Ethical Reflection</i>	31
Alexander Filipović	
<i>Ethical and Social Consequences of Artificial Intelligence: Insights from a Christian Social Ethics Perspective.....</i>	47
Robin R. Wang	
<i>Flowing of Life and Static of Machine: A Daoist Perspective on AI.....</i>	71
Sandra K. Alexander	
<i>Educatio Vitae: Person-centered Ethics Education in the Age of AI</i>	85
Second session. Artificial Intelligence and Human Health	
Walter Ricciardi	
<i>Artificial Intelligence and health.....</i>	93
Yuzo Takahashi	
<i>The Clinical Consequence of AI</i>	103
Alexandru G. Floares	
<i>Artificial Intelligence in Oncology</i>	117
Felix Hector Rigoli	
<i>Artificial Intelligence in the road of Health for All. Perils and Hope</i>	123

Contents

Shinya Yamanaka

AI in Medicine. Recent Progress in iPS Cell Research and Application... 141

Third session. Artificial Intelligence and Law

James A. Shaw and Leah T. Kelley

Policy and Governance of AI for Health: A Global Ethics Perspective ... 157

Federico de Montalvo Jääskeläinen

The Secondary Use of Health Data in the Context of Big Data and the New European Legal Framework: Have We Changed the Helsinki Paradigm? ... 177

Cédric Wachholz

UNESCO's Perspective ... 183

Amir Banitafemi

AI Common Projects ... 189

Francesco Profumo

A new RenAissance for the future of Education ... 195

Fourth session. The Rome Call: “RenAissance. A human-centric artificial intelligence”

Introduction ... 211

Vincenzo Paglia ... 213

Brad Smith ... 217

John E. Kelly III ... 223

David Sassoli ... 229

Dongyu Qu ... 235

Rome Call for AI Ethics ... 239

Abstracts of posters

Validation of Artificial Intelligence in Medical Diagnosis, utilizing models traditionally used in the Financial Industry (by Attard Trevisan A.)... 249

General Views of Bioethicists in Bulgaria about Artificial Intelligence in Medicine (by Aleksandrova-Yankulovska S.) ... 249

“If they asked you to jump off a cliff?”: AI and clinical decision-making (by Smith H.) ... 250

<i>Artificial Intelligence, Offender Rehabilitation & Restorative Justice</i> (by Alves Pereira A.C.)	251
<i>Ontological Plasticity and the Challenge to Anthropocentrism: Invoking Ethical Parity in Material Relations</i> (by Larrivee D.).....	252
<i>Human-Centric Algorithms in Healthcare 4.0: The Agenda of Campus Bio-Medico for a Good Polyclinic</i> (by Corti L. et al.)	253
<i>Fit for Purpose? The GDPR and the European Governance of Health-Related AI Technologies</i> (by Marelli L.).....	254
<i>ARTificial Intelligence</i> (by Lawitschka C., König P.)	255
<i>Ethical Problems of Using Artifical Intelligence in Medicine</i> (by Vvedenskaya E.)	255
<i>Recent Results and Activities in Trustworthy Artificial Intelligence</i> (by Lisi F.A.)......	256
<i>Components of the Digital Technological Revolution: Algorithm, Artificial Intelligence and Digital Communication, and its Impact between Young Mexicans</i> (by Huerta Vilchis F., Fernández Fernández Í.)	257
<i>The Dark Side of Consumer-Smart Object Relationship: a Non-User Perspective</i> (by Monsurrò L. et al.)	258
<i>Sociological View of Medicine of the Future</i> (by Prisyazhnaya N.V.)	259
<i>AI: Four Questions for the Great Challenge of the 21st Century</i> (by García-Tejedor Á.J., García Plá V.)	260
<i>A Taxonomy of Artificial Intelligence Opacity</i> (by Schneider M. et al) ..	261
<i>Artificial Intelligence and Sensitive Thought</i> (by Amendola G.)	262
<i>CA17124 DigForASP: A European cooperative action for AI Applications in Police and Digital Investigations</i> (by Olivieri R. et al)......	263
<i>Artificial Intelligence & Pluralistic Global Bioethics: Thomistic-Aristotelian Personalist Refinement of the United Nations' Social Contract View of Rights-duties in AI-genetic Engineered Nanotechnology</i> (by Monlezun D.J. et al.)......	264
<i>Ethical Problem of the Trademark Registration for "NEON Artificial Human"</i> (by Jin Y.).....	265
<i>Human-in-the-loop Artificial Intelligence</i> (by Zanzotto F.M.).....	266
<i>Does Artificial Intelligence Have a Purpose?</i> (by Gutierrez J.J.)	266
<i>Artificial Intelligence and the Future of Nursing Profession</i> (by Ahn S.H.)	267

Contents

<i>The Advent of Artificial Intelligence in Arts or the Creativity of Artifacts (by Mangione M.A., Carrara A.)</i>	268
<i>Accessible Numbers: Artificial Intelligence and Cultural Inclusion (by Baraldi L.)</i>	269
<i>In Tech we Trust...but we need Human as a Right (by Spiller E.)</i>	270

In Memory

Antonio G. Spagnolo <i>Remembering Cardinal Elio Sgreccia</i>	273
--	-----

Introduction

In a *Letter* addressed to the Pontifical Academy for Life on the occasion of its 25th Anniversary, Pope Francis described new technologies as one of the priorities that is “calling for study” (*Humana communitas*, 6th January 2020, n. 12). His request is rooted in the awareness that these are one of the most relevant features of the epochal shift that we are going through. In fact, current events are not only connected to the availability of new tools to carry out specific functions, rather they entail a transformation that affects the way we perform any activity. Digital technologies are in fact rapidly redesigning the world we live in, and creating a deep impact on many aspects of society: from the economy to security, from work organization to leisure, from the public administration to healthcare, from education to communication. Such an upheaval questions the perception of the reality and of ourselves that we take for granted, raising major ethical issues.

Given the complexity and novelty of the topic, the Pontifical Academy for Life deemed it appropriate to address the issues emerging in this field in two General Assemblies and their relative International Workshops. The 2019 General Assembly considered the relationship between robots and ethics, while the XXVI General Assembly (26th to 28th February 2020) concentrated on what is commonly referred to as “artificial intelligence” (AI). The latter was characterised not only by the current relevance of the topic, but also by a considerably new structure. In fact, after the traditional Workshop, an event was held to mark a concrete engagement, necessary to embrace and drive the transformations we are experiencing according to shared ethical criteria. This is the sense of the *Rome Call for AI ethics*. Its aim is to promote fundamental ethical principles for a greater common responsibility in this sector. However, before mentioning the contents and the sense of this event, let us briefly illustrate the Workshop structure that comprised three sessions.

Workshop Structure

A. Artificial Intelligence and Ethics

The first session aimed at exploring the anthropological and ethical scope of the transformations brought about by AI. The presentation of

Prof. Paolo Benanti, professor of moral theology at the Gregorian Pontifical University, raised some questions on the relationship between ethics and innovation. It is not simply a matter of identifying which of these values need to be protected and promoted in the current context, but also of designing new conceptual tools to drive technological evolution. The category of “algorethics” was illustrated as an example, in that it expresses the need to lead the whole development process of technological devices in a cross-disciplinary way. This is how the specific responsibilities of the stakeholders involved in each phase of the process are identified. The contribution of the Social Doctrine of the Church may be of particular importance, as highlighted by prof. Alexander Filipović, professor of media ethics at the Hochshule für Philosophie in Munich. In addition, in his speech, he identified some research areas that could be explored further to address the issues raised by data production and availability and their AI-based processing from the perspective of the Christian social Ethics. Prof. Robin R. Wang, professor of philosophy at Loyola Marymount University (California), offered an Asian perspective. According to Taoism, human life cannot be limited to data flows, as it is a complex system that can organise itself in a dynamic balance and non-linear procedures. Human intelligence is rooted in corporeality, that is why it should be separated from man-made, artificial intelligence that relies on computational procedures, instead. The Taoist approach paved the way for the subsequent presentation by prof. Sandra Alexander, professor of Human Science at the American University in Dubai (United Arab Emirates), who analysed a certain number of ethical consequences that digital technologies have on education. The pursued goal was to draft general guidelines to introduce the teaching of ethics in the age of AI, considering the multi-cultural and multi-religious context of today’s world.

B. Artificial Intelligence and Human Health

The first speaker of the session on human health was prof. Walter Ricciardi, professor of Public Health at the Università Cattolica del Sacro Cuore (Rome). His presentation showed the contribution that AI could offer to healthcare organizations, in terms of data management and an increasingly person-centred, custom-made medical practice. The impact of AI on the concrete clinical practice was addressed by prof. Yuzo Takahashi, Professor Emeritus of Dermatology at Gifu University (Japan). The key element of his presentation was the doctor-patient

relationship, and the diagnostic and therapeutic decision-making process. To safeguard a correct relationship, digital devices must be at the service of doctors, who maintain a role that is consistent with the respect they are owed, without being paternalistic. The two subsequent speakers, prof. Alexandru Floares, of SAIA Institute (Romania), and prof. Felix Rigoli, professor at the University of São Paulo (Brazil), addressed the strengths and weaknesses connected to the introduction of algorithms in data analysis for oncological diagnoses or in the access to care. Besides the greater ease of development, accessibility to tests and the greater predictive capacities of new technologies, the speakers highlighted the risks connected to the introduction of multiple biases. Care processes may produce distortions and exclude the most vulnerable. It is necessary to share an equal responsibility for device design in order to promote the universal right to health.

The session on health ended with a *Lectio magistralis* by prof. Shin'ya Yamanaka, 2012 Nobel Prize Laureate in Medicine, on the new perspectives opened up by Induced Pluripotent Stem Cells (iPSCs) also in view of the most recently developed AI techniques. The latest genome editing techniques and AI applications enabled the number of patients, as well as the range of diseases to treat to be extended thanks to this method.

C. Artificial Intelligence and Law

The great potential of digital technologies requires prudent regulation and governance strategies. Prof. James Shaw, from Toronto University, specifically discussed this theme, focusing on the relationship between health care systems and producers of AI devices. In particular, the need to strictly control health data transfers emerged quite clearly. In fact, there are risks related to people's security and profiling, with the relative consequences of economic exploitation. At the same time, as prof. Federico de Montalvo, professor of Constitutional Law at Comillas Pontifical University (Spain) stressed, we are questioning the very criteria produced and consolidated in recent history. The battle against disease that can be waged, as it currently is, by collecting, and analysing data prompts an evaluation of the balance, and the methods used to manage the relationship among the risks for personal rights, public health benefits and the common good from a new perspective. The new concept of "pseudonymization" is a good example of an emerging paradigm shift.

D. Education and Global Perspective

Education plays a key role. As prof. Francesco Profumo, Dean of Politecnico di Torino, highlighted, it is of crucial importance to reflect on how to best introduce technologies, and AI in particular, in educational processes. His presentation focused on three aspects – AI for education, Education for AI and Education to AI – which showed how relevant it is not only to have an operational knowledge of digital devices, but also to know and take up responsibility for their social and cultural impact. Dr. Cédric Wachholz, head of UNESCO's Digital Innovation and Communication Sector, set forth a global perspective, describing the steps that his Organization followed to support person-centred AI and to avoid any risk of exclusion, inequality and any breach of rights that AI may generate.

"Rome Call for AI Ethics"

As briefly mentioned above, the last morning of the meeting was devoted to the public presentation of the *Rome Call for AI Ethics*. A group of experts from different disciplines and backgrounds prepared the text for the Call, whose aim was to mobilize forces in order to address the profound changes that our world is experiencing, by promoting an ethical approach to AI and encouraging organisations, governments and institutions to take up their respective responsibilities in this regard. Indeed, only a broad collaboration among different stakeholders may build a future in which digital innovation and technological development are at the service of human creativity and genius, without gradually replacing them.

It is worth mentioning the literary genre of this document, which is neither a joint declaration, an agreement, nor a treaty. A *Call* cannot exist on its own; in fact, it only lives thanks to its partners' interaction. These partners identify a number of common difficulties that each encounter in their work and undertakes to address them collaboratively. Therefore, the text is not the property of anyone in particular, as it belongs to all those who commit to it. There is not one person exercising control over the others or in charge of enforcing the contents of the Call, rather each partner is one face of one and the same polyhedron on an equal footing with all the others. It may well be that not everyone is already aligned to the Call's recommendations. However, the signatories of a Call publicly take on responsibility for implementing such

recommendations in their concrete activities, even if this means, “paying the cost for it”. This process led to the call to recognise and take up the responsibility deriving from the options offered by the new digital technologies. In accordance with the exhortation of the Second Vatican Council (cf. GS, n. 3), the Pontifical Academy for Life set out on this common path, offering the contribution of the Church’s tradition and experience in the shared search for what is authentically human and can promote greater justice in today’s world.

A. Drafting of the text

The text is the result of a joint collaboration in which the Call’s partners identified common references, also in an effort to find a language that could be clear and shared by all. Human rights were a helpful inspiration, both in terms of content and a possible convergence at an intermediate level. In fact, although they do not lack fundamental anthropological elements, they nevertheless enabled the partners to find a possible convergence on different both cultural and religious worldviews. At the same time, they do not contain overly detailed and particular legal norms in their formulation.

The document follows three main lines. The first is ethical in nature and refers to the fundamental values of the Universal Declaration of Human Rights. Thus, this provides a reference framework that is necessary for any ethically valid technological development. In particular, the following elements are emphasised: inclusion, simultaneous focus on the good of the whole of humanity and of every human being, respect for and protection of the planet, ‘our common and shared home’. The second line concerns education for the younger generations, who will be deeply affected by the new technological resources, and shall need equal access to such technologies. Given the speed of the transformation, lifelong education will also be needed, especially for those at risk of being left behind. Education shall also include a focus on conscience and the motivations that enable it to promote the good of the community, even at the expense of its own interests. The final aspect is legal in nature: there is a clear need to translate the principles set out into effective regulations and to make them incisive through an ethical approach “by design”, i.e. one that accompanies every step in the technological production cycle from the very beginning.

B. Principles of reference

In order to achieve these objectives and provide more precise indications on how to operate in the field of AI ethically, a number of principles have been set out: 1) Transparency: in principle, AI systems must be explainable; 2) Inclusion: the needs of all human beings must be considered so that everyone can benefit and all individuals can be offered the best possible conditions to express themselves and grow; 3) Responsibility: those who design and implement these technologies must act with accountability and transparency; 4) Impartiality: avoid creating or acting according to prejudices, thus safeguarding fairness and human dignity; 5) Reliability: AI systems must operate reliably; 6) Security and privacy: AI systems must operate securely and respect users' confidentiality. These principles are fundamental elements for good innovation. We are pleased to observe that these principles are in line with the documents issued by various European Union bodies; and it is equally very interesting that large US companies praised their validity in various contexts, and embraced them.

The Congress opened with a letter of Pope Francis that was read out to the participants. In his letter, the Holy Father initially highlighted the risks of the new technologies that may not be intended as tools limited to individual sectors, but as forces that now go unnoticed, plunged as they are in our world. In this way, they can impose real forms of control and influence over mental and relational habits, which means that they do not only enhance cognitive and operational functions. Precisely for this reason, there is a need for joint research involving as many players as possible in this field. In this endeavour, the principles of the social doctrine of the Church provide a good starting point: "dignity of the person, subsidiarity, and solidarity. They commit to serve every person in her integrity and all the people, without discrimination, or exclusion".

The event concluded with the signing ceremony of the Call by the President of the Pontifical Academy for Life, Abp. Vincenzo Paglia, the President of Microsoft, Bradford Lee Smith, IBM Executive Vice-President, John Kelly III, FAO Director General, Qu Dongyu and the Italian Minister for Technological Innovation and Digitalisation, Paola Pisano. The President of the European Parliament, David Sassoli, also attended the ceremony and while expressing the interest of the European institutions in the process that had just been inaugurated,

stressed the importance of international collaboration among the Call's partners. The diversity of the signatories' roles and profiles clearly showed that stakeholders from the productive, institutional, political, scientific and academic worlds were involved. Each, in their specific role, as it emerged from their presentations that are collected in this volume, committed to disseminating the Call so that others may endorse it, thus participating in both the search for an ever deeper and common understanding of the changes we are experiencing and accepting the responsibility for them.

Address of His Holiness Pope Francis to Participants in the XXVI General Assembly of the Pontifical Academy for Life

*Distinguished Authorities,
Ladies and Gentlemen,
Dear Brothers and Sisters,*

I offer you a cordial greeting on the occasion of the General Assembly of the Pontifical Academy for Life. I thank Archbishop Paglia for his kind words. I am grateful too for the presence of the President of the European Parliament, the FAO Director-General and the other authorities and leaders in field of information technology. I also greet those who join us from the Conciliazione Auditorium. And I am heartened by the numerous presence of young people: I see this as a sign of hope.

The issues you have addressed in these days concern one of the most important changes affecting today's world. Indeed, we could say that the digital galaxy, and specifically artificial intelligence, is at the very heart of the epochal change we are experiencing. Digital innovation touches every aspect of our lives, both personal and social. It affects our way of understanding the world and ourselves. It is increasingly present in human activity and even in human decisions, and is thus altering the way we think and act. Decisions, even the most important decisions, as for example in the medical, economic or social fields, are now the result of human will and a series of algorithmic inputs. A personal act is now the point of convergence between an input that is truly human and an automatic calculus, with the result that it becomes increasingly complicated to understand its object, foresee its effects and define the contribution of each factor.

To be sure, humanity has already experienced profound upheavals in its history: for example, the introduction of the steam engine, or electricity, or the invention of printing which revolutionized the way we store and transmit information. At present, the convergence between different scientific and technological fields of knowledge is expanding and allows for interventions on phenomena of infinitesimal magnitude and planetary scope, to the point of blurring boundaries that hitherto were considered clearly distinguishable: for example, between inorganic and organic matter, between the real and the virtual, between stable identities and events in constant interconnection.

On the personal level, the digital age is changing our perception of space, of time and of the body. It is instilling a sense of unlimited possibilities, even as standardization is becoming more and more the main criterion of aggregation. It has become increasingly difficult to recognize and appreciate differences. On the socio-economic level, users are often reduced to “consumers”, prey to private interests concentrated in the hands of a few. From digital traces scattered on the internet, algorithms now extract data that enable mental and relational habits to be controlled, for commercial or political ends, frequently without our knowledge. This asymmetry, by which a select few know everything about us while we know nothing about them, dulls critical thought and the conscious exercise of freedom. Inequalities expand enormously; knowledge and wealth accumulate in a few hands with grave risks for democratic societies. Yet these dangers must not detract from the immense potential that new technologies offer. We find ourselves before a gift from God, a resource that can bear good fruits.

The issues with which your Academy has been concerned since its inception present themselves today in a new way. The biological sciences are increasingly employing devices provided by artificial intelligence. This development has led to profound changes in our way of understanding and managing living beings and the distinctive features of human life, which we are committed to safeguarding and promoting, not only in its constitutive biological dimension, but also in its irreducible biographical aspect. The correlation and integration between life that is “lived” and life that is “experienced” cannot be dismissed in favour of a simple ideological calculation of functional performance and sustainable costs. The ethical problems that emerge from the ways that these new devices can regulate the birth and destiny of individuals call for a renewed commitment to preserve the human quality of our shared history.

For this reason, I am grateful to the Pontifical Academy for Life for its efforts to develop a serious reflection that has fostered dialogue between the different scientific disciplines indispensable for addressing these complex phenomena.

I am pleased that this year’s meeting includes individuals with various important roles of responsibility internationally in the areas of science, industry and political life. I am gratified by this and I thank you. As believers, we have no ready-made ideas about how to respond to the unforeseen questions that history sets before us today. Our task is rather one of walking alongside others, listening attentively and seeking to link experience and reflection. As believers, we ought to allow

ourselves to be challenged, so that the word of God and our faith tradition can help us interpret the phenomena of our world and identify paths of humanization, and thus of loving evangelization, that we can travel together. In this way we will be able to dialogue fruitfully with all those committed to human development, while keeping at the centre of knowledge and social praxis the human person in all his or her dimensions, including the spiritual. We are faced with a task involving the human family as a whole.

In light of this, mere training in the correct use of new technologies will not prove sufficient. As instruments or tools, these are not "neutral", for, as we have seen, they shape the world and engage consciences on the level of values. We need a broader educational effort. Solid reasons need to be developed to promote perseverance in the pursuit of the common good, even when no immediate advantage is apparent. There is a political dimension to the production and use of artificial intelligence, which has to do with more than the expanding of its individual and purely functional benefits. In other words, it is not enough simply to trust in the moral sense of researchers and developers of devices and algorithms. There is a need to create intermediate social bodies that can incorporate and express the ethical sensibilities of users and educators.

There are many disciplines involved in the process of developing technological equipment (one thinks of research, planning, production, distribution, individual and collective use...), and each entails a specific area of responsibility. We are beginning to glimpse a new discipline that we might call "the ethical development of algorithms" or more simply "algor-ethics" (cf. Address to Participants in the Congress on Child Dignity in the Digital World, 14 November 2019). This would have as its aim ensuring a competent and shared review of the processes by which we integrate relationships between human beings and today's technology. In our common pursuit of these goals, a critical contribution can be made by the principles of the Church's social teaching: the dignity of the person, justice, subsidiarity and solidarity. These are expressions of our commitment to be at the service of every individual in his or her integrity and of all people, without discrimination or exclusion. The complexity of the technological world demands of us an increasingly clear ethical framework, so as to make this commitment truly effective.

The ethical development of algorithms – algor-ethics – can be a bridge enabling those principles to enter concretely into digital technologies through an effective cross-disciplinary dialogue. Moreover, in the encounter between different visions of the world, human rights

The “good” algorithm?

represent an important point of convergence in the search for common ground. At present, there would seem to be a need for renewed reflection on rights and duties in this area. The scope and acceleration of the transformations of the digital era have in fact raised unforeseen problems and situations that challenge our individual and collective ethos. To be sure, the Call that you have signed today is an important step in this direction, with its three fundamental coordinates along which to journey: ethics, education and law.

Dear friends, I express my support for the generosity and energy with which you have committed yourselves to launching this courageous and challenging process of reassessment. I invite you to continue with boldness and discernment, as you seek ways to increase the involvement of all those who have the good of the human family at heart. Upon all of you, I invoke God’s blessings, so that your journey can continue with serenity and peace, in a spirit of cooperation. May the Blessed Virgin assist you. I accompany you with my blessing. And I ask you please to remember me in your prayers. Thank you.

FRANCIS

Read by Abp. Paglia, President of the Pontifical Academy for Life, at the RENAISSANCE ceremony in Rome, on Friday, 28 February 2020.

© Copyright - Libreria Editrice Vaticana

Opening Address

Vincenzo Paglia *

Dear friends,

In the lives of each and everyone of us and the Academy as a whole, the past year has been intense and rich. From the bottom of my heart, I wish to thank you first of all for your daily scientific work, which, precisely because it is a laborious quest for truth, is at the service of human life. Thank you for your research studies, your publications, the many events you have promoted or participated in, and the working groups within the Academy. This immense, qualified and passionate research work is the main and truly precious service that every member of the Academy renders to the Church and the whole world.

Moreover, for the Academy, the period since the last General Assembly has been particularly intense and eventful. It was also the year of our 25th anniversary, on whose occasion Pope Francis addressed us the letter *Humana Communitas*, the latest of the many texts which the Pontiffs have addressed to the Academy in recent years, and which we have collected in a volume accompanied by a detailed index that will be published in the coming days.

Unfortunately, this year was also marked by the death of Card. Elio Sgreccia who for a long time generously devoted his energies to the Pontifical Academy for Life and Bioethics. We shall commemorate him together officially through the words of professor Spagnolo, whom I wish to thank in advance for this.

The Workshop which starts today was preceded by considerable reflections, listening and exchanges with a great number of stakeholders. As you will remember, our path was inaugurated last year with the workshop on Robo-ethics, whose Proceedings have been recently printed (Paglia V., Pegoraro R, (eds.), *Robo-ethics. Humans, Machines and Health*, Vatican City, Pontifical Academy for Life, 2020). In the next days, we shall proceed along the road paved to delve deeper into the specific topic of so-called Artificial Intelligence (AI). The strength and

* President of the Pontifical Academy for Life, Vatican City.

pervasiveness of the new technologies under whose shadow we all live, call for an in-depth anthropological and ethical reflection capable of withstanding the impressive speed of the progress of scientific research, and an exquisitely human wisdom, without which we risk an extremely serious process of dehumanisation, of which even stakeholders who are far from the Christian experience are beginning to be aware.

It is precisely thanks to the stimuli given to us by Pope Francis in his letter to the Pontifical Academy for Life, *Humanae Communitas*, that we have addressed the topic of Artificial Intelligence. This choice has involved a triple methodological process: we had to address new topics (to many of us virtually unknown), we had to learn a new language to try and re-articulate a vision of humankind, we had to find common ground making encounter and dialogue with the protagonists in this unprecedented technological development possible. A far from easy endeavour that has involved many of us, whom I warmly thank today.

I'll present more articulated and deep reflections on this issue of AI in my speech at the final session dedicated to "The Rome Call for AI Ethics", published further ahead in this volume at p. 213.

First session

ARTIFICIAL INTELLIGENCE AND ETHICS

Is a Good Algorithm also an Ethically Good Algorithm?

Adriano Pessina *

Good morning and welcome to the first session. I shall limit myself to a few introductory remarks, as it was asked of me, and I shall draw inspiration from the video¹ that we watched, which shall certainly raise many expectations. Technology has a huge potential and brings many novelties: that is why it is reasonable to imagine positive developments also in the fields of medicine and health care in order to provide support to people living in poor countries and challenging situations. However, the description of certain outcomes should not make us overlook the need to have, as already recalled, critical spirit. However, this expression needs to be clarified.

A critical spirit does not correspond to a negative, oppositional attitude, nor does it identify with an anachronistic refusal of technological advances, rather, it primarily expresses an intelligent willingness to analyse and assess the different aspects of information technologies. It is not sufficient to calculate the results, measure the potential, and highlight the problems that these new technologies entail: it is essential to understand how they transform our sensory, cognitive, emotional and relational experience.

In order to do that, I deem it very important to “describe” thoroughly the phenomena that we experience before expressing any judgement. In ethical matters, a description is not a tool for evaluation.

To this end, I deem it fundamental to overcome the idea of technology as a simple artefact, a neutral means to use for good or bad ends. As Günther Anders had already remarked in 1956, reflecting on the appearance of the radio and television, new technologies are not a mere means for the essential reason that they shape us, change the way we live, the way we think, our relationships, and our perception of the world. Information technology, the digital and virtual worlds are

* Professor of Moral Philosophy - Università Cattolica del Sacro Cuore, Milan (Italy). Ordinary Member of the Academy.

¹ The theme was presented by Jen Copestake in the video available at:
<http://www.academyforlife.va/content/pav/it/projects/robotics.html>.

new forms of reality that can have long-lasting effects on our life: they are, in fact, a new cultural milieu that influences our daily experience.

In addition, new technologies modify the way we relate to the human condition.

To this end, Günther Anders invented a beautiful expression: “Promethean shame”,² to indicate that the products that we build always look better than their authors. In a way, we “created” machines that seem more intelligent, more capable to govern life than we do. We have gone from “creating” to imitating our own creatures, while succumbing to their charm and power. Our products have gone from being a tool to increasingly becoming a model. In fact, defining an algorithm “intelligent” means to believe implicitly that intelligence can be compared to the capacity to compute; however, when we replace intelligence with calculations, we end up with confusing the truth with formal correctness. Nonetheless, what is formally correct may not be true. Even lies can be correct.

A living being’s intelligence is not comparable to the intelligence of a machine, and this difference should lead us to reflect, urging us to raise further questions.

Functions and performances increasingly enthrall us, to the point that any reference to those who carry out the functions remains indifferent: what matters is the result, they say. It seems that there is nothing special in being alive, in being flesh and bones, in living in a specific place and time. What do we make, then, of a doctor, a friend, a confessor in flesh and bones, if their functions can be performed – even much better – by a software, or artificial intelligence, that can always be at our disposal, do not fall ill, do not suffer from the coronavirus, and do not age or make mistakes?

We are always connected. However, can a connection replace a relationship? Our time is the era of proxy, and the technological proxy in treatment and care processes, that is the new frontier of artificial intelligence and robotics, makes us question the sense of our relationships. What do we mean by human relationships, care relationships, and affective relationships?

² Anders G. *The Outdatedness of Human Beings* (*Die Antiquiertheit des Menschen*). Munich, 1956.

Evoking the title of an interesting work,³ we have to go back to asking why we expect more and more from technologies and less and less from people.

We need to learn to question ourselves, without aligning to the famous theses of the “school of suspicion” on what lays behind new technologies, if we want to begin to understand them. What lays behind an algorithm?

First, people, computer scientists, and researchers lay behind an algorithm, behind artificial intelligence, as they decide and choose for us how to codify news, how to use and connect data.

What lays behind a team of software developers? Behind a team of developers always lay economic investors who expect to make profits.

And what lays behind each user? Behind each user, there is a possible consumer relying upon impersonal intelligence that is the result of multiple interests. And the whole technological production is directly addressed to this consumer of products, images, news and sensory stimuli.

Let me draw my conclusions: this conference bears a title that translates as “il buon algoritmo” in Italian. However, the Italian language is as rich in nuances as the human experience: if we slightly changed the sentence, we might ask: a good algorithm really is an ethically good algorithm? The two adjectives have different meanings: a good algorithm is a working system, an exact and correct process; while defining an algorithm “ethically good” means to describe it in ethical terms. A “good” rifle is a rifle that enables us to hit and kill with precision; however, this does not allow us to define a rifle “ethically good”.

I think that we should reflect deeply, because this title is so intriguing, especially in Italian, that it leads us to leave the question open: can a good algorithm really become ethically good?

We shall discover that over the next few days, we will learn by experience, as they say. Now, let me leave the floor to the speakers. Thank you for your attention.

(Translated from Italian by the Pontifical Academy for Life)

³ See Turkle S. *Alone Together. Why We Expect More from Technology and Less from Each Other*. New York, 2011.

Algor-ethics: Artificial Intelligence calls for an Ethical Reflection

Paolo Benanti*

From the ethics of technology to algor-ethics

Artificial intelligence is changing the world, as we know it: every human activity, from medicine to national security, is undergoing a profound transformation. AI systems not only help humankind, they create completely autonomous systems, bots or robots that deploy in an increasing number of scenarios. Before this downpour of artificial intelligence, it is urgent to address AI from an ethical perspective. The more universal AI, the more necessary the development of a new universal language to manage innovation.

We must begin by dispelling any possible misunderstanding. One of the most common misunderstandings in the field of ethics is considering it as a sort of chain to restrict freedom. Consequently, an ethics of technology would mean to impose limits on technology *a priori*. Is it really so? To understand what ethics of technology means, we must trace a path that started long ago.

According to anthropologists, 70 000 years ago, our species, homo sapiens, moved out of Southern Africa, the cradle of our existence, to colonize the world. We reached every corner of the globe in a unique way, exhibiting one of the peculiarities of our species. Up to that moment, every biological species lived in a special climate, its habitat. If a mammoth moved from the Siberian Steppe to Africa and India, it was because a member of its progeny underwent a genetic mutation, lost its fur, and thus became capable to survive in Southern, warmer climates. However, unlike mammoths, when man moved from Southern Africa, he reached every corner of the world, including the Siberian Steppe, without waiting for its descendants to grow a thick fur. In other words, he did not wait for the Homo sapiens hipster to appear. Man simply clothed himself in mammoth fur. Specifically, thanks to technolo-

* Professor of Moral Theology, Bioethics and Neuroethics, Pontifical Gregorian University (Rome, Italy); Corresponding Member of the Academy.

logical artefacts the human species succeeded in changing what other species received through their genetic codes. Other animals received their capacities thanks to their genes that can only change if the DNA changes. For us, it is different. We cooperate with one another; we pass on information on the world and educate the next generations to do things thanks to technological artefacts. While a dolphin can swim thanks to its DNA, man is different. Man changes himself and the world he inhabits through technological artefacts. Technology is the place where this is all condensed. We transform the world and ourselves through technology so as to inhabit the world. The ethics of technology is simply the natural basis for the technological background.

Every form of artificial intelligence is a technological artefact, each different from all the other artefacts produced thus far. All the tools that we have produced allow man to carry out some tasks. From primordial clubs to great industrial machines, all these tools were necessary to carry out precise tasks better, faster, and more effectively. AI, both in bots and robots, goes beyond the notion of artefact and machine that we have known so far. All the automated mechanisms that we built during the industrial revolution were developed with a specific purpose in mind. They only did what they were designed for. At present, AI is designed differently. It is not a programmed software, but rather a trained system. The classical if-this-then-that model in which a software engineer predicts any possible occurrence has been overcome. AI responds autonomously to the problem at hand. These artefacts are a new species of machines: Machine sapiens. Today, the world is no longer inhabited by Homo sapiens alone, but also by Machine sapiens. If the machine is autonomous, then who is accountable for its decisions? Its designer? Its user? Its seller? Its buyer? At present, the average AI machine can make a medical prognosis better than the average doctor can. Are we ready to delegate all these decision-making competences to machines? To respond to this question, we should clarify a basic issue: can artificial intelligence make the perfect choice?

Data scientists say that the problem lies in data quality and quantity. Machines shall make perfect choices only when we develop a perfect database to run AI services. Is it really so? We had this impression already in the past. Laplace stated that if at a specific moment, we had known the location of all the particles in the universe, then, we could have predicted the whole future and known the whole past of the universe. That is the famous Laplace's demon. Today, the question

applies to artificial intelligence and data-driven decision-making. What data does artificial intelligence use to make decisions? We can briefly say that data is a map of the world. Everything that exists in the world is mapped, recorded and placed in a database that represents its map. Can a map be the exact copy of the world? Let us leave the philosophical question on this possibility for the moment and address the matter from an operational standpoint. If we could create a map that is the exact copy of reality, and included everything in it, passers-by, tree leaves, etc., we should recognise that a map thus created is useless. In fact, it would be as complex as reality, too complex to make decisions, and therefore useless.

We would be confronted with the famous paradox told by Jorge Luis Borges in a fragment of "On rigor in Science", the last of A Universal History of Infamy, first published in 1935. In his usual style, the Argentinian author attributes the quotation to a book that does not exist in reality:

"...In that Empire, the Art of Cartography attained such Perfection that the Map of a single Province occupied the entirety of a City, and the map of the empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose seize was that of the Empire and which coincided point for point with it. The following Generations, who were not so fond of the Study of cartography, as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography."¹

Data is a map of reality, it represents a reduction of reality and as such, it is useful for decision-making. In addition, AI runs on databases and sensors. However, sensors cannot read the reality as a whole: they only take a portion of it and transform it into data. This is the key issue. As artificial intelligence makes data-driven decisions and since data is not a perfect copy of the reality, it is unthinkable *a priori* that an AI machine can make error-free choices. Machine sapiens shall always and integrally be fallible. AI needs ethics as its integral part. As artificial intelligence can make mistakes, it is necessary to understand how

¹ From the edition of Il Saggiatore, 1961.

these mistakes can be addressed. The ethical issue is fundamental, very important and urgent. We must find a shared ethical system so that the use of these systems does not produce injustices, harm people and create global unbalances.

What ethical trajectories can guide us towards developing a new human language that can connect Homo sapiens and Machine sapiens? The history of ethics can assist us in our search.

The first trajectory is what we might call Fear of the Uncertain. Any choice that we make has consequences. Everyone can choose freely, however, what happens once the choice is made does not always depend on us. Every free and conscious choice brings some uncertainty with it. One of the key ethical paradigms consists in finding ways to address uncertainty. This is the first ethical driver: the awareness that the choices made can also bring about unwanted effects, and the management of this risk.

A second, very important driver to consider is the tension between Equality and the Pursuit of Happiness. All the bloodiest wars of the 19th and 20th centuries were fought to achieve equality. As a matter of fact, such technologies risk producing inequalities. An ethics for AI must defend all this. Human dignity is the primary ethical value, not the value of data. Moreover, a state is legitimised if it enables individuals to pursue their happiness. These new technologies with their profiling possibilities and the capacity to predict the behaviour of human beings may make it difficult to live a free, individual life. We should not only consider the good and the bad that may arise for the individual (Fear of the uncertain), but for society as a whole: must we protect the equality among individuals and the possibility for each to pursue their happiness?

To conclude, we must be aware of a basic truth. Ethics is fragile by itself. Human dignity was crushed by 20th century totalitarian regimes, because no law protected it, at the same time AI ethics risks being ineffective, if it does not turn into binding policies to defend the individual and social coexistence.

The existence of Machine sapiens demands a new universal language to translate these ethical guidelines into guidelines applicable by machines. How can this come to be? Algorithms govern the world at the time of the Digital Age. Many talk of algo-cracy. To prevent algo-

rithms from dominating us, also thanks to AI, we must start to develop the common language of algor-ethics.

In order to shape algor-ethics, we must first clarify the meaning of value. In fact, algorithms operate according to numerical values. Ethics is about the moral value, instead. We must master a language that can translate the moral value into something the machine can compute. The perception of the ethical value is a purely human capacity. The capacity to elaborate numerical values is a skill of the machine. Algor-ethics can unfold, if we transform the moral value into something computable.

However, in the man-machine relationship, the true expert and bearer of value is the human side. Human dignity and values prove that man needs protection in the man-machine relationship. This provides a fundamental ethical imperative for the Machine sapiens: doubt yourself. We must enable the machine to experience a certain degree of uncertainty. Every time the machine does not know whether it is safely protecting the human value, then it should require man to step in. This fundamental directive is attainable by introducing statistical paradigms in AI. Google and Uber carried out similar attempts with special statistical bookshops. The capacity to be uncertain must be at the heart of the machine's decision-making skills. If every time that a machine is in a condition of uncertainty, it asks man, then we are producing a type of artificial intelligence that puts man at the centre or has, as experts say, a human-centered design. The fundamental rule is to build all AI machines in a human-centered way.

On the strengths of this basic rule, we might develop a new universal language: algor-ethics. This will have its own syntax and its specific literature. This is not the right place or the right time to explain what this language shall express; however, we believe that some examples could display its potential.

Anticipation - When two humans work together, one can anticipate and adapt to the other's actions, sensing his/her intentions. This competence is at the core of the ductility that characterizes our species: since the ancient times, it has enabled man to organise. In a mixed milieu, AI must be able to perceive what man wants to do, and adapt to his intentions, cooperating: the machine must adapt to man, not vice versa.

Transparency - Robots commonly work according to optimization algorithms: the energetic use of their servomotors, cinematic trajectories and operational speed are calculated for maximum efficiency in achieving their goals. If man wants to coexist with machines, their actions must be intelligible. The robots' main goal should not be to optimize their actions, rather to make them understandable and perceivable by man.

Customization - Through AI, a robot can relate to the environment, adapting its behaviour. If man and machines coexist, robots must be able to adapt also to the personality of the people they cooperate with. *Homo sapiens* is an emotional being; *Machine sapiens* must recognise and respect this unique and peculiar trait of their work partner.

Adequation - A robot's algorithms govern its line of conduct. In a shared milieu, the robot must adapt its objectives, observing the person and understanding the relevant goal in every specific situation. In other words, the machine must acquire "artificial humility" and give operational priority to the persons present, not to the achievement of a given goal.

In the age of AI, these four parameters set an example for the protection of human dignity. The problem is firstly philosophical and then, epistemological. AI "operates" according to data-connecting schemes. What sort of knowledge is this? What is its value? How should it be treated and considered?

Hence, the question is ethical before being technological: insofar as we intend to entrust the typically human skills of understanding, judgement and independent activity to AI software systems, we must comprehend the value, in terms of knowledge and capacity to act, of those systems that claim to be intelligent and cognitive.

At present, AI has developed in a market-driven or state-driven scenario. We must consider other development modes, for instance, deploying independently controlled algorithms that can certify these four capacities of machines. Or it is possible to think of independent third parties, who by writing dedicated algorithms can evaluate the suitability of AI to coexist with man. Only by respecting these indications, shall innovation proceed to promote an authentic human development.²

² From the blog paolobenanti.com, February 15th 2019.

How to use synthetic data

The large sets of visual data, made up of images and videos that are the assets accumulated by the most powerful Tech Giants in the AI market, provide a huge competitive advantage. These databases have created a rift that keeps the progress of automatic learning out of the reach of many players. This advantage seems destined to be reversed by the onset of synthetic data. What are the challenges for the philosophy and ethics of AI?

The most important tech giants in the world, such as Google, Facebook, Amazon, just to name some of the big players, are developing computer vision and AI to train computers. They collect immense visual data sets, composed of images, videos and other visual data from their consumers and use them to train their algorithms. These databases offer a competitive advantage to tech giants: these assets allow them to keep many competitors away from the advances of machine learning and the processes that enable computers and algorithms to learn more rapidly.

However, if we look at what is happening, this advantage might disappear thanks to the possibility for anyone to create and exploit synthetic data to train computers. Synthetic data can efficiently train algorithms in many scenarios, including retail, robotics, self-driving cars, trade and much more.

Synthetic data is computer-generated and reproduces – or rather simulates – real data; in other words, data is created by a computerised simulation, not by a human being or a real activity. Today, we can design software algorithms to create simulated data, or according to this wording, realistic synthetic data.

Data scientists and software engineers use synthetic data to teach a computer to respond to certain situations or criteria, replacing these to the training data captured in the real world. In the training process, one of the most important aspects both of real and synthetic data, is to have precise labels or tags, so that computers can translate visual data into specific meanings.

A number of companies use artificial vision, machine learning, and artificial intelligence to analyse the visual data of any business sector: healthcare, robotics, logistics, cartography, transportation, manufacturing

and many more. Many start-ups with really innovative ideas experience the problem of cold bootstrap, because they do not have enough, good-quality labelled data to train their algorithms: a system cannot draw any inference for users or items on which it has not collected sufficient information yet. Start-ups can gather relevant data contextually or collaborate with others to collect relevant data. For example, they may resort to retailers for data on consumers' purchasing habits or to hospital for medical data. Many start-ups in the early stage try to solve their cold bootstrap problem by creating data simulators to generate contextually relevant data with quality labels in order to train their algorithms.

This is not a problem for the Big Techs, as they expand their sources to gather unique and contextually relevant data exponentially.

The advances of synthetic data is remarkable. Serge Belongie, professor at Cornell Tech, who has researched computer vision for 25 years, said:

"In the past, our field of computer vision cast a wary eye on the use of synthetic data, since it was too fake in appearance. Despite the obvious benefits of getting perfect ground truth annotations for free, our worry was that we'd train a system that worked great in simulation but would fail miserably in the wild. Now the game has changed: the simulation-to-reality gap is rapidly disappearing. At the very minimum, we can pre-train very deep convolutional neural networks on near-photorealistic imagery and fine tune it on carefully selected real imagery."³

For example, AiFi is a start-up in its initial development phase and is building a platform for computer vision and artificial intelligence to provide a more efficient, check-out-free solution to family-run shops and large stores alike. They are developing a payment solution without check-out similar to Amazon Go. AiFi's solution to create synthetic data has become one of their defensible and differentiated technological advantages. Thanks to AiFi's system, buyers shall enter a retailer and select objects without using cash, a card or scanning barcodes. These smart systems shall continuously monitor hundreds or thousands of

³ Nisselson E. *Deep learning with synthetic data will democratize the tech industry*; TechCrunch (accessible on 10.10.2020 at: <https://techcrunch.com/2018/05/11/deep-learning-with-synthetic-data-will-democratize-the-tech-industry>).

buyers in a store and recognise or “identify” them again during a complete shopping session.

Ying Zheng, AiFi co-founder and chief science officer, had previously worked at Apple and Google. The entrepreneur said:

The world is vast, and can hardly be described by a small sample of real images and labels. Not to mention that acquiring high-quality labels is both time-consuming and expensive, and sometimes infeasible. With synthetic data, we can fully capture a small but relevant aspect of the world in perfect detail. In our case, we create large-scale store simulations and render high-quality images with pixel-perfect labels, and use them to successfully train our deep learning models. This enables AiFi to create superior checkout-free solutions at massive scale.⁴

Robotics is another sector leveraging synthetic data to train robots for various activities in factories, warehouses and across society. Josh Tobin is a research scientist at OpenAI, a non-profit artificial intelligence research company aiming to promote and develop friendly AI in such a way as to benefit humanity as a whole. Tobin is part of a team working on building learning robots. These machines have trained entirely with simulated data and deployed on a physical robot, which, amazingly, can now learn a new task after seeing an action done once. They developed and deployed a new algorithm called one-shot imitation learning, allowing a human to communicate how to do a new task by performing it in virtual reality. Given a single demonstration, the robot is able to solve the same task from an arbitrary starting point and then continue the task.

Their goal was to learn behaviours in simulation and then transfer these learnings to the real world. The hypothesis was to see if a robot could do precise things just as well from simulated data. They started with 100% simulated data and thought that it would not work as well as using real data to train computers. However, the simulated data for training robotic tasks worked much better than they expected.

Many large tech giants, car producers and start-ups are competing to produce self-driving cars. Developers have realized that there are not enough hours in a day to gather enough real data of driven miles

⁴ *Ibid.*

needed to teach cars how to drive themselves. May Mobility is a start-up building a self-driving microtransit service. Their CEO and founder, Edwin Olson, says about synthetic data:

“One of our uses of synthetic data is in evaluating the performance and safety of our systems. However, we don’t believe that any reasonable amount of testing (real or simulated) is sufficient to demonstrate the safety of an autonomous vehicle. Functional safety plays an important role. The flexibility and versatility of simulation make it especially valuable and much safer to train and test autonomous vehicles in these highly variable conditions. Simulated data can also be more easily labelled as it is created by computers, therefore saving a lot of time.”⁵

To date, the major platform companies have leveraged data moats to maintain their competitive advantage. Synthetic data is a major disruptor of these advantages, as it significantly reduces the cost and speed of development, allowing small, agile teams to compete and win.

The challenge and opportunity for start-ups competing against giants is to use the best visual data with correct labels to train computers accurately for diverse uses. Simulating data will level the playing field between large tech giants and start-ups. Over time, large companies will probably also create synthetic data to augment their real data, and one day this may tilt the playing field again.

However, the matter becomes philosophical and ethical in nature. The first point is epistemological. AI operates finding meaning and assigning correlations to large data sets. Now, this meaning and its epistemological value is problematic *per se* and should be thoroughly understood. The extent to which AI offers knowledge and the type of knowledge it offers is a matter we discussed several times on this blog and is a fundamental theme of my book *Oracoli*, collection Collassi, Luca Sossela editore.

Now, the matter at hand is even more complex: the virtual world offers knowledge on the reality with a pretence of truthfulness and guidance for the autonomous actions of AI algorithms.

Besides this further epistemological point, a major ethical issue arises. If AI and its automated systems raise ethical issues, as we have already

⁵ *Ibid.*

seen, AI trained on synthetic data plays out new unsettling scenarios: how can we grant a correct training of algorithms? How can we grant the safety of the systems currently in production if they have never been really tested?

Finally, we should ask if the consumer or user should be informed of these underlying features of AI: should we come up with a label to tell users that the system they are using was trained or is based on synthetic data? Is this a democratization of AI or an elegant marketing term that hides the willingness to produce business with technologically less appropriate or more fragile systems?

The scenario gets more complex: to manage these complexities an adequate philosophy of algorithms and algor-ethics are all the more necessary.⁶

The erosion of reality

In the face of the transformative and pervasive power of artificial intelligence, several people, not all technology experts, have started to launch appeals, cast shadows or demand that regulations be introduced to govern this new and fascinating technological development. However, there are several differing analyses of the phenomenon. To quote a famous essay of the 1960s⁷ by Umberto Eco, we may say that there are two equally large groups of alarmists and AI supporters. Alarmists raise the alarm about how AI shall put an end to our society or to the human species itself, many insist on the future of work and the advent of robots. Is this scenario to fear? Perhaps, the most radical and imminent transformations are of a different kind.

To be clear from the start, I must say that I am not convinced at all that the advent of what many call “super intelligent” artificial intelligence is anywhere near, or that the future generations of machines operating on deep learning models actually represent the most pressing threat for man. In fact, to be precisely frank, I am not sure at all that the whole notion of super intelligence is achievable and perhaps, it may be nothing more than an important philosophical hypothesis to

⁶ From the blog paolobenanti.com, May 22nd 2018.

⁷ Eco U. *Apocalittici e integrati*, 1964.

encourage our reflection, nothing more than an academic scenario to reflect on. We do not know if such AI shall ever be created, developed or implemented in the future - here on the Earth or somewhere else in the Cosmos. However, even though this author would not define himself an alarmist or supporter, the development of such a pervasive and transformative technology has the potential to change our society radically, as well as the relationships between man and the understanding that we as species have of ourselves. From my perspective, the most forthcoming and radical transformation that AI may produce is the radical distortion of what we consider truth and the ways that we, as men, share to seek it.

At present, in the relationship with what many call "intelligent" machines, we must recognise that we do not have a convincing quantitative theory of intelligence. There is no theory that tells us what we mean by intelligence ("look, it's intelligent, it can open a can of beans by itself"), nor one that gives us a measure of intelligence, effectively relating it with complexity, nor whether there is a theoretical maximum. Perhaps, the answers to these questions may come only from adequate experiments and the development of that fundamental theory of intelligence, which, as we said earlier, is still lacking.

For these reasons, I am not so worried about the advent of super intelligent AI on earth, however, I consider with great attention the spread of relatively stupid AI, whose goals – or better – incidental abilities – can manipulate our relationship with information, with what we call facts and with the reality, as we perceive it. This dimension of our lives may be the most threatened by AI.

To understand how, allow me to make a digression on the sociological concept of trust and on how the society and beliefs connect to this concept.

The notion of trust plays a primary role in the Western political and social thought. The contractual theories of the XVII and XVIII centuries considered trust as a fundamental pre-requisite of the political order and the foundation of the social contract. In addition, the founding fathers of sociology who were more interested in identifying the moral element permeating the social order implicitly refer to it. The contents of systemic or impersonal trust are generally defined as expectations of stability of a given natural and social order, a reconfirmation of the per-

formance of its rules. Therefore, these are far-reaching and generalised expectations of regularity.

From the cognitive standpoint, trust stands between complete knowledge and complete ignorance. Trust-based expectations act on uncertainty, not providing the missing information, but replacing it with a form of inner “certainty” that provides a positive reassurance vis-a-vis contingent events and experiences. Uncertainty is rendered tolerable with a replacement that reduces complexity in view of more gratifying predictions for the agent. Therefore, a trust-based expectation replaces uncertainty with a degree of “certainty” and inner reassurance that varies according to the degree of trust given. However, this represents a cognitive investment higher than mere trust. Consequently, in case of error, it ends in more severe, negative, motivational consequences.

The spreading AI capacities refer to the formation, development and maintenance of individual and social trust. Techniques such as conflicting learning have already produced AI that can imitate our voices to perfection. Similar approaches may presumably apply to our writing style, text messages and the posts we publish on social media. By spoofing our looks, thanks to the analysis of our photos, we can generate fake photos or videos in which we apparently do things that we have never done.

Probably, these systems shall continue to evolve (if they have not already done so, it is difficult to keep up with the developments). Then, why not generate news or entire columns of gossip with AI? Hollywood’s tabloids hardly ever report on facts and mainstream news at times seem to follow suit.

AI has the extraordinary potential to generate potentially misleading communication flows. Stealing our personal data swindling us, or creating an alternative version of ourselves that engages in any sort of anti-social, even criminal behaviour, or simply manipulating us to prompt us to desire certain goods or voting in a certain way or believing certain things. If we had to develop the first evangelical AI, this could easily outperform the best human preachers. And unlike hypothetical super-intelligence (whose motivations are hard to imagine), using AI to exploit people or lead the society follows a very ancient political model.

We, humans, have probably eroded the reality since our predeces-sors, Homo sapiens, started to communicate and tell stories. A good story told around a fire may help keep a verbal story alive or structure moral and social rules, bringing unity to our families and groups. Inevitably, though, it may mislead, distort and manipulate.

Let us follow a reflection on the theme by Yuval Noah Harari. Some 70 000 years ago, our ancestors were insignificant animals. The most important thing to know on prehistoric man is that he was not important at all. His impact on the world was slightly higher than that of jellyfish. Today, we dominate this planet. The question is: how did we make it? How did we go from insignificant apes, busy minding their own business in a corner of Africa, to rulers of the planet Earth?

Humans control the planet because they are the only animals capable to collaborate flexibly and in large groups. Yes, there are other animals, such as social insects – bees, ants – that collaborate in large numbers, but they do not do it flexibly. They collaborate in rigidly pre-set schemes. Other animals, such as social mammals, wolves, elephants, dolphins, chimpanzees, can collaborate more flexibly, but they do it in small groups, because the collaboration among chimpanzees relies on an intimate mutual acquaintance.

The only animal that can combine these two abilities, i.e. collaborating flexibly and in large groups, is Homo sapiens. But how do we do it exactly? What allows us, unmatched in the animal world, to collaborate thus? The answer is our imagination. We can collaborate flexibly and with an infinite number of strangers, because only us, in the animal kingdom on this planet, can create and believe in fiction, in imagined stories. If everyone believes the same story, then everyone obeys and follows the same rules, norms and values.

However, biologically, altering the reality is a behaviour that is not limited only to our “intelligent” species. Deceit is largely widespread in the natural world. Animals camouflage, or pretend to be what they are not – from mimicking poisonous species to fluffing up their feathers, scales or skin, the males of many species decorate themselves or build seductive structures or resort to subterfuge to disseminate their genes. Deceit seems to be an integral part of Darwinian selection, as much as honesty. The capacity to mislead is a measure of evolutionary adaptability.

We should not have much hope of obtaining machines that are qualitatively different from the models they are trained on. If we create artefacts that are only designed to optimize the result or win in a decision-making simulation game, perhaps we should urgently question ourselves on the social consequences that such systems may have in the hyper-connected social fabric that we inhabit, if they were to spread. Naturally, as physicist Niels Bohr said, it is terribly difficult to make good predictions, especially on the future. However, one thing is for sure, we will learn a great deal about the path that AI can take – clearly, assuming that we can see, know, and tell the truth.⁸

(Translated from Italian by the Pontifical Academy for Life)

⁸ From the blog paolobenanti.com, March 07th 2018.

Ethical and Social Consequences of Artificial Intelligence: Insights from a Christian Social Ethics Perspective

Alexander Filipović*

The continuous process of global social change seems to be gaining momentum in recent decades. The effects of industrialization and globalization have been an ongoing challenge for societies. Digitization can be seen as yet another significant social shift calling for an ethical assessment and political governance.

While industrialization has laid the foundations of Christian Social Doctrine (*Rerum Novarum*, 1891) and globalization belongs to major focal points of theological Christian social ethics as well as social teachings of the church, digitization seems to be lagging behind these phenomena in gaining the attention of theological social ethics. How Christian social ethics could contribute to the ethical and political discourse on digitization and artificial intelligence remains an open question.

By combining the issues of personality, solidarity, and subsidiarity with ethical considerations on technology, this paper aims at exploring the field of algorithms, data, and artificial intelligence (AI) through a prism of Christian social ethics. Therefore, the research goal is to discuss links between Christian social ethics, digitization, and artificial intelligence in order to shed some light on the current position of these digital phenomena in the tradition of Christian social ethics and to highlight the most promising directions for further research.

Given this rationale, three interrelated questions seem to be worth consideration: how to embed social ethics of artificial intelligence into social ethics of technology; how the issues of power and justice, which can be regarded as core challenges of artificial intelligence from a moral perspective, should be approached; and, how ethics of responsibility, which assigns different areas of responsibility to different actors, could serve as a possible solution to the aforementioned concerns.

* Professor of Social Ethics, Institute for Systematic Theology and Ethics, Faculty of Catholic Theology, University of Vienna (Austria). For assistance with this text I thank Cindy-Ricarda Roberts.

Thus, this paper consists of three main parts. First, a brief description of Christian social ethics will be presented. Secondly, an analysis of the challenges that technology may pose to Christian social ethics will be discussed. To round up, specific socio-ethical concerns over artificial intelligence will be addressed.

The following considerations are primarily of an exploratory, i.e. investigative character. Drawing attention to the aforementioned issues and collecting relevant approaches is at the core of this work.

Christian social ethics as theological ethics with a focus on problems of justice

The roots of explicit ecclesiastical social doctrine originate in the encyclical letter *Rerum Novarum* from 1891 in which pope Leo XIII raised concerns not only over the industrial revolution but also over the “transformation of the world”, as Jürgen Osterhammel pointed out.¹ Nevertheless, the church has a long tradition of observing and commenting on social challenges and shifts.

To mark the jubilee years of *Rerum Novarum*, the social encyclicals have been issued; among these, *Quadragesimo anno* (1941), *Octogesima adveniens* (1971) and *Centesimus annus* (1991) seem to be of the greatest importance. This series focused mainly on the themes of work and economy. A second series of social encyclicals started with *Populorum Progressio* (1967), and the anniversaries are accompanied by *Sollicitudo rei socialis* (1987) and *Caritas in veritate* (2009).² This series focused on the ethics of development in the context of globalization. In *Laudato si'* (2015), Pope Francis explicitly takes up the ecological question.

This kind of church social proclamation can be accompanied by two other dimensions of ecclesiastical social teaching. Firstly, practicing solidarity by local churches, the church associations, and federations as well as the parishes (social Catholicism) should be highlighted. These activi-

¹ Osterhammel J. *The transformation of the world: A global history of the nineteenth century*. America in the world. Princeton, Oxford: Princeton University Press; 2014.

² Originally the encyclical was to appear in 2007, but revisions delayed publication. On *Caritas in veritate* as social encyclical see issue 3 (2009) of the journal *Amosinternational*, online at https://www.amosinternational.de/user/pages/02.magazine/issue-2009-3/Amos_1275_2009_3.pdf.

ties unveil the applicable and real-life dimension of social practicing and its importance to social life of people and local communities. A vivid example of this kind of practice can be seen in the actions undertaken by the Christian workers' associations around 1900 or the contemporary local environmental groups in the parishes, which implement and follow what is proclaimed by the Magisterium.

Another expression of the church social teaching is Christian social ethics, which has been established as a scholarly subdiscipline of theology in the German-speaking world at the end of the 19th century. The theological subject is known under various names, including Christian social ethics, Christian social teaching, or Christian social sciences. The first professorship for "Christian Social Doctrine" was filled in 1893 in Münster by Franz Hitze, what was also a reaction to the first papal social encyclical *Rerum novarum* (1891).

Although both social ethics and moral theology hold established and prominent places in the canon of theological subjects, they differ significantly from each other, as social ethics focuses thematically on genuinely social or socio-structural questions. Therefore, Christian social ethics does not deal with individual questions of lifestyle, but rather with the question of "fair social institutions"³ – a statement that is also reflected in Arno Anzenbacher's "The central question is thus: Are given institutional structures just?"⁴ Christian social ethics reflects how goods, opportunities, rights and duties are distributed in (world) society. It deals with the question of when people should be treated equally or when they should be treated unequally. The perspective of power is thus included: Who has power, who is powerless?

The focus on justice is biblically founded: "As a theologically founded reflection on social institutions [...] it is situated in a certain horizon of world understanding shaped by the biblical faith in God. In it a relationship between God, man and the world is unfolded that opens up a meaning for human existence [...]."⁵

³ Heimbach-Steins M. *Wozu dieses Buch?* in *Christliche Sozialethik. Ein Lehrbuch. Bd. 1. Grundlagen*. Regensburg: Pustet: 2004; 7-18: 7.

⁴ Anzenbacher, A. *Christliche Sozialethik: Einführung und Prinzipien*. 1998; Paderborn u. a.: Schöningh: 15. Own translation. In the German original: «Die zentrale Frage lautet also: Sind gegebene institutionelle Gebilde gerecht?»

⁵ Heimbach-Steins M. *Biblische Hermeneutik und christliche Sozialethik*. in *Christliche Sozialethik. Ein Lehrbuch. Bd. 1. Grundlagen*. Regensburg: Pustet: 2004; 83-110: 88. Own

Social ethics as a theological-ethical discipline does not, therefore, make appeals to people about individual good behaviour, but considers on a theological, philosophical and social scientific basis how structures can be changed "politically" so that people can lead a good and just life. In doing so, it works in a strongly interdisciplinary way and lets itself be challenged by the *signs of the times*. Its normative foundations are personality, solidarity, subsidiarity, and sustainability, all of which are specifically directed towards justice. Personality as a basic Christian social-ethical principle can therefore be easily summarized in the term *justice to people as persons* (in german: "Persongerechtigkeit").

A final commentary on the relationship between the three dimensions of church learning and teaching on social issues (magisterium, social practice of the congregation, social ethics as theological science) concludes this sketch on the concept of Christian social ethics: Social ethics as theological-ethical science is mutually connected with magisterium and congregational practice. It stands in a critical-loyal relationship to the social proclamation of the Magisterium: Through the social proclamation it receives impulses and hints, but also reveals its inconsistencies and weaknesses. Not infrequently theological social ethicists collaborate on the texts of the church's social proclamation (in the first half of the 20th century for instance Gundlach and Nell-Breuning). The critical reception of the Church's doctrinal social proclamation through theological social ethics makes it accessible to scientific theology, but also to the critical public.

The issues of social learning and teaching of the Church (Magisterium, social community practice and theological-scholarly social ethics) have already been addressed and cover the whole range of social problems. It is interesting, however, that technics and technology are rarely explicitly addressed. It is true that Furger and Heimbach-Steins formulate in the preface to the 1996 Yearbook for Christian Social Sciences: "The 37th volume of the Yearbook is dedicated to the complex of topics 'technology ethics', certainly one of the most important future topics

translation. In the German original: «Als theologisch gegründete Reflexion auf die gesellschaftlichen Institutionen [...] ist sie in einem bestimmten, durch den biblischen Gottesglauben geformten Horizont des Weltverständens situiert. Darin wird ein Beziehungsgefüge von Gott, Mensch und Welt entfaltet, das eine Sinngebung für menschliche Existenz [...] erschließt.»

that a Christian social ethics must face".⁶ However, little has happened since then. The relevant textbooks do not explicitly deal with technology, I am not aware of any dissertations in the field of Christian social ethics that are intensively devoted to the subject, and the Compendium of the Social Doctrine of the Church hardly compiles anything on the subject.⁷ Obviously, up to now, technology and technics have not been understood as urgent questions of justice. In the encyclical *Laudato si'*, however, Pope Francis formulates a few impressive paragraphs on the topic, which clearly go beyond the field of environmental issues.⁸ We will come back to this.

Technology and artificial intelligence as a specific challenge to Christian social ethics

What has been described in the previous section as the social learning and teaching of the Church, gives us in the following the view with which we approach our theme. Christian social ethics is not interested in objects and (technical) artefacts, but in social structures, orders and institutions. In order to break down the phenomenon of artificial intelligence from a socio-ethical point of view, we need a view that helps us to find out how AI influences human coexistence on a structural level.

Socio-structural problems of justice and power are nowadays mainly influenced by engineering and technology. This has already been problematized in various ways. This can be followed up. An essential preliminary decision for the further procedure is therefore to understand "artificial intelligence" as a technique or technology and to break down this topic from there. Thus, we do not take the current "hype" about the topic of "artificial intelligence" or individual achievements of computer systems, which appear to us today to be astonishing, as our starting

⁶ Furger F., Heimbach-Steins M. *Vorwort. Jahrbuch für Christliche Sozialwissenschaften* 1996; 37: 7-10: 7. - Own translation. In the German original: «Der 37. Band des Jahrbuchs ist dem Themenkomplex „Technikethik“ gewidmet, sicher einem der wichtigsten Zukunftsthemen, denen sich eine christliche Sozialethik stellen muß.»

⁷ A few indications are given in Chapter 10, cf. Pontifical Council for Justice and Peace, ed. 2004. *Compendium of the social doctrine of the Church*. Vatica City.

⁸ See in particular Chapter 3, Section I: Technology: Creativity and Power and Section II: The Globalization of the Technological Paradigm (Numbers 101-114) from Franciscus PP. *Litterae Encyclicae Laudato Si': De Communi Domo Coldenda*. *AAS* 2015; 107: 847-945.

point, but let the specific technological nature or technological nature of "artificial intelligence" and other digital technologies guide our view.⁹

Since no one can have a disinterested view of technology and since, moreover, we can *find* the language used to talk about technology and cannot invent it, we have to take a look at the subject. Part of the scientific responsibility is that we explicate what view we have and why we have it. *Thus, from a Christian social-ethical perspective, we are given a person-centred, justice-oriented and power-critical perspective with which we reconstruct technology as a specific challenge.*

But first we prepare this perspective with a general technical-philosophical and -ethical perspective (2.1.), a theological-hermeneutical reflection on AI as a sign of the times (2.2.) and finally we collect these results with a social-ethical intention (2.3.).

Technology and Society: Technical-ethical Perspective

In order to get to the bottom of the specifically technical or technological or to get to the bottom of the technicity or technologicity, a whole range of approaches are available. The discipline in question is the philosophy of technology, which as a theoretical philosophy of technology analyses what we can know about technology (with the aim of explaining this knowledge) and which as a practical philosophy of technology asks what we should do about technology.

Philosophy of technology is traditionally not an outstanding subject of philosophy. It is rather implicit in philosophy of nature, anthropology and other classical subjects. Some philosophers even diagnose fears of contact between philosophy and technology.¹⁰ According to the relevant diagnoses of Karl Marx in the middle of the 19th century, a cultural-

⁹ I owe an impulse for this approach to Clifford Christian. In his book "Media Ethics and Global Justice" he criticizes that media ethics takes the technical revolution (digitalization) into account intensively but does not pay enough attention to the philosophy of technology. For this reason, he says, it falls short of its potential. He formulates: "In presenting a new perspective on international media ethics, this book demonstrates why and how our theorizing gives the philosophy of technology intellectual priority". (Christians. C.G. *Media Ethics and Global Justice in the Digital Age*. Cambridge: Cambridge University Press: 2019; 3). Here I suggest analogously: If Christian social ethics wants to deal competently with artificial intelligence, *it should give intellectual priority to the philosophy of technology*.

¹⁰ Cf. Bahr H.-D. *Über den Umgang mit Maschinen*. Tübingen: Konkursbuchverl: 1983; 14f.

philosophical tradition (Max Scheler, Karls Jaspers, Ortega Y Gasset), an anthropological (Arnold Gehlen, Lewis Mumford) and a social tradition can be distinguished (Jacques Ellul, Helmut Schelsky, Herbert Marcuse, Jürgen Habermas, Andrew Feenberg). In addition, Martin Heideggers, Stanislaw Lems, Günter Anders' and Hans Joas' analyses of technology should be mentioned.

Inevitably, even this extremely rough compilation of philosophical approaches to the phenomenon of technology poses problems of selection; all approaches differ, in part quite considerably, and all would be worthy of analysis. Since, as described, Christian social ethics refers back to problems of justice and social phenomena, I will briefly outline three perspectives that are continuative and connectable to this.

1) In contrast to instrumental theories, which presuppose or emphasize the neutrality of technology, the technical philosophies of Jacques Ellul and Martin Heidegger can be characterized as "substantial" theories. Such a substantial theory of technology would like to draw attention to the difficulties of that construction which sees technology as a neutral means of a desirable increase in efficiency and as modernization. The substantial theories assume that such understandings have a cultural character and thus draw attention to the constitutive effects of technology itself: "The issue is not that machines have 'taken over,' but that in choosing to use them we make many unwitting commitments. Technology is not simply a means but has become an environment and a way of life."¹¹ The term "technique" in Ellul's work, on which I concentrate in these short sections, then logically stands not for objects but for methods: "The term technique, as I use it, does not mean machines, technology, or this or that procedure for attaining an end. In our technological society, technique is the totality of methods rationally arrived at and having absolute efficiency (for a given stage of development) in every field of human activity. Its characteristics are new; the technique of the present has no common measure with that of the past."¹²

In this sense, technology is "a distinct pattern of reality whose consequences are real, whose formations and deformations are obvious and tangible."¹³ Ellul sees technology not as an isolable fact, but as deeply

¹¹ Feenberg A. *Transforming technology: A critical theory revisited*. New York, NY: Oxford Univ. Press: 2002; 7f.

¹² Ellul J. *The Technological Society*. New York: Vintage Books: 1964; xxv.

¹³ Langenegger D. *Gesamtdeutungen moderner Technik - Moscovici, Ropohl, Ellul, Heidegger: Eine interdiskursive Problemsicht*. Epistemata, vol. 75. Würzburg: Königshausen &

connected to every factor of modern life. Therefore he understands technology as a sociological phenomenon and pleads for studying technology in this way.¹⁴ Conceptually, he therefore makes a strict distinction between *technique* and *technology*: For Ellul, *technique* is a basic pattern of reality and *technology* is technical processes and objects as a whole.

Ellul structures the sociological field of technology with philosophical intention and philosophical methods. Technology as a pattern of generating reality plays a role in all areas of social reality: economy, politics as well as human orientation and culture. Ellul is thus in search of the basic provisions of the present, he is looking for the "key of modernity."¹⁵ His work is ontological and metaphysical, less anthropological and sociological-theoretical: Ellul's "direction of questioning can be understood as an ontologically probing phenomenology of the technical order or technical system, as a structural analysis of omnipresent technicity."¹⁶ Ellul sees the value of ex-post analyses, for example of work organization, alienation and media effects, as they help to understand individual aspects of "technology". But these studies must not be the starting point, rather: "It is necessary to start at the highest level of abstraction and then reach the reality constituted by the relationship between the Technique and man or Society."¹⁷

From this perspective, technology is *power*: technology is a powerful, determining factor. Technology is an "enforcement process [...] that takes possession of all areas of people's lives."¹⁸ For the technical order "the specifically human experiences of the body, language, interaction, togetherness and opposition, the involvement in space and time appear

Neumann: 1990; 109. Own translation. In German original: Technologie ist «ein distinktes Erzeugungsmuster der Wirklichkeit, dessen Folgen real, dessen Formationen wie auch Deformierungen augenfällig und handgreiflich sind.»

¹⁴ Cf. Ellul. *The Technological...* xxvi.

¹⁵ Ellul J. *Le système technicien*. Liberté de l'esprit. Paris: Calmann-Lévy: 1977; 7. Own translation. In French original: «clef de la modernité».

¹⁶ Langenegger. *Gesamtdeutungen moderner...* 110. Own translation. In German original: Elluls «Fragerichtung lässt sich als ontologisch sondierende Phänomenologie der technischen Ordnung, bzw. des technischen Systems verstehen, als eine Strukturanalyse allpräzenter Technizität.»

¹⁷ Ellul. *Le système...* 38. Own translation. In French original: «Il faut commencer au plus haut niveau d'abstraction pour ensuite rejoindre le réel constitué par la relation entre la Technique et l'homme ou la Société.»

¹⁸ Langenegger. *Gesamtdeutungen moderner...* 144. Own translation. In German original: Technik ist ein «Durchsetzungsvorgang [...], der sich aller Lebensbereiche der Menschen bemächtigt.»

as material that can be modelled.”¹⁹ With the implementation of the technical system, human action is finally also fixed and restricted to power: “Technique is a realisation, therefore an accomplishment, therefore an increase, of the spirit of power, which led to a polarisation of man on power.”²⁰

In later writings, Ellul applied his structural analysis of ubiquitous technicity to information and communication technologies (ICTs).²¹ Here the power and *justice relevance* of technical systems become clear, especially for digital technologies. It is to be feared that precisely these effects will be further enhanced by artificial intelligence technologies.

2) Romano Guardini, the great theologian and philosopher of religion, dares to say in 1950 that the modern age is over. He presents an analysis of the preconditions for the emergence of the modern age in the Middle Ages and describes the formations of the modern world view on the basis of three elements: The modern age is characterized by 1) naturalness or nature resting in itself, by 2) the self-perception of man as an autonomous person and 3) by culture, which produces itself from its own norms:²²

“By seeing the world as ‘nature’, man places it within itself; by understanding himself as ‘personality’, he makes himself the master of his own existence; in the will to ‘culture’ he undertakes to build up existence as his work.”²³

These elements of modern times, however, are lost to us, according to Guardini. The concept of nature is changing, the reverence for nature is being lost, and this has to do with the advent of technology:

¹⁹ Ibid.: 144f. Own translation. In German original: Für die technische Ordnung «erscheinen die spezifisch humanen Erfahrungen des Leibs, der Sprache, des Mit-, An- und Gegeneinanders, der Eingelassen-heit in Raum und Zeit als modellierbares Material.»

²⁰ Ellul. Le système... 80f. Own translation. In French original: «La technique est une réalisation, donc un accomplissement, donc en accroissement, de l'esprit de puissance, ce qui conduisait à une polarisation de l'homme sur la puissance.»

²¹ Cf. Marlin R. *Propaganda and the ethics of persuasion*, 2nd edn. Peterborough: Broadview Press: 2013; p. 312.

²² Guardini R. *Das Ende der Neuzeit: Ein Versuch zur Orientierung in Das Ende der Neuzeit / Die Macht*. Mainz: Matthias-Grünewald-Verlag: 1986; 9-94: 47.

²³ Ibid.: 40-41. Own translation. In German original: «Indem der Mensch die Welt als „Natur“ ansieht, stellt er sie in sich selbst; indem er sich als „Persönlichkeit“ versteht, macht er sich zum Herrn der eigenen Existenz; im Willen zur „Kultur“ unternimmt er es, das Dasein als sein Werk aufzubauen.»

"This can be seen in that epitome of knowledge and ideas of form, efficiency and procedure which we refer to by the word "technology". It slowly grew up in the course of the 19th century, but for a long time it was carried by a non-technical kind of man. It seems as if the human being assigned to it has only broken through in the last decades, finally in the last war. This human being again perceives nature as a valid norm, still a living salvage. [...] The new age loved to justify the measures of technology by their benefit for the welfare of mankind. In doing so, it covered up the devastation wrought by its unscrupulousness. The coming time, I believe, will speak differently. The person who wears it knows that technology is ultimately not about utility or welfare, but about domination; domination in the extreme sense of the word, expressed in a new world form."²⁴

According to Guardini, *technology is domination as a world form*. This also applies to the self-image of man as a person. Guardini sketches the idea of the "humane human being" that characterizes the modern age. He does not want to understand this humanity as a moral judgement, but rather as the correspondence of human activity with human experience. This specifically humane relation now falls apart at the end of the modern age: in the technical age human recognition, will and ability "exceeds the area of its immediate organization". Man plans effects "which he simply can no longer feel."²⁵ Thus the relationship to nature changes and, aimed at here, humanity changes:

"For man is what he experiences – but what is he if his actions can no longer be an experience for him in terms of content? After all, responsibility means standing up for what one does; the transition

²⁴ *Ibid.*: 50f. Own translation. In German original: «Das zeigt sich in jenem Inbegriff von Erkenntnissen und Formvorstellungen, Tüchtigkeiten und Verfahrensweisen, die wir mit dem Wort „Technik“ bezeichnen. Diese ist im Laufe des 19. Jahrhunderts langsam heraufgewachsen, war aber lange Zeit hindurch von einer nicht technischen Menschenart getragen. Es scheint, als ob der ihr zugeordnete Mensch erst in den letzten Jahrzehnten, endgültig im letzten Krieg, durchgebrochen sei. Dieser Mensch empfindet die Natur wieder als gültige Norm, noch als lebendige Bergung. [...] Die neue Zeit liebte es, die Maßnahmen der Technik mit ihrem Nutzen für die Wohlfahrt des Menschen zu begründen. Damit deckte sie die Verwüstungen zu, welche ihre Skrupellosigkeit anrichtete. Die kommende Zeit wird, glaube ich, anders reden. Der Mensch, der sie trägt, weiß, dass es in der Technik letztlich weder um Nutzen noch um Wohlfahrt geht, sondern um Herrschaft; um eine Herrschaft im äußersten Sinne des Wortes, sich ausdrückend in einer neuen Weltgestalt.»

²⁵ *Ibid.*: 60. Own translation. In German original: «Das Feld des Erkennens, Wollens und Wirkens des Menschen überschreitet [...] den Bereich seiner unmittelbaren Organisation. [...]. Er vermag Wirkungen zu planen und durchzuführen, die er einfach hin nicht mehr durchfühlen kann [...].»

of the respective event into ethical appropriation – but what is it, if the process no longer has a concrete form, but runs in formulas and apparatuses?"²⁶

These changes in the relationship to nature and the self-perception of the human being as a person have an impact on what Guardini calls culture. For him, the basis of cultural creation is the power over being and in it still an element of modern times. But with the end of the modern age, power is equated with "progress". Power is something ambivalent, it depends on how it is used. Guardini sees in the technical age a rapid increase of power, but a stagnation in dealing with power. There is no adequate growth of an ethos of the use of power, which leads to the view that the use of technical possibilities "must be seen as a natural process for which there are no norms of freedom, but only supposed necessities of use and security."²⁷ This results in independence of post-modern power, ultimately a consolidation of technology as a form of rule as a world form:

"More: the development gives the impression that power is becoming more objective; as if it is basically no longer held and used by man at all, but is developing independently from the logic of scientific questions, technical problems, political tensions, and determines its own actions."²⁸

3) A final impulse focuses on the interrelation between science and technology. In the Handbook for Christian Ethics Zimmerli and Wolf show the consequences of the "technological age":

"Technologization in the narrower sense means that (a) basic scientific research, technical application and economic use can no longer be clearly separated, but merge into a new type of scientific activity,

²⁶ *Ibid.*: 61. Own translation. In German original: «Denn der Mensch ist doch, was er erlebt – was ist er aber, wenn sein Tun ihm inhaltlich nicht mehr zum Erlebnis werden kann? Verantwortung bedeutet doch das Einstehen für das, was man tut; den Übergang des jeweiligen Sachgeschehens in die ethische Aneignung – was ist sie aber, wenn der Vorgang keine konkrete Gestalt mehr hat, sondern in Formeln und Apparaturen verläuft?»

²⁷ *Ibid.*: 70. Own translation. In German original: «Da es ein wirkliches und wirk-sames Ethos des Machtgebrauchs noch nicht gibt, wird die Neigung immer größer, diesen Gebrauch als einen Naturvorgang anzusehen, für welchen keine Freiheitsnormen, sondern nur angebliche Notwendigkeiten des Nutzens und der Sicherheit bestehen.»

²⁸ *Ibid.*: 71. Own translation. In German original: «Mehr: die Entwicklung macht den Eindruck, als ob die Macht sich objektiviere; als ob sie im Grunde überhaupt nicht mehr vom Menschen innegehabt und gebraucht werde, sondern sich selbstständig aus der Logik der wissenschaftlichen Fragestellungen, der technischen Probleme, der politischen Span-nungen weiterentfalte und zur Aktion bestimme.»

and that (b) not only science is becoming more and more technical, but also technology is becoming more and more scientific, i.e. technological.”²⁹

In the course of this technologization, however, technological civilization itself becomes a problem and turns against the goals of modernity. The empirical sciences lose their innocence, so to speak. Curiosity and the search for truth as driving forces of the sciences are becoming more and more implausible. The status of the findings of the empirical sciences is also changing: experiments are taking on a technological character through computer simulations and the computers are “gaining a theory-generating status within certain questions of the empirical sciences (e.g. in economics).”³⁰

A new type of technological knowledge is emerging. This has high ethical relevance: “If it is true that scientific research and technological action converge, then it is also true that the gain in knowledge must already be morally responsible.”³¹ One can clearly see how much the curiosity and freedom of research have lost their innocence in the technological-scientific age, although these modern ideas continue to be defended and the problems are thereby (intentionally?) overlooked.

These three outlined perspectives on technical ethics illuminate only a part of the many debates of the last 80 years or so, but they seem appropriate for a Christian social ethics that puts social questions, problems of justice and power at the centre of its interest in knowledge. In the next step, we sharpen the cognitive interest of Christian social ethics with the help of the heuristics of the signs of the times.

²⁹ Zimmerli W. C., Wolf S. *Die Bedeutung der empirischen Wissenschaften und der Technologie für die Ethik* in *Handbuch der christlichen Ethik*. Bd. 1. Hertz A., Korff W., Rendtorff T. Ringeling H. (eds.). Freiburg i. Br.: Herder: 1993; 297-316: 299. Own translation. In German original: «Unter ‚Technologisierung‘ im engeren Sinn ist zu verstehen, daß (a) wissenschaftliche Grundlagenforschung, technische Anwendung und wirtschaftliche Nutzung nicht mehr scharf zu trennen sind, sondern zu einem neuen Typus wissenschaftlichen Handelns verschmelzen und daß (b) nicht nur die Wissenschaft immer technischer, sondern auch die Technik immer wissenschaftlicher, eben technologischer, wird.»

³⁰ *Ibid.* Own translation. In German original: Computern «wächst ein theoriengenerierender Status innerhalb bestimmter Fragestellungen der empirischen Wissenschaften (z. B. in der Wirtschafts-Wissenschaft) zu.»

³¹ *Ibid.* Own translation. In German original: «Wenn gilt, daß die wissenschaftliche Forschung und technologisches Handeln konvergieren, dann gilt auch, daß schon der Erkenntnisgewinn moralisch verantwortet werden muß.»

Technology and society: Christian social-ethical perspective (signs of the times)

Christian social ethics as science sees itself as theological ethics. As such, it is a necessary part of the theological endeavour. At its core it is about the connection between faith and responsibility for the world, as it was outlined in the Pastoral Constitution *Gaudium et Spes* of the Second Vatican Council: "To carry out such a task, the Church has always had the duty of scrutinizing the signs of the times and of interpreting them in the light of the Gospel."³² The formula of "signs of the times" used in this document is essential for Christian social ethics. The challenges reveal themselves as such only in turn towards the world.

According to Marie-Dominique Chenu, signs of the times are "general phenomena that encompass a wealth of events and express the needs and expectations of contemporary humanity. But these general phenomena [...] are only 'signs' in that they bring a leap or even a break in the continuity of the human sense of history."³³ The epochal marks a rupture in which, as Karl Lehmann put it, a "form of life emerges in a new way"³⁴ and which challenges a Christian-ethical perspective in a special way. "The task of recognizing the signs of the times thus refers to a work of interpretation to be carried out from the standpoint of Christian social ethics: to perceive reality in such a way that the concrete needs, fears and insecurities of people are got to the bottom of it."³⁵

³² *Gaudium et spes*, No. 4.

³³ Chenu M.-D. *Les signes des temps* in *Nouvelle Revue Théologique*. 1965; 87: 29-39: 33. Own translation. In French original: «Ainsi sont "signes des temps" des phénomènes généralisés, enveloppant toute une sphère d'activités, et exprimant les besoins et les aspirations de l'humanité présente. Mais ces phénomènes généraux ne sont "signes" que sous la commotion d'une prise de conscience, dans le mouvement de l'histoire.»

³⁴ Lehmann K. *Neue Zeichen der Zeit: Unterscheidungskriterien zur Diagnose der Situation der Kirche in der Gesellschaft und zum kirchlichen Handeln heute*. Eröffnungsreferat bei der Herbst-Vollversammlung der Deutschen Bischofskonferenz 2005 in Fulda. Der Vorsitzende der Deutschen Bischofskonferenz, vol. 26. Bonn: 2006; Sekretariat der Deutschen Bischofskonferenz: 45. Own translation. In German original: «Manches kann auch als ein „Zeichen der Zeit“ erscheinen, das einfach neu bedacht werden muss: Eine Gestaltetes Lebens entpuppt sich auf neue Weise.»

³⁵ Heimbach-Steins M. *Sozialethik in Orientierung finden: Ethik der Lebensbereiche*. Arntz K., Heimbach-Steins M., Reiter J., Schlägel H. (eds.). Theologische Module, vol. 5. Freiburg im Breisgau: Herder Freiburg: 2008; 166-208: 178. Own translation. In German original: «Die Aufgabe, die Zeichen der Zeit zu erkennen, verweist also auf eine vom Standpunkt der christlichen Sozialethik aus zu leistende Deutungsarbeit: die Wirklichkeit so wahrzunehmen, dass den konkreten Nöten, Ängsten, Verunsicherungen der Menschen auf den Grund gegangen wird.»

This attitude of Christian social ethics based on faith is at the same time marked by the turn towards the world and by the critical perspective on the present, so it cannot be classified in the scheme of optimism/pessimism. The technical perspectives described by Ellul, Guardini and others may be depressing because of their analysis of a comprehensive penetration of the world with the technical paradigm. A hopeless and end-time mood, however, is not guiding either Ellul or Guardini, but I recognize above all an enlightening impulse in their critical diagnoses. Therefore it is important to emphasize that technology criticism can also appear as an ideology or can be integrated ideologically.³⁶

Interim social-ethical result: Techniques and technologies of artificial intelligence as a sign of the times

The techno-ethical perspective of this chapter has decoded technicity as a special form of domination: *technology is domination as a world form* (Guardini), technology is, following Ellul, a specific, generally asserting *pattern of reality production*. The interactions between science and technology give rise to a precarious modernity in which a technological type of knowledge emerges that is no longer securely supported by the scientific ethos of curiosity and the quest for truth.

Undoubtedly, the latest technological successes, especially in synthetic biology and artificial intelligence, are signs of the times in so far as they can be interpreted as intensifications and dynamizations of technology as domination and as specific patterns of production of reality. Guardini's categories of nature, man and culture as elements of the modern age are subject to increased change in the present through these technologies. Although this presented technical-ethical analysis is not complete and shows only an excerpt from the debate, it has become clear that the social-ethical problems of modern successes in the field of artificial intelligence should be found and dealt with at the level of their technicity.

Artificial Intelligence - Impulses of Christian Social Ethics

Artificial intelligence as a topic of Christian social ethics

In the field of Artificial Intelligence we are dealing with innovations in computer science and computer technology; the context is therefore

³⁶ See for example Lenk H. *Technokratie als Ideologie: Sozialphilosophische Beiträge zu einem politischen Dilemma*. Stuttgart: Kohlhammer; 1973.

"digitization". By digital technologies, we mean today computer systems based on algorithms, data and artificial intelligence. Particularly in the area of machine learning (AI), great progress has been made in recent years. This is due to a) the availability of data (through digital communication and powerful sensor technology everywhere, for example in mobile phones and in medicine), b) the greatly increased storage capacity, c) recent research successes in the field of machine learning (deep learning), especially in mathematics by Yan LeChun and Geoffrey Hinton, and d) a multiplication of computing power. Especially "autonomous systems" on this technical basis (robots, vehicles, cancer detection systems, military drones ...) are in the focus of a social debate.

With good reasons, the philosophy responds to the challenges of current innovations from different perspectives. Philosophy of mind, anthropology, philosophy of nature, general or metaethics and applied ethics are the typical philosophical disciplines in this field. However, it is unclear to what extent these approaches are connected or interrelated. On closer inspection, the philosophical challenges of Artificial Intelligence are quite heterogeneous, depending on the philosophical discipline from which one approaches the matter. Time also plays a certain role: Since about the 1950s, the philosophy of mind has often found one of its touchstones in the field of artificial intelligence. Theoretical considerations of computer science were accompanied by theoretical considerations of metaphysics, for example. Applied ethics is only now coming into play, because many existing data (increased storage capacity, ubiquitous sensor technology) and computing power mean that AI systems are ready for use and pushing into the markets.

Christian social ethics as social ethics is oriented towards social phenomena, towards structures of human coexistence and towards organizational and institutional conditions. As ethics within the horizon of faith and as theological ethics it is characterized by a special sensitivity for questions of power and justice. The considerations so far have shown that social ethics should not start with individual technical objects but should analyze artificial intelligence in its technical character.

Theological-ethical basic impulses

Christian social ethics as theological ethics is connected with basic concepts of Christian theology. Its ethics, which as such wants to be generally agreeable in the space of reason, acquires a specific character through the processing of theological motives. If one considers how

artificial intelligence *as technology* can be reflected theologically and ethically, I think that three theological motives in particular come into question: creation (and with it also freedom), anthropology and eschatology.

1) *Creation and freedom:* Creation means theologically: "All reality ultimately comes from the hand of God and is held by him."³⁷ For Sacred Scripture does not focus on a divine quality of the cosmos, but on man as the centre of the world: he is the image of God and partner in his covenant. The created world is entrusted to man as environment, he has to shape it. Therefore, the term creation does not mean the divinity of being or gives an indication of the physical origin of the world, "but points to the significance of the world as the scene of God's history with man."³⁸ Not only the creation but more the creative preservation of the world is in the focus of the Bible. Theologically this is understood as "*creatio continua*": In the great history as well as in the everyday actions of man God works and remains the "origin of all events."³⁹ This insight leads to an understanding of creation, in which God creates this world together with man: "By working on the creation of a peaceful world, man wins himself, but also gains a corresponding image of the author, of the Creator God."⁴⁰ Essential for this understanding of creation is that creative preservation of the world by God and human freedom belong together. Man does not have this freedom, it only takes place by acting. Freedom is creation. In his free action in and on the world, man proves himself as a person, and he does this as God's partner in the covenant. Freedom out of the idea of creation means the "ability to accept one's own being in the free execution and to see in the fulfilment of God's demand not an opposition to one's own freedom, but its true facilitation."⁴¹

³⁷ Gründel J. *Die Kategorie der Schöpfung* in *Handbuch der christlichen...*; 407-421: 407. Own translation. In German original: «Alle Wirklichkeit geht letztlich aus der Hand Gottes hervor und wird von ihm gehalten.»

³⁸ *Ibid.*: 408. Own translation. In German original: Der Begriff Schöpfung «weist auf die Bedeutung der Welt als Schauplatz der Geschichte Gottes mit dem Menschen hin».

³⁹ *Ibid.*: 415. Own translation. In German original: «Ursprungsgrund allen Geschehens».

⁴⁰ *Ibid.* Own translation. In German original: «Indem der Mensch an der Gestaltung einer friedfertigen Welt arbeitet, gewinnt er sich selbst, gewinnt aber auch ein entsprechendes Bild vom Urheber, vom Schöpfergott.»

⁴¹ *Ibid.*: 417. Own translation. In German original: «Fähigkeit, in dem freien Vollzug das eigene Wesen anzunehmen und in der Erfüllung der Forderung Gottes nicht einen Gegensatz zur eigenen Freiheit zu sehen, sondern deren wahre Ermöglichung.»

2) *Anthropology*: In the context of technology and artificial intelligence, there is always talk of the fact that “man must be at the centre”. Which person is meant by this and what is good for him and does him justice remains open. This is not necessarily a bad thing, because with the category of the human being, despite or precisely because of an ambiguity of content, a moral perspective can nevertheless be made strong. A theological anthropology does also not provide a rigid definition of the conception of man, but it emphasizes the dignity of man as a free person:

“If today we speak of man as a person, we mean both his endowment with the spirit, which goes beyond materiality and physicality, and the independence and freedom that is founded in it. The human being as a person is not merely part of a larger whole, but retains its autonomy, its value in itself and must not be seen merely in terms of function or purpose”.⁴²

Artificial intelligence and digital information technology in the broad sense of the term represent an enormous challenge to the question of human beings and their identity.⁴³ Man himself is becoming a bundle of information, human rationality must be comparable with the precision of machine “decisions” in more and more areas, man-machine interactions are increasing more and more and place man next to the machine.

3) *Eschatology*: Eschatology is about the completion of the individual as well as the completion of the whole creation. Whether we find in the theology of hope (Moltmann and Pannenberg) or in political theology (Metz) an orientation towards this world or whether we are classically oriented towards the hereafter does not play a decisive role. What is challenging is that completion narratives today come from the field of technology. It is about an immanentization and trivialization of the hope of completion, for instance by mind-uploading, cognitive enhancement or the transfer of world destiny to a super intelligence. A theological eschatology, on the other hand, refers to the divine promise, the Easter

⁴² *Ibid.*: 414. Own translation. In German original: «Wird heute vom Menschen als Person gesprochen, so ist damit sowohl seine über die Materialität und Körperlichkeit hinausgehende Begabung mit Geist wie auch die darin gründende Eigenständigkeit und Freiheit gemeint. Der Mensch als Person ist nicht bloß Teil eines größeren Ganzen, sondern behält seine Eigenständigkeit, seinen Wert an sich und darf nicht nur funktional oder unter dem Gesichtspunkt einer Verzweckung gesehen werden.»

⁴³ Cf. Benanti P. *Artificial Intelligences, Robots, Bio-engineering and Cyborgs: New Challenges for Theology?* in *Concilium. International Journal for Theology*, 2019; 55: 267-279.

event and the Kingdom of God and thus salutarily relativizes the technical eschatologies.

These short sketches indicate how theologically and ethically fruitful it is to encounter the present technologized world. With the creation motive one can appreciate the free work of man and does not have to reject his technical achievements *per se*. Anthropology emphasizes the purpose of man and keeps the question of what is good for man and does justice to him up-to-date and emphasizes its importance. Eschatology, finally, sees completion in connection with God's promise and is sceptical of technical completion fantasies of man or the world.

Social-ethical perspective: justice of persons

This theological background results in a social-ethical perspective in the concept of *justice to people as persons* (in german: Persongerechtigkeit); "The social conditions are to be set up and formed in such a way that the personhood of man is taken into account, his personal self-development is made possible and demanded."⁴⁴ The Christian social-ethical principle of the person raises the "claim of justice to people as persons to the social institutions."⁴⁵

Pope Francis also gives us impulses for the question of just social structures in this world of ours, which is characterized by its technical character. In Laudato Si', Pope Francis names the criticism of the forms of power "derived from technology"⁴⁶ as a central theme that runs through the whole Encyclical. In the numbers 101 to 114 of his encyclical he becomes concrete. First he emphasizes the value of technical progress (102f.), but then in No. 104 he deals with the "tremendous power" which for instance also computer science gives us. Francis then goes on to focus on his central theme, the forms of power: "More precisely, they have given those with the knowledge, and especially the economic resources to use them, an impressive dominance over the whole of humanity and the entire world." The power that information technology, for example, gives us is unequally distributed.

⁴⁴ Heimbach-Steins, *op. cit.*, note 35, p. 183. Own translation. In German original: «Die gesellschaftlichen Verhältnisse sind [...] so einzurichten und auszustalten, dass dem Personsein des Menschen Rechnung getragen, seine personale Selbstentfaltung ermöglicht und gefordert wird.»

⁴⁵ *Ibid.*: 187. Own translation. In German original: «Anspruch der Persongerechtigkeit an die gesellschaftlichen Institutionen».

⁴⁶ Franciscus PP., *op. cit.*, note 8, p. 854. (Nr. 16).

In paragraph 105, Francis then quotes from Romano Guardini those passages that were already discussed above. He takes up from Guardini above all the thesis that the growing technical power of man has no controlling possibilities: "In this sense, we stand naked and exposed in the face of our ever-increasing power, lacking the wherewithal to control it. We have certain superficial mechanisms, but we cannot claim to have a sound ethics, a culture and spirituality genuinely capable of setting limits and teaching clear-minded self-restraint."

In the following, Francis continues his analysis and deepens it in relation to the theses that we have already highlighted above in terms of technical ethics. In doing so, he understands the ecological crisis as an element of a general enforcement of the technological paradigm: "We have to accept that technological products are not neutral, for they create a framework which ends up conditioning lifestyles and shaping social possibilities along the lines dictated by the interests of certain powerful groups. Decisions which may seem purely instrumental are in reality decisions about the kind of society we want to build." (107)

Here Francis again quotes Guardini when he characterizes the hard logic of technology as a comprehensive form of domination (108). This ultimately does not do justice to the human being as a person: "Our capacity to make decisions, a more genuine freedom and the space for each one's alternative creativity are diminished." He nevertheless recognizes cautious "fog" of free human creativity, "authentic humanity" (112), which could overcome this logic of technology. In our social-ethical terminology, it is then about making the world as a technical world more just to people as persons.

Social-ethical problem dimensions and solution perspectives

The study "*Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*"⁴⁷ is in line with the social-ethical perspective. Especially the focus on social implications helps to bring together the basic questions of the technical character of our reality with concrete challenges of artificial intelligence technologies. It is true that the study does not clearly distinguish between algorithms, data and artificial intelligence, so that the authors speak of ADA technologies (ADA = Algorithms, Data, Artificial Intelligence).

⁴⁷ J. Whittlestone et Al. 2019. *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. London: Nuffield Foundation.

Therefore, the following problem areas arise for Christian social ethics when dealing with artificial intelligence:

- Impact of ADA on the economy and economic growth.
- Impact of ADA on jobs and labour markets, developing policies around technological unemployment.
- Impact of ADA on global inequality.
- How ADA changes power in a society.
- Impact of ADA on international relations, conflict, and security – including impact of autonomous weapons and risk of a global arms race.
- What new methods of global governance might be needed to deal with the challenges posed by increasingly powerful technologies

Because of its fundamental nature and scope, this list is a description of a comprehensive research programme for Christian social ethics. It is important that problems of power and justice must be addressed thematically in these fields. From the extensive catalogue of social-ethical perspectives on the topic of artificial intelligence, I would like to briefly outline four questions in more detail:

1) The international political dimension: A global competition for the leading position, especially in the field of "artificial intelligence", has emerged that is reminiscent of an arms race. The libertarian, venture capital-driven American model is contrasted with a state-totalitarian top-down system with mass surveillance. Europe is trying to develop a specifically value-oriented dynamic. For this reason, international relations (geopolitical, economic, etc.) must be given more intensive social and ethical consideration. Cooperations and agreements would have to be demanded in a social-ethical way with a view to the global common good. An important problem in this context is that the data hunger of an "artificial intelligence policy" sees people as data suppliers: Both the more Anglo-Saxon version of surveillance capitalism⁴⁸ and, for example, the social credit system in China are based on a comprehensive data validation of human behaviour so that artificial intelligence systems can function better and better. While it is true that, for example in the health sector, the sharing of personal data is relevant to the common

⁴⁸ Cf. S. Zuboff. 2019. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.

good and thus also to justice, this tendency must not be at the expense of informational self-determination.

2) Future of democracy: Social media as a business model lead to strong political and cultural distortions, at least in Western democracies. The significance of a discursive public sphere for democracy as a political form of social self-determination is increasingly under debate. Digital communication technologies are ideally suited for manipulation, surveillance and populism. Information supply, debate and opinion-forming are difficult to ensure. Strong media organisations remain central. But democratic institutions, political processes, and the enforcement of European and national law are also current social-ethical hotspots.

3) New knowledge and insights: In addition to the technical control of knowledge flows through artificial intelligence, this technology also gives us the power to generate new knowledge and new forms of knowledge. Benanti points out in a nice analogy that the invention of the microscope and the telescope has created completely new dimensions of knowledge.⁴⁹ The technologies of artificial intelligence are now the invention of a "macroscope". This enables us to create equally precise and granular knowledge on a huge database that was previously inaccessible to us. And clearly, this knowledge is also about power. Examples are the prediction of voting decisions, the diagnosis of health conditions on the basis of eye images or the analysis of sexual orientation on the basis of portraits.

4) The future of work: Most studies show that in total, no more jobs will be lost than will be created. Nevertheless, the transformation of the world of work through automation will be massive. This change will cause social problems. The dynamics are immense; however, I cannot imagine an alternative to strong investments in automation and digital technologies for the economies. These must therefore be accompanied by investment in education and training, climate protection, adaptation of social security systems and adaptation of tax systems.

In all these fields, the status of the human person is at stake. It is high time that Christian social ethics and also the Church's social proclamation deal more intensively with the technicity of our world.

⁴⁹ Benanti, *op. cit.*, note 43.

Actors and responsibilities: Multilevel governance for artificial intelligence

It was already pointed out at the beginning that from a Christian social-ethical perspective the shaping of the world, which functions on the logic of the technical, must be done in orientation to the human person. *The social-ethical perspective understands Artificial Intelligence as the current form of the technical imperative.* Thus, Artificial Intelligence is a concretion of the comprehensive changes of those external technical and social conditions that shape the human being, that are the preconditions for the questions about himself, that influence answers to the questions about the good, that order his knowledge and shape coexistence. These external conditions must do justice to man as a person.

There is no way around an ethics of responsibility for achieving this goal, although actors and their responsibilities are increasingly difficult to define. Intelligent systems are already permeating our everyday life and our lives. They open up opportunities to improve our lives, for example by reducing road deaths through self-steering vehicles or improving communication skills. They also harbour dangers such as the loss of our privacy or the deliberate manipulation of people and society. The current issue is, therefore, the question of what standards we should use as a basis for designing artificial intelligence.

To start from the concept of responsibility here means to take up an integrative approach that has proven itself in the field of ethical theory formation. Legally and ethically, responsibility is associated with individual, corporate and social subjects of responsibility (actors): Who is responsible for what, to whom, before which authority, why and in which time perspective? The increasing complexity of the technical world, however, makes this assignment extremely challenging. When the processes behind the human-(system)-machine interface become more and more complex and technologies increasingly make "autonomous" decisions, relations of responsibility are difficult to determine. Platform operators on the Internet like to reject responsibility for content. Although something is slowly happening here in the context of fake news and hate speech, controlling these systems is extremely difficult, especially in an international context.

A concept that reacts to this situation is currently being discussed under the label "Corporate Digital Responsibility".⁵⁰ However, where

⁵⁰ Cf. Koska C., Filipović A. *Blackbox AI - State Regulation or Corporate Responsibility?* In *Digitale Welt* 2019; 3: 28-31.

this responsibility is not recognised and perceived by companies, where ethics only act like a washing machine, this is where politics must take action. A courageous regulation of technical systems beyond a mere consideration of economic factors is and remains central. However, it is not always clear whether and how politics will assume its regulatory primacy. Here, too, the admonishing voice of church and science is needed.

Flowing of Life and Static of Machine: A Daoist Perspective on AI

Robin R. Wang *

This essay brings ancient Daoist philosophy into a conversation to address the challenges proposed by AI technology. Today, all cutting-edge AI technologies exist not just in research labs but already easily penetrate into all aspects of human lives. It is difficult to argue that we do not yet inhabit a world with AI, which has become a pervasive and effective technology woven into the fabric of everyday living. AI challenges what it means for humans to be humans; what our moral society means, and how our societal values are shifting? Our values, society, and laws are centered around humans, and under the current revolution of AI technology, we must ask what impacts it would have on each of these aspects? This essay will discuss these issues from two Daoist philosophical aspects.

Life as Qi flow and Beyond: Daoist view on the Nature of Human Being and Beyond

In order to comprehend the Daoist view of a human being we need to begin with the Chinese notion of *qi* 氣 (vital energy). This term is among the most important, cherished, and widely applied concepts in Chinese intellectual history. As a shared notion underlying all schools of ancient Chinese thought, *qi* is believed to be a dynamic, all-present, all-penetrating, and all-transforming force that animates every existence in the universe. Although *qi* is an abstract idea, it also is a common and integral part of our perception, experience, and existence. It is woven into language we speak, the air we breathes, the very force that drives the fusion of our blood, the food we eat, the strength of our mind, the flow of our thoughts, and the deepest urges of our hearts. *Qi* is the very fabric and force of life. As the Daoist classic text *Zhuangzi* puts it, "Human life is all about generating *qi*. When *qi* is gathered there will

* Full Professor of Philosophy, Loyola Marymount University, Los Angeles (USA).

be life; when *qi* is dispersed there will be death.¹ When *qi* declines, one will become sick; when *qi* is lost, one will die.

On further analysis, *qi* is a complex of different energies, each animating and controlling various aspects of human life and the human body. We read from other early Chinese text *Huainanzi*:

Human beings can see clearly and hear acutely; they are able to protect their own body and bend and stretch their one hundred joints. In their discrimination they are capable of distinguishing white from black, beautiful from ugly. In their intelligence they are capable of distinguishing similarity from difference and clarifying right from wrong. How can human beings do so? This is because the *qi* infuses these activities and the spirit (*she* 神) regulates them.²

This primacy of *qi* lies in its self-generating and self-operating power. *Qi* is the *Dao* in its sense of the origin of the myriad things and the basic material of universe. *Dao* is materialized in *qi* and, thus, in space and time. In fact, in parts of the *Guanzi*, the *Zhuangzi*, and many Neo-Confucian texts, *Dao* and *qi* are practically interchangeable. The pulse and rhythm of *qi* give rise to all things. As a force embedded in nature, *qi* guides, shapes, and directs natural processes from within.

The classical Chinese medical text *Huangdi Neijing* (*Yellow Emperor's Internal Classics*) attributes the particular state of physical and mental entities to *qi*. *Qi* is stored in the five *zang* 腸 (organs, storehouses) of humans. The quality and quantity of *qi* in these organs as well as their transformations—the way the *qi* moves and interacts with other organs as well as the outside world—determines one's physical and mental state. *Qi* is causally responsible for all of our mental and physical states.

After all, life is a form of *qi* flow. Daoist practitioners throughout Chinese history are like *qi* engineers, capable of taking a variety of *qi* flows into a directed system and configuring the 12 *qi* flow channels, namely *jingluo* 经络 in the human body.³ In Chinese medical practice, acupuncture needles, hand-message and natural herbs can increase one's

¹ Ziporyn B. (trans.). *Zhuangzi: The Essential Writings*. Indianapolis: Hackett Publishing Company, 2009, p. 86.

² Ames R. (trans.). *Yuan Dao, Original Dao: Trace to Its Root*. New York: Ballantine Books, 1998, p. 26; Roth H. (trans.). *The Huainanzi, A Guide to The Theory and Practice of Government in Early China*. New York: Columbia University Press, 2010, p. 75. This is a modified translation from both and Ames and Roth.

³ Ziporyn. *Zhuangzi...* p. 22.

qi flow while others will decrease or block it. *Qi* can be measured and quantified through contemporary technological devices today.

Arguably, computational algorithms can be seen as exhibiting a form of *qi*-flow. Consider a computer algorithm, an organizing and arranging of data, which is a means of turning inputs into outputs.⁴ Recently, AI has shown the ability to defeat humans at chess or Go because, it was programmed how to play the game with access to huge amounts of computing power enabling it to make billions of calculations about the best possible move in a game. AI can analyze the consequences of certain move, to remember the outcome of this move in past games, and determine if this move can and should beat a human opponent.

More interestingly, cognitive psychology images the human brain as a machine, from which complex behaviors arise or as the aggregate result of multiple simple responses. Similarly, Daoists conceptualize the human brain itself as a *qi*-flow network, from which complex behaviors arise as the aggregate result of multiple *qi* responses.⁵ The action is an emergent behavior between humans or machines and the environment. This principle is seen in Daoist teachings as well as in the cybernetics movement. However we may aspire to ask, in what ways do we know that biology does not conform to the algorithmic view of organism behavior? Certainly the human body has algorithmic parts, but to call the whole human being an algorithm is far too reductionist. More specifically, Daoist philosophy sees human life as something more than simply information, data and network. Human life is regarded as a complex, nonlinear, dynamic, self-organizing system, exchanging energy with information on multiple levels of organization in order to survive and thrive.

Let's consider a story from the early Daoist text *Liezi*.⁶ Jiliang was sick but he refused to undergo any medical treatment, and after seven days his situation became serious. His seven sons stood in a circle and begged him in tears to seek medical attention. In order to teach his sons a lesson about life, he agreed to call in three doctors, Qiao, Yu, and Lu, to take his pulse and make a diagnosis. Doctor Qiao explained that Jiliang's hot and cold temperatures, the invisible and visible forces in

⁴ Dormehl L. *Thinking Machines: The Quest for Artificial Intelligence and Where It's Taking Us Next*. Tarcherperigee, 2017; p. 12.

⁵ Ziporyn. *Zhuangzi*... p. 86.

⁶ *Liezi jishi* 列子集釋, Beijing: Zhonghua Shuju Press, 1979; p. 204. My own translation.

his body, were out of order. According to him the illness was the result of improper diet, sexual indulgence, and lifestyle stressors. However it could be cured. Jiliang responded, "This is a *zhongyi* 罷醫 (common doctor), get rid of him now."

Next, doctor Yu offered his diagnosis and interpretation: "The current condition started even in your mother's womb. Your mother suffered a deficiency of embryonic *qi* and an excess of breast milk. This illness was not a matter of one day or one night. It has gradually been developing." Jilang responded, "This is a *liangyi* 良醫 (good doctor), serve him a dinner." Lastly, doctor Lu offered his diagnosis: "The illness is not from heaven, not from a human, and not from a ghost. Your life was generated and endowed with a form 粟身授形 (*bing shen shou xing*). However, it also came with a 制者 (*zhizhe*) governor. You should know it. What can all medicine do for you?" Jilang responded: "This is a *shenyi* 神醫 (spiritual doctor), give him a great gift."⁷

How can we make sense of this metaphorical story? The first *zhongyi*'s diagnosis was an accurate description of common human life in which: dietary indiscretion and life style choices produced a set of syndrome. The second *liangyi*'s diagnosis sees the interdependencies in human life, examining human life in a genetic context. In fact, even today, we are told that some distress may be the result of a DNA defect in the human genome. Advanced AI can effectively perform the tasks of the *zhongyi* and *liangyi*, exhibiting of conventional medical practices and hereditary theory.

The distinct from the first two, which operate at the level of human bodies, the *shenyi*'s diagnosis points to a "governor" of human life. Lu indicates that there is a *shen* 神 (spirit, force, power), a ruling and managing entity in one's life. If one can cultivate this *shen* the illness will be cured without any medications. A commentary claims that "the stupid ones will be perplexed when they hear it, but the intelligent ones will be enlightened when they learn it."⁸ The cultivation of this *shen* within the human body through persistent efforts and practices later canonized in the Daoist tradition eventually led to the formation of *neidan* 內丹 (inner alchemy).

This classical Daoist vision of human being was pursued and actualized in later Daoist texts, rituals, and practices. One of the most important ways to understand and analyse the human body was through the

⁷ *Ibid.* 205.

⁸ *Ibid.* 205.

distinction of three elements: physical form (*xing* 形), *qi* (vital energy), and spirit (*shen* 神). *Xing* refers to shape or form – the physical, visible form of the body. It is the house or abode of life and a vessel for the Dao.⁹ *Qi*, as mentioned before, is the invisible foundation and the source of life. Spirit (*shen*), just described, is the psychological and spiritual aspects of human life that regulate it.¹⁰ If *qi* is gathered, the form possesses life; if *qi* is lost, the form loses life. If spirit has purified *qi*, it will be in order (*zhǐ*治), however, if *qi* is turbid, the spirit will be chaotic. The crucial aspect of this understanding is that *qi* is the mediator between spirit and bodily form. It is through *qi* that mind and body are united and interact.

Other early Daoist text, *Taiping Jing* (*Classics of Great Peace*) illuminates this point: "Spirit is embodied from heaven, refined essence (*jīng*) is endowed by earth; *qi* comes from balance and harmony (*zhōng hé*). Spirit rides *qi* to move, and refined essence (*jīng*) inhabits balance. The three of them assist one another. This will lead to longevity. This is caring for *qi*, respecting spirit and valuing refined essence."¹¹ A. C. Graham translates the word *shen* into a wide range of meanings, such as spirit, daemon/daemonic, numinous, and the locus of more prosaic aspects of awareness.¹² *Shen* can mean spirit, the divine, the obscure and immeasurable, or that which happens as if by magic. *Shen* was frequently associated with other Chinese character "*ming* 明" which encompasses meanings from brightness and illumination to insight. The character itself groups images of the sun (*ri* 日) and of the moon (*yue* 月). Both *shen* and *ming* can be either an attribute of the world or of human beings. They, thus, connect human beings with the cosmos, and together they constitute the spiritual and intellectual core of a human being. *Shenming* literally refers to a kind of intelligence possessed by the spirits, however, it is also attributed to sages.¹³ It is a "spiritual-like intuitive clarity" attainable by human beings. Harold Roth takes the phrase *shenming* 神明 as marvellous influence, magical efficacy, or spiritual illumination. It is a way to explain the world beyond the narrow

⁹ Kohn L. (ed.). *Daoism Handbook*. Leiden: Brill, 2000; p. 96.

¹⁰ Roth. *The Huainanzi...* p. 74.

¹¹ Yang Jilin 杨寄林 (ed.). *Taipingjing* 太平經 *Classic of Great Peace*. Shijiazhuang: Hebei People's Press, 2002; p. 1730.

¹² Graham C. (trans.). *Chuang-tzu, The Inner Chapters*. Indianapolis: Hackett Publishing Company, INC. 2001, p. 58.

¹³ According to Kenneth E. Brashier, "In pre-Han and Han texts, 'spirit illumination' [*shenming*] could refer to either a divine being or a spirit-like intelligence ..." (Brashier, K. E. *Han Thanatology and the Division of Souls in Early China*, 21, 1996; p. 149).

vision of human beings, pointing toward mystical experience, ineffability, noetic equality, transiency, and passivity.¹⁴ *Shenming* is attained only through a process of cultivation and the elimination of bias.

This *shenming* is also closely connected with *qi*. As a verb, *shenming* communicates two opposite qualities in the transformation *qi*: condensing and extending. Condensing *qi* begins with *shen*, and extending *qi* becomes *ming*. *Shen* frequently means that the *qi* is condensing or absorbing the nature of the earth; *ming* denotes that the *qi* is extending or issuing the nature of heaven. Thus, *shen* is the non-manifest, inscrutable aspect of *qi*, whereas *ming* involves the manifestations of *qi* as phenomena and explicit influences. We have encountered the multi-layered word *shen*, which can refer to a spirit or what goes beyond our present, cognitive ability. *Shen* can come to inhabit the human body, and this is one goal of body cultivation.

Although translated as “spirit,” *shen* cannot be identified with a soul. However, it is important to point out that human being also possess spirits or souls, known as *hun* 魂 and *po* 魄. In the Western tradition, the soul is usually spoken of in terms of radiance and light.¹⁵ In the early Chinese context, the soul is not a single entity but these two interrelated forces. *Hun* and *po* are active in both the human body and consciousness. In Han medical texts, *hun* and *po* inhabit the human body and play an essential role in the body’s physiology.¹⁶ Ying-shih Yu argues that the ancient Chinese generally believed that breathing (from heaven) and eating (from earth) were the two basic human activities governed by the souls: *po* as bodily soul (*xingpo*) and *hun* as breath-soul (*hunqi*).¹⁷ *Hun* and *po* are living forces that form a union with the human body when one is alive; at death, they depart and leave the body. *The Elegies of Chu* (*Chuci* 楚辭) refers to this as “*hun* and *po* separating and leaving” (*hunpo lisan* 魂魄離散). They have their own fate. The *hun*-soul as *qi* moves quickly up to heaven, and the *po*-soul, as the heavier physical form, moves downward to earth. Therefore, one death ritual called the *fu* attempts to “summon the *hun* and return the *po*” (*zhaohun*

¹⁴ Roth. *The Huainanzi...* p. 127.

¹⁵ Tansley says that according to the Greek tradition, “soul is a radiant body of light, which they called *augeoides*, meaning ‘form of radiance.’” (Tansley D. V. *Subtle Body: Essence and Shadow*. London: Thames and Hudson 1977; p. 6).

¹⁶ Brashier. *Han Thanatology...* p. 145.

¹⁷ Yu Y.-S. *O Soul, Come Back! A Study in the Changing Conceptions of the Soul and Afterlife in Pre-Buddhist China* in *Harvard Journal of Asiatic Studies* 47, no. 2, 1987; p. 374.

fupo 招魂复魄).¹⁸ *Hun* and *po* are of great concern in (bodily) cultivation, because improper actions can cause them to leave the human body. One must avoid this “losing *hun* and destroying *po*” (*diuhun shipo* 丢魂失魄), a common expression even today.

With this in mind, at a deep level, the machines cannot flow like Dao. The “flow” of Dao, relies on the *shen*: the spirit, as the capacity to flow like Dao—but is spirit something capable of being reproduced in AI? In the case of AI, you have a machine that is made from inanimate substances, which you program to perform human-like tasks. While a central goal of AI is to design computers capable of recognizing and understanding human consciousness, the possibility of AI actually exhibiting human consciousness is another question. It is difficult to imagine the way in which one can upload *shen*, as *shen* is not an object, a computation, an algorithm, a piece of software, or a program, but rather something embedded in bodily transformations and social interactions. *Shen* flows, spirals, and transforms. Like *Zhuangzi* states in an anecdote in which a seasoned cook is cutting ably through the flesh of a cow, “My understanding consciousness comes to a halt...the promptings of the *spirit* begin to flow (神欲行).¹⁹

Another significant aspect of the Daoist vision of a human being is the ability to gain and lose self-awareness. If AI is at a point of developing self-awareness, can it also lose this awareness, as Daoism dictates? As the *Zhuangzi* also says, “Absorb yourself in the realities of the task at hand to the point of forgetting your own existence,” and “Let yourself be carried along by things so that the mind wanders freely.”²⁰ Like the loss and gain of self-awareness human traits like creativity and social intelligence cannot be easily replaced by AI. Daoist philosophy prizes these skills in humans, and it adds on one more trait to this difference, namely the *shen*. It is the *shen* that makes human beings different from AI, objects, and all other things. *Shen* is rooted in a social structure but it is also connected with cosmos. Clearly Amazon’s Alexa and Apple’s Siri have limited power to do so.

¹⁸ *Ibid.* p. 363. There is a T-shaped painting excavated in the Mawangdui tomb in Changsha that symbolizes this kind of summoning of the soul and offers an archaeological confirmation of the *fu* ritual.

¹⁹ Ziporyn. *Zhuangzi*... p. 22.

²⁰ *Ibid.* p. 29.

An Ultimate Quest for Genuineness/Trueness (求真 *qiuzhen*): Daoist Ethical Framework for AI Technology

Some AI scientists believe that next-generation AI, or Artificial General Intelligence (AGI), will be able to handle virtually any human task. AGI has no uniform definition, but experts say it should have the ability to reason, use strategy, solve puzzles, make judgments under uncertainty, represent knowledge, plan, learn, communicate in natural language, and integrate all these skills for achieving common goals.²¹ Should Daoists be worried about this revolutionary movement of AI?

On a superficial reading, one might conclude that Daoism opposes AI due to its theory of *wuwei*, non-action. That is, one could claim the development of AI represents purposeful human action that makes the world more complicated than the simplistic version Laozi imagined. Another argument is that AI is centrally concerned with giving machines the power to reason, representing the ultimate departure from *Zhuangzi's* idea of poetic dwelling. AI represents calculative thinking at its finest.

At a deeper level of understanding Daoist teachings, we will find that Daoism, in its profound grasp of the world, can support and validate AI's development, albeit with some distinctions.

Daoist teachings provide the conceptual tools that are well-positioned to deal with all kinds of changes. *Zhuangzi* advises us always "to go along with time" (与时共进 *yushi gongjin*): "When it was time to arrive, the master did just what the time required, and when it came time to go, he followed along with the flow. Resting content in the time and finding his place in the flow, joy and sorrow has no way to seep in."²² "To recurrently revert to the way" is actually "to go along with time." In other words, time refers to continuous changes and transformations. As heavenly time is constantly creating and recreating, in the same manner, humans carry out ongoing creative activity. The constantly emerging novelties of continuous changes demand that human beings go along with *shi* 时(time), locating themselves in the great flux of changes and transformations of the cosmos.

²¹ Walsh T. *Machines That Think: The Future of Artificial Intelligence*. Prometheus Books, 2018; p. 20.

²² Ziporyn. *Zhuangzi*... p. 24.

In Daoist classic text *Daodejing*, chapter 51, Laozi remind us “The Dao engenders all creatures, *de* 德 (virtue, power) nourishes them, *wu* 物 (material reality) confers physical forms on them, *shi* (particular circumstances) brings them to fruition.”²³ *Shi* offers the key to the actualization of things by adapting to the particular circumstances that defines various stages of a particular process and development. The popular slang of *flowing like water* captures this open spirit of Dao.

From a Daoist perspective, the ultimate goal of AI should be to build machines that can mimic human intelligence for the sake of allowing humans to be closer to nature or Dao. Daoists will not fear AI because it can never exceed human beings as a whole, only in specific programmable aspects. The flowing like Dao is to proceed on the basis of the conjecture that every aspect of human being should be alight with the movement of Dao.

For example, Jawbone or Fitbits build a personalized data set that includes information related to your identity, your profile, your biometrics, your age, your height and weight, your gender, your food preferences, your mood, your activities, your burned calories, and the quality of your sleep. By plotting over time, these devices construct a contextualized data set focused on your well-being. This data can be parsed by AI algorithms in a way that makes contextual sense. Wearable devices will tirelessly monitor our heart rate, blood oxygen levels, physical activity, breathing patterns, facial expressions, lung function, voice inflection, brain waves, posture, sleep quality, and more, in addition to taking external measurements, such as air quality and noise level. Using AI, these data points cannot just be turned into generalized information about your life as a whole, but rather into an actionable insight capable of improving health on a moment-to-moment basis. Carrying out both prediction and diagnosis, we will learn through these measurements what conditions are necessary for a particular illness or episode to occur and can take proactive and preventative steps to improve the quality of human life.

If Jawbone and Fitbit devices or in these ways are intended to serve as our biometric biographer capable of keeping us healthier and happier, why not? Daoists welcome these new devices. In fact, Daoists are

²³ *Daodejing*, translated with illuminating explanation by Hans-Georg Moeller, Open Court, 2007, 121.

well-known for their interest in and ability to develop very straight and specific guidelines for daily routines and selective food intake.²⁴ These guidelines include information on when one should go to bed or get up, what and when one should eat, and how best to manage daily one's rhythms. Daoists have used acupuncture needles and wild herbs to improve the human life span. AI optimizes measures that were previously unmeasurable, so biometric biographies offer life enhancing opportunities. These devices can only make our lives easier, more effective, and more comfortable and convenient in following Daoist dietary suggestions. They can progress toward developing what we might now deem super-normal human abilities, culminating in the search for immortality (in the same sense that body alchemy does). The quest for immortality allows Daoist practitioners to use all kinds of tools or technologies to extend human life.

However we can question whether Daoist teachings can offer ethical insights on AI? What is the ethical framework for a Daoist AI? At a broad and abstract level, Daoist teachings will insist on a non-interruptive technology that permits the natural rhythm of things. With the increased intelligence of AI comes the increased level of moral and legal accountability and responsibility.

The kinds of moral and ethical questions that Daoist philosophy focus on with respect to AI is whether AI will lead human beings closer to the Dao or if it will simply work in accordance with human reason by finding optimal problem-solving paths and furthering the qualities of natural human reasoning.

What kind of challenges does AI pose to human beings? To address this problem, Daoist teachings make a distinction between natural intelligence and artificial stupidity. Daoism warns human beings to avoid the fake intelligence (智 *zhi*, cleverness²⁵) that we are creating. It is not something that can compete with our sophistication. *Zhuangzi* makes a distinction between *renxin* 人性 (humanly nature) and *tianxing* 天性 (heavenly nature). Genuine human being “did not intrude into the heavenly with the Human.”²⁶ *Daodejing* chapter 5 claims that we should be guided by our natural stomach and not by socially constructed conven-

²⁴ Yongfeng H. *Introduction to Daoist Dietary Food and Method*. Religious Culture Press, Beijing 2007.

²⁵ On the question “*zhi*” see the *Daodejing* chapters 18, 19, 27, 65.

²⁶ Ziporyn. *Zhuangzi*... p. 106.

tional standards. Technology can be misused, operating against human nature and pushing us far from our inborn nature.

The ultimate pursuit for Daoist philosophy is the search for genuineness, which is quite different than satisfying merely desires. Daoist teachings draw a distinction between a quest for that which is genuine desire (求真 *qiuzhen*) and mere satisfactions (求欲 *qiuyu*). The *Daodejing* advises to “extend your utmost emptiness as far as you can and do your best to preserve your equilibrium”²⁷ in order to return to the original nature and be united with the Dao. Daoist teachings are aimed at an ultimate journey of being *zhen*, authentic and genuine human being.

The *Zhuangzi* celebrates a special type of moral human being called the *zhenren* 真人, “genuine person,” which is the highest rank of human being. *Zhenren* are capable of lifting heaven and earth, grasping *yinyang*, breathing pure *qi*, relying on spirit (*shen* 神), enjoying longevity, and mastering the timing of heaven (*tianshi*).²⁸ The *Zhuangzi*’s conception of the *zhenren* is a person who acts completely in accord with the natural patterns inherent in the Dao. The *Zhuangzi* teaches problem-solving, patterns recognition, information processing, method, models, and metaphors, capable of taking human experience in a variety of fields. It is for this reason that the “genuine person” (*zhenren* 真人) has the ability to act spontaneously (*ziran* 自然) in a manner that others cannot. The true person can act in this way because such a person follows along with the natural patterns, subsuming one’s own agency in that of the patterns themselves. A number of passages across a range of early Daoist writings discuss this ability to follow of natural patterns.

The idea that mirroring the patterns of nature makes one *zhen* is something we first see in the *Zhuangzi* in a number of phrases: “following heaven (*tian*),” “following *Dao*,” and “adhering to the natural propensities.” The general idea is that nature itself, or the ground of nature, has certain normative patterns such that when we align our conduct with them (which is most often a matter of getting rid of our own biases, etc), we act effectively.

In the Outer Chapters of the *Zhuangzi*, the idea is discussed in terms of “following the pattern of *tian* (Heaven)” (*xun tian zhi li* 循天之理), and linking this to potency or virtue (*de* 德). The criterion of actually living

²⁷ *Daodejing* chapter 10, p. 25.

²⁸ Guying C. 陳鼓應, 莊子今注今譯, *Commentaries on Zhuangzi*. Beijing: Chinese Press, 1983; pp. 168-169.

in accordance with (or even promoting) the natural order of the universe of which human flourishing is an integral part of what it means to “follow.” The assumption that rhythmic order and on-going generation are implicit in the fabric of existence is most apparent within the concept of *ziran* (自然), which most literally means, self (*zi*), so (*ran*). *Ziran* is often translated as “spontaneity” or “naturalness,” But it refers to what is so of itself, without any external force or coercion. *Ziran* is not only an element of the world but also the most potent mode of action for human beings. AI cannot grasp *ziran*. Human qualities like wisdom and love can be simulated but not duplicated in non-biological systems. The human brain is not merely a calculating machine operating on binary Boolean logic. It is embedded in a biological system with both analog and binary processes, with organs, tissues, a bloodstream, metabolism, and sensorimotor functions.

The human biological system is part of multiple energy fields in nature and the cosmos. If we consider the world as heavenly and earthly *qi* influence it is clear that the human brain is not a machine with a reset button. It is part of a process rather than a thing. The claim that *is *ziran* encompasses a view of the world grounded in uncertainty and novelty, that is, a “mysterious efficacy” (*xuande* 玄德).²⁹ The *Zhuangzi* makes many claims directing our awareness to this, such as that: “In motion be like water, in stillness be like a mirror, in responding be like an echo.”³⁰*

Final Remark

We are living in a globalized world where the scale and speed of technological and social change has been ever escalating and challenging the human ability to respond. Intensifying AI has been drastically transforming the mode of our lives. One of the most worrisome phenomena that we feel in our life today is uncertainty. We confront an ever widening range of uncertainty in our personal lives, in our working places, in our governments and around the world. On the other hand, we have not yet grown into the most needed mindset for contending with uncertainty. A Daoist framework hold that unpredictability and change are unavoidable and actually dominate everyday life. Daoist teaching

²⁹ *Daodejing* p. 121.

³⁰ Ziporyn. *Zhuangzi...* p. 123.

can assist us to challenge the liner cause-effect thinking that privileges order and stability.

Change is not merely a controlled move from one stable state to another. Change is integral to life and we should not balk at the realities of technological change. Consider *Zhuangzi* claims, "Whenever formation is going on, destruction is also going on."³¹ Hence all things are neither formed nor destroyed, for these two also open into each other, connecting to form a singular oneness. Change is an intrinsic and everlasting condition of all configurations, regardless of human desire, will, or planning. Thus, uncertainty is a vital part of our lives. It is not something external or temporal. Uncertainty is not a problem that needs to be corrected, but rather a condition to be prepared for and accepted. One can bear or react to uncertainty passively or embrace it and deal with it in an active and spontaneous way. Real strength entails flexibility; real wisdom entails uncertainty; real endurance entails resilience; real power entails humility.

The question of whether human beings will be rendered obsolete by AI is predicated on the extent to which new social relationships between human beings and machines can be imagined, thereby to develop a more philosophical understanding. A Daoist needs neither reject nor accept technological development without reservation. Based on the core values of Dao, a Daoist would neither resign himself/herself to whatever forms of technological development ensue nor unreservedly reject all forms of technological development. So the question for the Daoist is the extent to which technological development facilitates the human awareness of the Dao through the exercise of *ziran*. If the development of AI is currently premised on profit incentive, and if AI largely tends toward human obsolescence, the alienation of human beings from one another and alienation of human beings from the common good and natural world, then it can only be at odds with the Daoist teaching and practice.

³¹ Ziporyn. *Zhuangzi*... p. 27.

Educatio Vitae: Person-centered Ethics Education in the Age of AI

Sandra K. Alexander *

Since the 1963 encyclical letter *Pacem In Terris* by Pope St. John XXIII, numerous encyclicals have begun not just with warm benedictions to the faithful but also with the greetings, “to all people of good will.” The importance of these words – good and will – is worth reflecting on for students of ethics, as these words go to the heart of what they are asked to consider and define: what is the good, what is the will, and what makes a good will “good”? Over the last four decades, other authors of these encyclicals have increasingly asked all people of good will to consider the role of technology in society, the benefits and threats various technologies pose, and the overall role of technology in human flourishing and development.

For students of ethics in the age of artificial intelligence, these considerations take on an added urgency. Although many fears about the dangers of AI are overblown (the total replacement of human workers by machines comes to mind), students are still compelled to consider numerous implications of AI for human labor. The relationship between AI and work, where work is understood as an expression of human dignity and purpose, is worthy of sustained consideration by students today. Likewise, at a time when AI is expanding the reach of healthcare for those in need around the world, AI research and development still remain in the hands of a relative few and this raises serious questions about the disproportionate nature of access to this technology. Lastly, as the age of artificial intelligence is also, as Pope Francis describes it, one of technological “*rapidification*,¹ it is worth considering the contribution, if any, of AI to the “degradation” of our “common home” as well as to the degradation of human bonds.² Such are just two of the issues facing students of ethics today, but issues that I believe point to “goods” to be pursued at this time.

* Assistant Professor of Humanities, American University in Dubai (United Arab Emirates).

¹ Francis. *Encyclical Letter “Laudato Si”* (May 24, 2015). Acta Apostolicae Sedis 2015; 107: 854.

² *Ibid.* p. 852.

Ethics education in the age of AI: considering the goods of dignity and community

So, returning to that appeal first made in the *inscriptio* of *Pacem In Terris* but now viewed in the context of our present situation, we might ask: what should “all people of good will” reflect on in the age of AI? More so for the student of ethics, how should we define “the good” in the age of artificial intelligence? Set within a consistent narrative addressing the fate of humanity and all creation, the views set forth in various encyclicals by Popes St. John Paul II, Benedict XVI, and Francis suggest compelling answers to these questions.

I would now like to turn my focus to a few of their observations on the impact of technology on the world of work and on social bonds. For those like myself who teach ethics, these observations offer a clear framework for ethics education in the age of AI, a framework where the preeminent concerns are for human dignity and community. It is a person-centered education that can be described as *educatio vitae*.

Let us begin with the comments of Pope St. John Paul II on the impact of emerging technologies on human labor, those found in his encyclical *Laborem Exercens* of 1981. It is “through work”, he begins, that “man must earn his daily bread”.³ He continues that, “from the beginning... he [mankind] is *called* to work. *Work is one of the characteristics that distinguish man from the rest of creatures [...]*. Thus work bears a particular mark of man and of humanity”.⁴ It is for this reason, he cautions us, that the issue of human labor is “a perennial and fundamental one, one that is always relevant and constantly demands renewed attention and decisive witness”, for it is from work that a person’s life claims a “specific dignity”.⁵ The particular technological threats to human labor and dignity mentioned by John Paul II may appear somewhat outdated to us in the twenty-first century, but it is striking how his concerns over the rise of automation prefigure our concerns today over the impact of autonomous and semi-autonomous systems on skilled workers.⁶ The introduction of automation in various industries at the time of this

³ John Paul II. *Encyclical Letter “Laborem Exercens”* (September 14, 1981). *Acta Apostolicae Sedis* 1981; 73: 577.

⁴ *Ibid.*

⁵ *Ibid.*, p. 578.

⁶ *Ibid.*, pp. 578-579.

encyclical, which the Pope says may “mean unemployment [...] or the need for retraining”,⁷ confronts the fact that work helps man “realize his humanity, to fulfil the calling to be a person that is his by reason of his very humanity.”⁸ The implications of this confrontation between labor and automated systems, and its particular impact on human dignity, are as significant today as they were in 1981, and it is this good of human dignity that students of ethics must affirm in the age of AI.

While *Laborem Exercens* focuses in part on concerns over human dignity in the context of work, in *Laudato Si'* Pope Francis invites (again, all people of good will) to consider the loss of human dignity and freedom in our era of technological “rapidification”. Although much of this letter addresses the problem of human-driven environmental change and other threats to our “common home”, the Holy Father also reflects on the impact of unnaturally rapid technological development on human bonds and community. As he warns: “(our) immense technological development has not been accompanied by a development in human responsibility, values and conscience”, and that “(it) is possible that we do not grasp the gravity of the challenges now before us.”⁹ What examples can we find of this disconnection between technological development and conscience? Numerous. I would suggest it is manifested in the use of semi-autonomous robots in caring for the elderly and others in need of special assistance. On this topic David O’Hara explains, through the use of such “care machines”, “(m)aybe our intention is to distance ourselves from the difficult work of care. Our machines might offer one kind of care, while being the physical expression of our lack of interest in those who need the care.”¹⁰

The essential role of ethics education in the age of AI

For students with the ambition to design such machines and systems – machines that will play a role in our “caring” for one another or displace the human workforce – these are pressing concerns. For ethics educators, the age of AI faces us with a somewhat different, but

⁷ *Ibid.*, p. 579.

⁸ *Ibid.*, p. 584.

⁹ Francis. *Laudato Si'...* p. 889.

¹⁰ O’Hara D. *How Robot Priests Will Change Human Spirituality*. Medium OneZero. January 3, 2020 (accessed on 02.25.2020 at: <https://onezero.medium.com/how-robot-priests-will-change-human-spirituality-913a19386698>).

no less significant concern, one that could undermine our attempts to address these very issues with students: the very role of ethics study *in education* today.

As noted earlier, ethics courses typically ask learners to consider what defines “the good” and “the good will”, but the study of ethics also challenges students with questions concerning truth. It is on the role of truth in human development that Pope Benedict XVI focuses part of his 2009 encyclical *Caritas in Veritate*. Here, the Pope Emeritus affirms the Church’s narrative on human dignity while highlighting elements essential to integral human development: charity, truth, “justice and the common good” amongst them.¹¹ On the subject of truth, he cautions that without truth “social action ends up serving private interests and the logic of power, resulting in social fragmentation”.¹² For an ethics educator like myself, this leads me to ask, what role does asking questions about truth play in the lives of students in the age of AI or in any age? The answer is, as it was in the early days of the first universities, that it plays an essential if not the most important role in the education of the whole person.

Yet, the study of truth by way of ethics, moral philosophy, and theology is under pressure in the age of AI, as is the student-centered, person-centered approach to inquiry it requires. We might say that this pressure is the result of the techno-scientific demands being put on university curricula, where students are asked to engage in more complex, technical training at the expense of studies in the humanities.¹³ But is it realistic to return to a kind of education where ethics is at the foundation, or is it as Matthew Milliner describes a kind of “educational romanticism”¹⁴? I would answer that putting the study of ethics at the centre of education is as realistic as it is necessary. From St. Augustine, through Clement of Alexandria, to St. Thomas Aquinas, there is agree-

¹¹ Benedict XVI. *Encyclical Letter “Caritas in Veritate”* (June 29, 2009). Acta Apostolicae Sedis 2009; 101: 644.

¹² *Ibid.*

¹³ Cf. recent events at Saint Louis University (USA), a Jesuit institution, and the proposed removal of ethics courses from its core curriculum. See Bishop J. SLU’s core meltdown runs the risk of losing its distinctiveness. St. Louis Post-Dispatch. February 8, 2020 (accessed on 02.25.2020 at: https://www.stltoday.com/opinion/columnists/jeffrey-bishop-slu-s-core-meltdown-runs-the-risk-of/article_63647ea1-5fa3-5d70-b8da-2b8f0831fd73.html).

¹⁴ Milliner M.J. *Medieval Wisdom for Modern Universities*. Public Discourse. April 6, 2011 (accessed on 02.25.2020 at: <https://www.thepublicdiscourse.com/2011/04/3106/>).

ment on the centrality of theology and moral philosophy in education. I would say that all of these men understood that, “if knowledge is not planted in the seedbed of wisdom, it [will] either never take root, or – far worse – grow into something dangerous”; and “that ethics [is not] a sub-discipline of the educational curriculum, but [is], in a way, its entirety.”¹⁵

In conclusion: a call to work toward the goods of dignity, community and person-centered “*educatio vitae*” in the age of AI

But now, to offer some concluding observations, directed once again “to all people of good will” in the age of AI. What goods should we pursue, as individuals and as a community? And by “we” I mean those of us in education, in industry, in the public sphere, whether religious or secular. Two of the pre-eminent “goods” affirmed by these encyclicals are those of human dignity and the bonds of community itself. A third “good”, the study of truth, allows us to defend passionately the role of ethics education the age of AI. But the purpose of ethics education and its practice is not simply the concern of educators, academics and students. Ethics education must be part of a larger public encounter focused on considering the “good” in the age of AI, one that I believe both the Church and this very workshop are committed to leading.

¹⁵ *Ibid.*

Second session

ARTIFICIAL INTELLIGENCE AND HUMAN HEALTH

Artificial Intelligence and health

Walter Ricciardi*

Artificial intelligence plays a part in the profound evolution of the world's healthcare systems, which I believe must be prepared to avoid what could become a perfect storm.

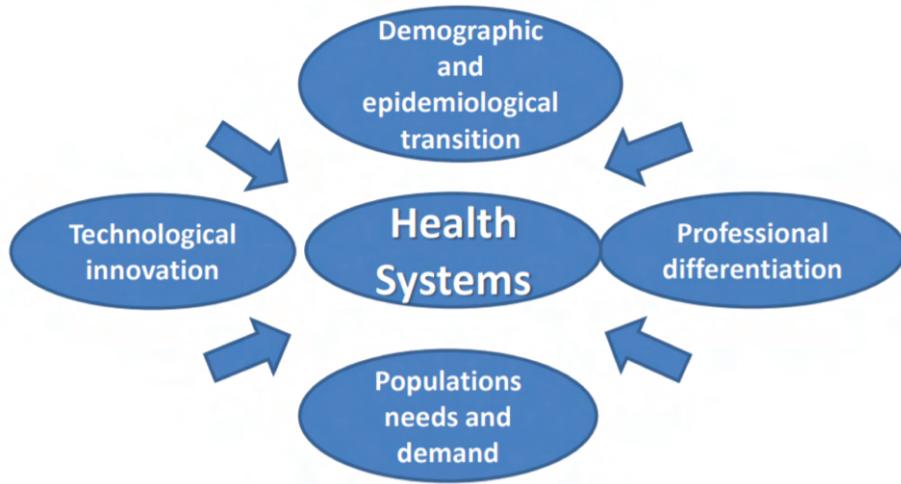
What is a perfect storm? In addition to a very famous Hollywood film, it is an expression that describes an event in which a rare combination of circumstances occurs at the same time, making the situation drastically worse. In the seas, all over the world, there are storms every day, but we rarely read in the newspapers about shipwrecks and casualties, because even if the sea is rough and the winds are strong, ships are solid, captains are good at what they do, there is enough fuel, they have the radio, and, possibly, if ships are not far from the coast, they may be rescued and no one dies; however, if adverse events occur simultaneously, as in the film, everyone perishes.

What are the waves of this perfect storm that can hit the healthcare systems all over the world, even the strongest ones? They are the "waves" of supply and demand for healthcare services.

The waves of demand are:

- the unprecedented demographic and epidemiological transition we are experiencing the world over (because today's challenges are shared by the richest countries and low middle-income ones);
- as a consequence, a huge increase in people's needs;
- unprecedented technological innovation (that we are discussing, which naturally includes artificial intelligence, but also new drugs, new vaccines, highly sophisticated technologies, that however are expensive and must therefore be somehow carefully managed);
- the lack of physicians, without whom healthcare services cannot be provided (this is happening for the first time because while a shortage and professional differentiation of doctors and healthcare workers have always existed, this is now a problem also in developed countries).

* Professor of Public Health, Università Cattolica del Sacro Cuore (Italy).



In 1970, when the Servizio Sanitario Nazionale, the Italian National Health Service, was established, the Italian average family was made up of two children, two parents, three grandparents; the situation was difficult, but it was sustainable.

Today, Italian families are made up of one child, two parents, four grandparents, two great-grandparents...

You do not need to be a demographer or an economist to realize that this situation is unsustainable, particularly if young people live in a precarious economy and if the elderly become ill. Although they live longer, the elderly are affected by chronic illnesses and, for reasons linked to inappropriate eating habits especially during childhood, they become ill at an increasingly earlier age. Years ago, it was middle-aged people who developed diabetes, while today even teenagers may be affected by it.

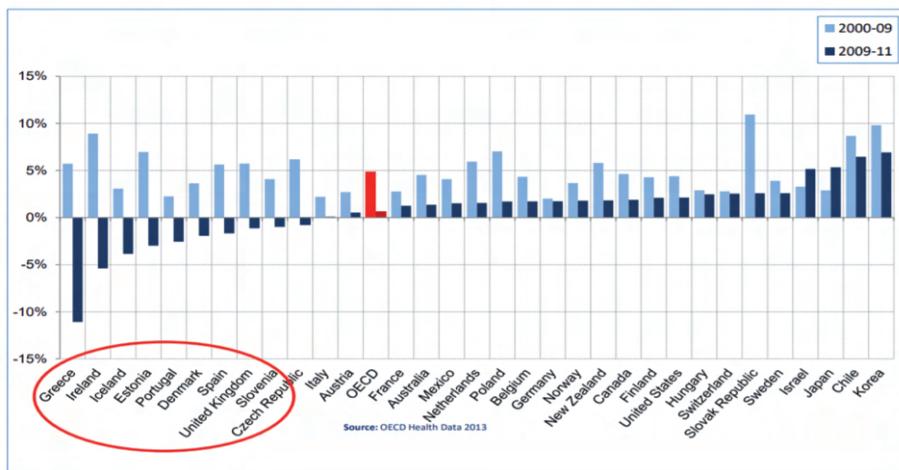
As a consequence, in their old age, people are often not only affected by one chronic disease but by three or four, which entails huge costs because treating a large number of people affected by chronic diseases throughout their lives is expensive. Today, to treat the five main diseases alone (dementia, diabetes, stroke, cardiovascular disease and cancer), the European Union spends one trillion euros, an amount so large it is difficult to grasp; however, if we do nothing to manage this situation, within a short time this figure will increase to six trillion.

All this is happening in a context of great financial constraints, in the sense that the global crisis, which originated in 2007–2008, started

having such an impact on countries that, instead of making more resources available for healthcare, have made them scarcer.

Instead of increasing healthcare budgets, left-wing countries have cut them.

Average annual growth in health spending in real terms



So, what can we do for our healthcare systems?

I was the rapporteur for the White Paper on Sustainability for European countries, which advised all EU politicians to invest in healthcare, because return is not only reaped on health conditions but on society as a whole.

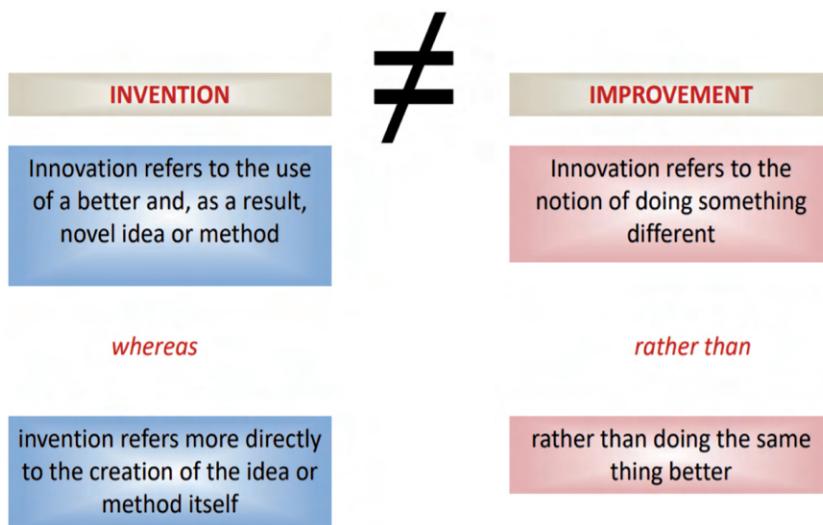
We told citizens to improve their health and try to change their behaviour.

We told the healthcare services to try and promote prevention and act at an earlier stage, to try to make citizens responsible and to reorganise healthcare.

What does it mean to reorganize healthcare? It means solving the problems that affect citizens: delays in treatment; the same treatment for everyone; variability in health conditions that is the result of choices and not of natural developments; keeping patients waiting; uncertainty about what really works; frequent medical errors (they occur naturally, it is only human: the problem lies in the fact that they are not notified, hence organizations are not structured as to learn from mistakes); irrational workflows and patients who increasingly ignore doctors' instructions.

We no longer see human beings holistically, what we see are organs and people as "fragmented", based on their diseases.

These are the problems, but how to deal with them? By innovating, by changing! Because we cannot make use of old solutions for old problems, and, all the more so, we cannot use old solutions for new problems. Hence, we need to change somehow; innovations are not inventions, they are not creations: innovation is the process of translating an idea or an invention into a product, a service that adds value, doing something better in an original way. Therefore innovation is profoundly different from invention, because while the former refers to the use of a better idea of a method, the latter refers to the creation of the idea of the method itself. Improvement is basically something more than doing the same thing better: "doing more" of the same things or "doing them better" is not the answer.



There are two types of innovations.

What we are discussing today is disruptive innovation, which creates new pathways, new organizations, it even creates new values, it involves new players, it makes better health conditions possible and other important objectives such as equal access to be achieved. We want our citizens to be taken care of in a way that is appropriate to their conditions, everywhere, not only in rich countries. However, innovation – and this is something we must be aware of – disrupts the old systems and the old way of doing things; it is disruptive and has deep

economic implications, because it completely alters the markets and, as it invariably happens, this means that there are losers and winners, which can lead to imbalances.

As far as healthcare is concerned, the main effects of disruptive innovations (such as artificial intelligence) are:

- improving health conditions;
- creating new services and overcoming challenges such as accessibility to existing or new services;
- leading to something that is cost-effective, because resources are limited, and, during crises, it is even more important to use them appropriately;
- promoting person-centred healthcare;
- enabling people, especially the elderly or those in need of long-term care, to be increasingly able to maintain, as far as possible, good quality of life;
- creating new professionals at the service of the community.

In short, innovations change the way we think, radically change the way we behave and therefore disrupt the old systems; of course, not all innovations should be accepted and financed, but only those that add value: at a time when resources are scarce, we cannot afford to finance creations, inventions, technologies that do not add value, those values that all of us here share.

Some classic examples can illustrate this classification; innovations can be technological, organizational, concern products, human resources, but what we are considering today is an innovation that involves what the European Commission considers the five strategic areas for disruptive innovations:

1. translational research, that is, research that develops new drugs and new technologies to solve problems;
2. technology per se, and (which is the next topic I am going to address) artificial intelligence technology;
3. precision medicine;
4. health promotion (i.e. helping people to stay healthy or trying to ensure they get sick as late as possible);
5. new training for health workers (we continue, for example, to train doctors with the same approach adopted two centuries ago in most universities, and we do indeed train professionals who may be very good from a technical point of view, but who do not have a systemic vision and find it difficult to fit into the contexts we are discussing).

Artificial intelligence refers to systems that, in some way, strengthen the intelligent behaviour of humans by acquiring a certain degree of autonomy to achieve precise objectives. As you know, artificial intelligence is used not only in the healthcare sector, which is one of those in which the greatest developments are taking place, but also in the defence and in the security sector... in short, artificial intelligence is something that pervades all areas of knowledge and all human behaviours. But it is gaining gradually greater importance in the field of healthcare.

Indeed, this can have extraordinarily positive effects, but some caution is necessary: first of all, artificial intelligence will lead to the creation of new jobs, which from an economic point of view is an extremely positive fact. Moreover, these solutions will give the countries, which adopt them and as a consequence change their organization, the opportunity to build their own artificial intelligence capacities.

Will doctors be replaced by algorithms? No, but organizations that do not use algorithms will be replaced by organizations that do, so basically those who will not adapt will be excluded, marginalized. Artificial intelligence will emerge to create an increasingly sophisticated system; everybody agrees on the idea that computers (or better robots) will almost entirely replace some activities, so human intelligence will have to devote itself to things (healthcare being one of them) in which the human relation component is irreplaceable: no robot will ever replace a hand, a caress, a hug, a moment of sharing, this will never be replaced; but those who will only do that will be marginalized by those who will use artificial intelligence to better treat their patients. Of course, this then raises unprecedented issues from an ethical and from a medico-legal point of view: who has made a mistake when something goes wrong, the robot or the doctor? Is it the living person or the artificial person? This, of course, requires a whole different system. And how can you introduce artificial intelligence based on the same standards used to introduce drugs or medical devices? How do you go about reimbursing them? It is evident that challenges are huge in all respects.

Of course there are also great difficulties: if all over the world different standards were to be used, this would turn into a non-interoperable Babel and this is to be avoided at all costs; of course there are moral

and ethical implications – which is what we are discussing – as well as equal opportunity issues. If everything is only driven by the market, rich countries will benefit, while poor countries will be even more marginalized; and again, within rich countries, the more accomplished, the more educated people will be able to reap the benefits, while poorer, less educated people will lag behind. It is astonishing to think that artificial intelligence may promote inequalities rather than equal opportunities.

And of course there are concerns about privacy, security, also consequently to the fact that few people are able to handle these problems.

When I was President of the Italian National Institute of Health (Istituto Superiore di Sanità), a meeting was held with all the Presidents of the Institutes of Health in the world, and what emerged was that public systems, for example, lack lawyers who (at prices that a public administration can obviously pay) contribute to drafting laws. If capable lawyers are only to be found in the private sector, while there are no people dealing with these issues in the public sector, it becomes a problem.

This is a crazy market in terms of growth, there is no other human sector that is marked by a pressing, remarkable and impetuous growth as the healthcare sector. And, in the healthcare sector, there is probably no area that experiences higher growth than that of technological development: we are talking about 20-30% per year, incredible rates, that undoubtedly see North America, and especially the United States, at the forefront. But this is self-evident because the United States has a university system that is clearly marked by an approach deeply linked to market development.

At the moment I am chairing the Mission Board for Cancer of the European Commission, which has allocated 100 billion euros for its five missions, of which 20 billion for the fight against cancer. We want to follow this model, but there is no doubt that Europe is lagging behind the United States and it is no coincidence that 50% of these developments will take place in the United States.

What are the challenges we have to face in some way? The challenges are the following:

- How are we going to integrate old health workers, people of a certain age who have been trained in a completely different way, in clinical common sense, observation, benevolent paternalism? How can these people adapt to such a rapid, massive, broad change?

- How can we deal with regulatory issues? This has to be done at the international level, it cannot be addressed only nationally. In our case, it is the task of the European Union.
- How do you get clinicians and patients strongly involved? Both need to be involved because there can be no healthcare system that works well for a patient unless also the healthcare provider is at ease with it; if the work of healthcare providers is made difficult in some way or if healthcare providers do not understand the system, patients will be the first to pay for it.
- Then there is the great problem of data quality and privacy.

In this connection, I invite you to look at the huge effort that Tim Kelsey of the Digital Health Agency has made in Australia: 24 million Australians have handed their data over to the State (which, of course, guarantees they are protected and used wisely). It was a process that had not progressed for ten years and was resumed thanks to the leadership of Tim Kelsey, who managed to give a concrete answer to all these questions.

There are also instances of hospitals that, while waiting for healthcare systems to deal with these complex problems, are working hard to become operational. By transforming its organization and digitizing it completely, Langone Hospital has demonstrated that a healthcare centre can move from ranking 60th to second in the list of the best hospitals in America, from the 34th position to the top ten in the education of its professionals, and move from a loss of \$150 million a year to an annual gain of \$240 million. This is a process that must be managed, this is a process that must be coordinated.

In conclusion, artificial intelligence can be (and I believe will be) an important tool; it can provide new and different perspectives that tend to reduce complexity in favour of empowerment, the empowerment of citizens and patients; it must be seen by policy-makers as a new way to solve old problems, even though healthcare systems must act responsibly and be careful in facilitating the introduction of disruptive innovations. These must be tested, the market cannot choose freely, because it would create winners and losers, and the losers would be the usual ones: the poor, the marginalized, those who do not have access to quality education, those who do not have access to better tools.

Therefore, both politicians and technical experts will have to be open-minded, but also possess strong managing skills. There cannot be a solution appropriate for everyone even though the challenges affect-

ing the world are common to all; today, African countries face a double burden of disease: the infectious diseases we have been discussing these days, for instance in connection with the coronavirus, are increasingly being accompanied by chronic diseases, so also African countries have to deal with diabetes, overweight and obesity, cancer, and cardiovascular diseases.

However, I believe that by working together, and this is mandatory, we will be able to find an answer, which is ultimately not only technical, but above all ethical and moral. Thank you for your attention.

(Translated from Italian by the Pontifical Academy for Life)

The Clinical Consequence of AI

Yuzo Takahashi *

Introduction

With the implementation of AI in medicine and the drastic changes it will bring about, we are entering a new historical era marked by an exponential transformation of health care systems.¹ The major concerns resulting from AI implementation are summarized in BOX 1, and this contribution mainly deals with the clinical consequence of point 3, i.e. doctors losing monopoly over medical skills.

It is widely acknowledged that access to health care is determined by the availability of resources such as money, technology, human resources, etc. In other words, health care providers and patients have always had to deal with restrictions, to some degree. Some of these restrictions may be reduced thanks to the implementation of AI technology in medicine, especially in the medical skill and business domains, therefore improving health care and human welfare. To take full advantage of this incredible technology, we must be familiar with the way in the medical system works, and how we can incorporate the available AI technology to maximize its beneficial effects and avoid adverse ones.

By virtue of its accuracy and limitless use, AI provides for two advantages. First, it can level medical services across patients, and second, it promotes a transition from hospital-centered to patient-centered medicine, which cannot be achieved by resorting to conventional technology.

* Professor Emeritus, Gifu University, Gifu (Japan). Visiting Professor, Hyogo College of Medicine, Nishinomiya (Japan).

¹ Loh E. *Medicine and the rise of the robots: a qualitative review of recent advances of artificial intelligence in health*. BMJ, Leader, 2018; Vol. 2-2. (Accessed on 01.06.2020 at <http://dx.doi.org/10.1136/leader-2018-000071>).

There are, however, adverse effects. One minor disadvantage is the loss of jobs, which has already been documented by many authors and is seen as manageable in the long term. The major problem, however, is represented by the relationship between doctors and patients. Thanks to AI software, patients are likely to obtain almost the same knowledge and skills possessed by doctors. Should this be the case, how can doctors continue to be respected by their patient as they used to?

AI technology is a doubled-edged sword, and as such, the consequences of its implementation in clinical medicine seem to depend largely on the response of stakeholders. Usually, the health care system works best when it is based on rapport, with minimal law involvement. The time has come to think about the importance of face-to-face communication in order to establish rapport, a prerequisite for offering the best health care.

BOX 1 Major concerns in medicine following AI implementation

1) Appearance of new “human doings”

Human beings establish or determine their personality and behavior through personal experience gained after birth. The older generations did so through interaction with family and friends. However, future generations will do so through their interaction with AI software, and may develop a new type of personality and behave differently. Such new “human doings” are beyond the comprehension of the old generation. Nevertheless, doctors will have to interact with the new “human doings” and may be confused as to how to proceed.

2) Transition from direct evidence to indirect evidence

Currently, doctors make decisions based on a few pieces of direct evidence, but AI software will do so based on numerous pieces of indirect evidence. As a consequence, the strategy adopted by medical science and practice will be drastically influenced by this. Doctors will have to reassess the value of big data.

3) Doctors lose monopoly over medical skills

The high esteem in which modern doctors are held is the result of professional skills that patients have no access to. However, in the future, diagnostic ability will be shared with patients and cause a rebalancing of powers between them and doctors.

AI implementation in medicine is beneficial except for some adverse effects

An amazing technology that improves medical services

It is well known that health care is a sector that already benefits from AI technology. AI-assisted medical equipment, diagnostic software and administration algorithms are now being implemented in medical providers' routine workflows, drastically improving medical services from the viewpoint of quality and quantity.² See BOX 2.

BOX 2 Possible consequences of AI

Beneficial effects

Quality; accurate

Quantity; cost effective

Adverse effects

Minor; job loss

Major; doctor-patient relationship

AI can improve the quality of health care for two reasons. One is that some AI software provides more accurate and earlier diagnoses than human doctors. The second is that AI allows medical providers to save time, thereby allowing them to concentrate on more essential work and avoid carrying out routine work every day.

Generally speaking, the quality of the medical services available to patients is not equal because of constraints related to money and resources. High costs represent a big problem in medicine; as a consequence, medical services are provided in a hospital-centered manner so that as many people as possible can receive standard health care at an affordable price. On the other hand, patient-centered medicine is convenient but not common because of its high cost, and therefore is beyond the reach of most people.³ See BOX 3.

AI implementation reduces the cost of medical services, and as a consequence it may trigger a historical transition from hospital-centered to patient-centered medicine. Furthermore, cost cutting will bring about

² Amisha, Malik P, Pathania M, et Al. *Overview of artificial intelligence in medicine*. J Family Med Prim Care. 2019; 8(7): 2328-2331.

³ Beauchamp TL, Childress JF. *Justice*. Principles of Biomedical Ethics. 5th ed. Oxford University Press. New York. 2001; pp. 225-272.

another important benefit by enabling more people to receive standard health care regardless of their financial situation. Naturally, equal health care for all patients is what medicine should aim for and achieve.

BOX 3 Two types of health care systems

- (1) Hospital-centered medicine at provider's convenience and priority, less costly
- (2) Patient-centered medicine at patient's convenience and priority, costly

An example of the beneficial effects of AI

The notion of “social hospital” is an example of the patient-centered medicine of the future. Medical equipment is embedded in a city as a social infrastructure.⁴ In other words, the hospital function is expanded beyond the boundaries of the hospital, and patients receive medical services at their own convenience and based on their priorities.

When this becomes a reality, even a patient with a heart problem will be able to walk along the road freely. In the case of hospital-centered medicine, a patient with an unstable condition should be hospitalized and be kept strictly under the care of doctors.



BOX 4 The Social Hospital

The hospital function is expanded beyond the boundaries of the hospital, allowing patients to obtain health care in town.

The patient and the doctor are at the center, a remote blood test unit is bottom right, and equipment in a remote area is to the upper left.

The mechanism underlying the social hospital is the following: the patient has a wearable monitor, and vital signs such as pulse rate, blood pressure and temperature are transferred to the hospital via the internet

⁴ Kuroda T. *Kinmirai no iryou no Sugata Social Hospital*, in Japanese Zinkoutinou zidai no iryou to igakukyouiku. Shinoharashinsha. Tokyo, Japan. 2016; 3-4.

in order to analyze data on a 24-hour basis, using AI software. In case of emergency, the primary doctor is alerted, and, if necessary, the doctor sends an ambulance to the patient to provide hospital-based care.

A remote blood test unit distributed in the town is another essential component of the social hospital. Patients go and give a blood sample, and the blood test data is sent to the hospital and processed for remote diagnosis purposes. This system enables patients to go anywhere freely. See BOX 4.

When this becomes a reality, patients will visit hospitals less frequently, and use their free time to enjoy life. “Cure the disease and save the patient’s biological life” is the primary goal of medicine, but the QoL of patients should be prioritized in the health care provision process.

Adverse effects of AI

Although AI implementation is beneficial to patients and medical providers alike, there are both minor and major problems to be considered.⁵ The former is job loss affecting some professionals, including doctors with excellent skills which took many years to develop. This kind of issues have been common throughout history as civilization evolved, and are the inevitable byproduct of technological progress. The job loss problem is manageable, but the major problem is the adverse effects of AI on the doctor–patient relationship. This will be discussed in relation to the main topic of this article in the following subchapters.

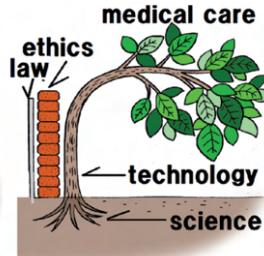
Factors that influence decision-making in medicine

How modern medicine works

Decision-making in health care is a key process, and decisions should take many factors into account. For a layperson, the process is not easy to understand because of the range of multi-faceted factors involved, including law, ethics, medical technology, and medical science. See BOX 5.

⁵ Rigby M.J. *Ethical Dimensions of Using Artificial Intelligence in Health Care*. AMA Journal of Ethics. 2019; Vol. 21, 2. E121-124.

BOX 5 Functional structure of medicine



Best health care needs rapport and medical technology. Technology depends on science. The medical tree needs supporting structures (ethics, law) as it grows taller.

This complex functional role in medicine can be best depicted using the analogy of a tree. Daily health care represents the leaves, medical technology the trunk, medical science the roots, and ethics and law the supporting structures of the tall tree. Medical technology needs science just like a tree needs its roots. A tall tree needs a supporting structure such as a stack of bricks and a metallic pole.

Although the relationship among the various factors may be depicted in the form of the elements making up a tree, factors often conflict with each other. This will be discussed in the paragraph below. Of course, medicine used to be simple in its primitive, early history, and it comprised few laws, ethics, and simple technology based on experience.

Therapeutic decision-making involves many conflicting factors

First of all, we have to consider that while making a diagnosis is a purely scientific matter, therapeutic decision-making is a social matter, in which ethical principles and values as well as cost and law all play a part.

Therapeutic decision-making is the most difficult of all decision-making in health care. A typical example is a life-threatening disease, either chronic or acute, in whose connection doctors have to make critical decisions every second. See BOX 6.

BOX 6 Factors influencing therapeutic decision-making

Diagnosis

Available resources; costs, human resources

Patient factor; physical condition, best interest, autonomy

Unforeseen (unknown, unpredictable) factors

Specific factors that should be considered are diagnosis, patient preference, costs, as well as unforeseen factors. These factors are often changeable; therefore, decision priority should be made tentatively, taking conflicting factors into account. The decision should be lawful and ethical, but which ethics has priority? The patient's or the doctor's? The patient's best interest and autonomy should be respected, but patient preference is not always achievable.⁶

Examples of a conflicting case are given below:

- (1) The disease is curable, but the cost is not affordable for the patient.
- (2) The disease is curable by amputation surgery, but the patient does not accept such a radical operation.
- (3) There is a technology that can save an infant patient's life, but parents do not accept it because of their religious beliefs.

Who should make the therapeutic decision, and how?

Quality health care is the result of rapport; the involvement of law should be minimal

What is the best way to regulate daily medical practice? We know that should the law and contracts stipulate medical practice at every level, no good would come of it. In medical practice, in which both verbal and nonverbal communication prevails in the doctor–patient interaction, it is not practical to adhere strictly to verbal-written law.

BOX 7 Doctors need to be more approachable and professional

- To be respected and trusted by patients
- To establish and maintain rapport
- For health care to promote wellbeing without misinterpretation
- For patient's adherence to doctor's decision

Daily medical practice should be based on rapport, keeping the involvement of law or contracts to a minimum. If rapport is well established, most routine health care will be managed at the discretion of

⁶ Beauchamp TL, Childress JF. *Respect for autonomy*. Principles of Biomedical Ethics. 5th ed. Oxford University Press. New York. 2001; pp. 57-103.

doctors, resulting in very efficient and high-quality care. In this sense, the importance of rapport should be emphasized. See BOX 7.

We need a judge who makes wise therapeutic decisions, taking every relevant aspect into account

In the case of a complex disease, there is no “one size fits all” solution. The patient has to give up something in order to keep something more important, in a kind of trade-off. This prompts the controversial question: Who should make a therapeutic decision? The doctor? The patient? The family? Third parties? Or AI? No conclusive answer has yet been found, but one thing is certain, we do not need dictators who decide everything based on their own thinking.

On the contrary, we need a person who acts as a judge taking every relevant aspect into account and making wise balanced decisions in daily health care, asking the opinion of institutional ethics committees or multidisciplinary teams, depending on the situation. See BOX 8. In most cases, doctors are the best candidates to act as judges.

BOX 8 Therapeutic decision-making policy

Category

major principle; institutional ethic committee
difficult case; multidisciplinary team
routine work; doctor's discretion

Meet the requirement

lawful, ethical, scientific

Patient involvement

best interest of patient, autonomy of patient

Doctor's function as a medical judge

It is desirable for laypersons to be involved in the institutional ethics committee in which the principle of ethics is discussed and created. When this principle is applied to make a decision for a difficult individual case, a multi-disciplinary team should be involved. In all cases, doctors will play a crucial role in providing the medical knowledge necessary for the discussion. These situations are relatively rare, and in daily practice, the decisions which doctors make are minor in comparison, often routine work needed to provide daily health care. These repetitive duties are well managed within ethical principles. As such,

it would be practical that routine medical decisions are made at the discretion of doctors.

To make a wise decision, the medical judge needs to consider the situation in a balanced manner, taking into account relevant factors, including both known and unknown factors that change over time. Furthermore, the judge should always consider whether the decision is within or beyond ethical principles, or, in the presence of a borderline case, whether discussion within a multidisciplinary team is required. This is a typical role played by the medical judge, and it works well if rapport is well established. Of course, whether a medical judging system works at a satisfactory level or not depends on the establishment of a close rapport among the stakeholders.

Once the decision is made by the doctor, the patient supposedly accepts it. If not, the patient should no longer be placed under the doctor's care.

Doctors lose monopoly over medical skill after AI implementation, resulting in changes in the balance of powers with patients

Doctors earn their patients' respect for two main reasons: their excellent skill and good personality. Doctors of modern medicine have had a monopoly over medical skill and knowledge to which patients have had less access; this resulted in great respect for doctors. However, doctors are losing this advantage. Suppose patients installed AI diagnosis software in their smartphones, and then presented themselves at the hospital. In this situation, patients would have access to the same medical knowledge and diagnostic skill as doctors.

This would cause a historical change in the doctor-patient power balance relationship. This rebalancing should not be overlooked when considering the clinical consequences of AI, one of them being that patients tend to respect doctors less than they used to because one of the main reasons for that respect is diminishing. How doctors can continue to be respected by patients is therefore a big issue.

Doctor's judgment is valid only when rapport is established

As a medical judge, a doctor makes a decision at his or her discretion in daily health care provision. The decision is supposed to be wise, but despite its necessity, it is not sufficient for patients to obey the doctor. Who would listen to what he or she is told in the absence of warm human touch? Who would obey doctors lacking noble character?

For this reason, doctors should be trusted, respected and approachable. In other words, doctors should establish rapport with patients for their decisions to be valid and effective, otherwise their decisions would be of no use.

BOX 9 The three domains of medicine

- (1) business; cost, law, contract
- (2) skill; technical, science
- (3) humanity; trust, respect, rapport, ethics, responsibility

AI can greatly contribute to the business and the skill domains of medicine, and plays a major role towards improving health care and human welfare, but not towards improving humanity. The new AI technologies would not be able to replace any element of the humanity domain.

As mentioned in the previous subchapter, diagnosis is made based on purely scientific evidence. On the other hand, therapeutic decision-making is based on both scientific and non-scientific factors, and is more of a social matter encompassing diagnosis, possible prognosis, available resources, patient best interest and the autonomy of patients. Furthermore, health care provision needs a warm human touch to be fully successful. A doctor's words alone can instill hope in patients. Human factors are important and are a characteristic feature of medicine, an aspect which should always be considered in order to improve the medical system and allow it to offer efficient and high-quality services.⁷

We know that close and positive relations between doctors and patients based on their humanity (rapport, and mutual trust & respect) have a positive influence on fighting against disease and maintaining good health. See (3) in BOX 9. This is still true even after AI implementation. Therefore, AI should be used so as to enhance, not impede, the formation of rapport, which shall be further discussed in the next subchapter.

⁷ Beauchamp TL, Childress JF. *Veracity Respect for autonomy*. Principles of Biomedical Ethics. 5th ed. Oxford University Press. New York. 2001; pp. 283-292.

We need the human touch in medicine after AI implementation

A possible crisis in a close patient–doctor relationship

The time when physicians were held in the highest esteem is possibly when people addressed them as “doctor” to express their respect for their remarkable skills, well beyond a layperson’s comprehension. During the past decades, respect for doctors has been decreasing to a certain extent, with differences from country to country.

Many social factors and/or health care systems are probably the cause of this erosion in the doctor–patient relationship. Additionally, the advent of AI can also potentially alter this traditional relationship. As mentioned in the previous subchapter, doctors are losing their monopoly over medical skills, and patients may not respect doctors as much as they used to. Considering the professional responsibility which is part of the medical profession, erosion of the close relationship between doctors and patients is a crisis which cannot be ignored. Doctors must be aware of the impact of this crisis and look for ways to avoid such a disaster.

Personable doctors

To continue to be respected and trusted by their patients even after AI prevails in health care systems, doctors will need to be more approachable and professional than ever before.

It is well known that personable doctors are highly respected and trusted by their patients. How should we then define this trait? How can doctors become personable? This differs depending on time and place; therefore, any attempt at generalizing this definition seems to be rather pointless. Instead, it is more appropriate to look at a couple of specific examples on how doctors acquire implicit patient trust.

Dr. Yamagishi is an eye doctor in a small village in the remote mountain village of Nara, Japan. His family has been in private medical practice for generations in this small community, and he is considered by his patients a good neighbor. His personal history



is of paramount importance in establishing rapport with patients. See (1) in BOX 10.

Another example is Dr. Motonaga, who is a physician in a base hospital on an isolated small island, Miyakojima, far away from mainland Japan. He can speak the local dialect, which is highly valued by patients in his community. I am convinced that speaking the same dialect means that a cultural understanding is shared, and helps lower the guard of patients, resulting in a very close relationship between doctor and patient. See (2) in BOX 10.

The Showa University School of Medicine in Japan is very impressive in terms of its professionalism. Since its foundation about one hundred years ago, the school has centered its medical training on “confronting everything with a sincere heart”. The slogan “Shisei Ikkān”, in Japanese, was proposed by the founder of this institution.⁸ He was aware of the core ethics of health care.

As a consequence, graduates from this medical school are aware of the importance of being extremely sincere in performing their duties as doctors. It is clear that this professionalism has helped doctors build ideal relationships with patients and taking professional responsibility. See (3) in BOX 10.

BOX 10 Implicit trust in doctors

- (1) being a good neighbor for many generations
- (2) local culture sharing, thus causing patients to lower their guard
- (3) “Shisei Ikkān”, confronting everything with a sincere heart



⁸ Web site of Showa University. (Accessed on 01.06.2020 at http://www.showa-u.ac.jp/about_us/mission/establishment.html).

Face-to-face communication for establishing and maintaining rapport

Laypersons tend to think medicine is science based, but when it is provided in the form of health care, it seems to be based on social relationships in which the human touch plays a major role. Traditionally, medical communication was performed face-to-face and both verbal information and non-verbal cues are reciprocally transferred between doctors and patients. This is the basis on which rapport is established and maintained.

In the course of the past decades, many diseases have become curable thanks to the introduction in medicine of highly specialized machines. The achievements made possible by these incredible machines should be highly appreciated, but patients visiting modern hospitals often become confused trying to understand doctors, because there are too many machines between them. The examination of patients is often carried out without any physical contact and results are interspersed with figures out of blood tests. This may sometimes cause misunderstandings or misinterpretation, with an adverse effect on rapport. In the worst case scenario, patients challenge and sue doctors.

In order to promote good relationships, the time has come to go back to the good old days by putting machines aside in order to regain sufficient human touch. It is noteworthy that after AI implementation, the user will be afforded more time. Doctors should use this gift of time to spend more hours on face-to-face communication with difficult cases. This type of human-centric health care is the first step for doctors to become more approachable. We should not overlook the opportunity for reconstructing close relationships that AI implementation has brought about by freeing up time.

Conclusion

AI technology brings great benefits to the health care system. However, a possible main adverse effect is represented by the erosion of the doctor-patient relationship, as a result of doctors' losing monopoly over medical skills. To maintain rapport with patients and provide them with the best health care, doctors have to incorporate the human touch to improve communication with patients.

Artificial Intelligence in Oncology

Alexandru G. Floares*

Artificial Intelligence (AI) has seen an explosive evolution in the last decade, and the first medical applications, outperforming human experts, have since appeared. It is merely the beginning of a revolution, without precedent in human history, that will radically change the whole of humanity. Moreover, the integration of Digital Automation with Artificial Intelligence has resulted in a new field - *Intelligent Automation*. Due to the COVID-19 pandemic, we expect this field's evolution to be highly accelerated, especially its digital automation component. So, we are entering the Era of *Augmented Intelligence Automation*, where AI is enhancing, not replacing human intelligence. It became clear that physicians and AI working together as a team outperform both physicians and AI working alone.

The urge to adopt Intelligent Automation in hospitals has several motivations. Considering the very high total cost of the European healthcare systems, the analysis of its containment is relevant.¹ Two important facts are the progressive aging of the population and the increase of chronic diseases (e.g., cancer). Chronic conditions account for 75% of health care costs, and the impact of chronic illness will grow over time.^{2,3} Aging societies represent a significant challenge for healthcare systems worldwide, not only due to the increase in the number of older people⁴ but also because older people tend to be more physically

* President Solution of Artificial Intelligence Applications (SAIA), CEO Artificial Intelligence Expert & ONCOPREDICT (Romania).

¹ Mossialos E., Le Grand J. *Health care and cost containment in the European Union*. Routledge, London (Eds.). 2019.

² Parks A.C., Williams A.L., Kackloudis G.M. et Al. *The Effects of a Digital Well-Being Intervention on Patients With Chronic Conditions: Observational Study*. *Journal of Medical Internet Research*, 2020, 22(1), e16211.

³ National Center for Chronic Disease Prevention and Health Promotion. Centers for Disease Control and Prevention. *The Power of Prevention: Chronic Disease...The Public Health Challenge of the 21st Century*. Atlanta, GA; 2009. (Accessed on 03.01.2020 at: <https://www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf>).

⁴ Huang S., Yang J., Fong S., Zhao Q. *Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges*. *Cancer Letters*, Elsevier, 2019, 471:61-71.

inactive. Thus, the economic costs of inactivity are likely to increase significantly.⁵

Wrong or delayed diagnoses and the consequent inappropriate prescriptions and therapies, unnecessary diagnostic tests, and unwanted patient outcomes significantly contribute to healthcare costs. Diagnostic errors make it difficult to provide effective, patient-centered, high-standard healthcare while maintaining at the same time cost-effectiveness.⁶

One of the most relevant causes of healthcare malpractices and the consequent lawsuits is the failure of diagnosis,⁷ further increasing total expenditure of the healthcare. Chronic and widespread diseases have high impacts on healthcare and social systems, and always require innovative strategies. Among these diseases, cancers are one of the most harmful, but a correct and prompt diagnosis can dramatically increase treatment success rate and decrease the costs.

Another set of reasons, revealed by the recent COVID-19 pandemic, is represented by the imperious necessity to integrate the healthcare resources distributed through the territory, better exploitation and coordination, transparency for consumers, cost reduction, and sharing of outcome data.⁸ The existing healthcare systems cannot fully integrate all their distributed resources, causing a waste of money and worse healthcare services.⁹ They must be able to quickly coordinate and rearrange all their resources, defining healthcare pathways that start from General Practitioners (GPs) and following patients through hospitalization, clinic visits, and home follow-ups, expanding their services beyond the hospital facilities. Thus, the patients' pathways, especially for chronic diseases such as cancer, are distributed among different organizations, and so is the biomedical workflow, including multiple medical actors.

This workflow is far from optimal. However, its digital automation is the first step in facilitating optimization. For example, using Robotic Process Automation (RPA), we can model, analyze, and optimize the

⁵ Friedberg M.W., Hussey P.S., Schneider, E.C. *Primary care: a critical review of the evidence on quality and costs of health care*. Health Affairs, 2010, 29(5), 766-772.

⁶ Chaudhary M.A.I., Nisar A. *Escalating Health Care Cost due to Unnecessary Diagnostic Testing*. Mehran University Research Journal of Engineering and Technology, 2017, 36(3), 569-578.

⁷ Medscape, *Medscape Family Physician Malpractice Report 2019*, New York, NY, 2019. (Accessed on 03.01.2020 at: https://www.medscape.com/slideshow/2019-m_alpractice-report-fm-6012446#3).

⁸ Baffert S., Hoang H.L., Brédart A. et Al. *The patient-breast cancer care pathway: how could it be optimized?* BMC Cancer, 2015, 15(1), 394.

⁹ Bentley T.G., Effros R.M., Palar K. et Al. *Waste in the US health care system: a conceptual framework*. The Milbank Quarterly, 2008, 86(4), 629-659.

biomedical workflow. In this digital framework, each biomedical actor is assisted by a digital twin, e.g., a digital oncologist, radiologist, pathologist, etc., and AI can be embedded instead of only applied to specific patient data. By considering the hospital as the center of a medical organizations' ecosystem, working together by using a digital distributed workflow, where AI is embedded, we reach the new *Smart Distributed Hospital* concept. The application of this AI-based *Smart Distributed Hospital* paradigm provides better care - prompt and accurate diagnoses and personalized treatment, and dramatically reduce the costs. To this end, current state-of-the-art AI methodologies require an integration and customization process.

Cost containment must be coupled with an improvement in the new smart distributed hospital approach's provided services. The improvement process is manifold and must embrace different aspects. A commonly recognized issue for physicians is related to best practices and medical guidelines: it is tough to read hundreds of PDF documents pages to discover the rules embedded in these documents. It is likewise challenging to memorize and retrieve them when visiting a patient, maybe in an emergency. Consequently, adherence to best practice is lower than expected, further increasing the medical error rate. US studies¹⁰ show that between 250,000 and 400,000 hospitalized patients each year experience preventable harm, and medical errors cost approximately 20 billion USD a year. Supporting the physicians in the exploitation of clinical guidelines – via digital automation of the best practice rules – will not only improve primary and secondary healthcare, but will also allow physicians to spend more time on interacting with their patients humanely. The automated best practice rules and workflow will act as a recommender system, suggesting to the physicians the best actions or decisions, tailored to an individual patient, during the doctor visit, with no need for interruptions. Thus, adherence to best practice will dramatically increase, with all the benefits for the patients and physicians.

Moreover, the recommender system can use a *conversational agent interface*, which is the most human-like interface. Conversational agents understand medical language and can make their recommendations in a pleasant and explanatory way. AI systems for diagnosis, prognosis, and response to treatment prediction, embedded in the automated clinical

¹⁰ CNBC. *The third-leading cause of death in US most doctors don't want you to know about*. 2018. (Accessed on 03.30.2020 at: <https://www.cnbc.com/2018/02/22/medical-errors-third-leading-cause-of-death-in-america.html>).

workflow, will follow the entire life cycle of the medical tests. First, they will learn from patients' data, stored in the Electronic Medical Records software database, in an automated way, as RPA selects and prepares the proper inputs and outputs for them. After choosing a particular research theme, AI will automatically research it in parallel, with the physicians completing their routine daily work. After learning a predictive model from data and validation, the AI system will become a Clinical Decision Support System embedded in the workflow. We expect that following this Intelligent Automation roadmap will dramatically improve healthcare quality, as best practice will be easy to follow, and AI will boost precision and reduce errors. Physicians will have more time for human interactions with their patients, as patients now tend to be seen as collections of clinical, lab data, and images instead of human beings.

We have been addressing these problems and many others in a recent project proposal, together with a large European consortium (Consiglio Nazionale delle Ricerche, Engineering Ingegneria Informatica SpA, Sphynx Technology Solutions AG, Fundatia Ana Aslan International, UNINOVA - Instituto de Desenvolvimento de Novas Tecnologias-Associacao, Knowledgebiz Consulting-Sociedade de Consultoria em Gestao LDA, Telecommunication Systems Institute, Università degli Studi di Milano, DEN Institute, Aegis IT Research GMBH, Privanova SAS, Istituto Nazionale Tumori "Fondazione Pascale", Klinikum Nürnberg, Faculty of Medicine, University of Belgrade), within the context of the EU projects competition "AI for the Smart Hospital of the Future".

We believe that both Information and Communication Technologies and biomedical and societal mindsets are ready for the revolution mentioned above. Information Technology facilitated the rise of modern medicine by its profound impact on medical imaging and molecular biology. Both fields produced vast amounts of high-throughput data. We hoped to find answers to important biomedical questions, but we have just began to formulate more meaningful questions. These problems should be reframed as Data Science (data-driven) problems and solved with AI, instead of the conventional hypothesis-driven and statistics approach. To be more specific, let us focus on cancer, which is a significant health problem.

Globally, more than 8 million people die from cancer every year, but early-detected cancers can be cured. However, the existing tests are mainly invasive (surgical procedures) and for later stages or non-inva-

sive but with deceptively low accuracy for early stages. Furthermore, cancer risk increases with age.

Some imaging methods could be used for early cancer detection, e.g., mammography for breast cancer. As most new powerful AI algorithms were developed in computer vision, it is not surprising that some superhuman AI applications were proposed for medical imaging analysis. However, the molecular alterations related to cancer development precede the formation of a tumor size detectable through imaging techniques.

To develop molecular cancer tests, tissue biopsies could be replaced by "liquid biopsies," e.g., a blood drop. We started realizing that blood circulation is like a *liquid nervous system*. These *non-invasive* tests make it possible to reduce fear, pain, and risks for patients.

While it is clear that the best strategy is to exploit the complementarities between Human and Artificial Intelligence and cooperate, this is not easy to implement and prone to mistakes. For example, a common mistake is to use biomedical knowledge to select the subset of relevant molecular alterations from Big Data or impose a model on the data. Instead, let data speak to AI first (and not to us!) and then use knowledge to interpret AI findings. Using this strategy, we can develop highly accurate predictive models - molecular tests for diagnosis, prognosis, or response to treatment prediction.

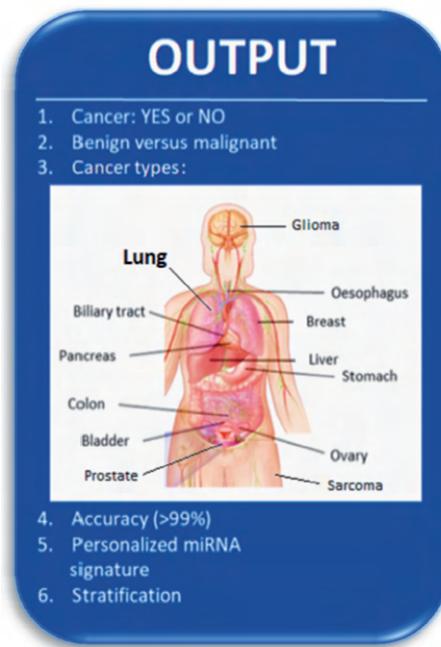
These tests should satisfy what we called the ART criteria:

1. Highly Accurate, with performance > 95%
2. Robust, having similar accuracy for different groups of patients
3. Transparent instead of "black-box."

For illustrative purposes, we will shortly present our AI-based non-invasive multi-cancer diagnosis and early detection test. To our knowledge, it is the best, working on thirteen cancer types with an accuracy greater than 99% (see, for example¹¹). Starting from a single drop of blood ("a liquid biopsy") we obtain an output discriminating between cancer, benign and malignant, working on multiple cancers as shown in *Figure 1*.

¹¹ Cordis Europa, *Artificial Intelligence System for Multi-Cancer Detection Support*. 2019. (Accessed on 08.30.2020 at: <https://cordis.europa.eu/article/id/411554-artificial-intelligence-enhances-cancer-diagnostic-testing>).

Figure 1. The output of the multi-cancer test



Compared to the competition, our test works on more cancer types, it has the highest median and lowest accuracy, and was tested on a higher number of cases, as shown in *Table 1*.

Table 1. Comparison with the competition

Company	Cancer types	Median accuracy	Lowest accuracy	Number of cases
AIE	13	>99%	99%	>6000
CancerSEEK	8	~77%	33%	~1000
Delfi	7	~73%	57%	208

AI medical applications pose new and complex ethical problems. If clearly and pragmatically formulated, they can be more or less easily solved. However, the benefits for patients, physicians, and healthcare systems could be so great that it is more unethical not to use AI to revolutionize medicine than to do so.

Artificial Intelligence in the road of Health for All. Perils and Hope

Felix Hector Rigoli*

"If we just let machines learn ethics by observing and emulating us, they will learn to do lots of unethical things. So maybe AI will force us to confront what we really mean by ethics before we can decide how we want AIs to be ethical." Pedro Domingos

"Our ethical evolution still lags behind the technological revolution"
Virginia Eubanks

Introduction

Most societies in Latin America and in other regions of the world have made progress towards health systems available for all people. This march towards health as a right for all is somewhat contrasted by other trends that seem to be fostering inequities and exclusion in most societies. The gradual advances using artificial intelligence in many aspects of health services enable to expand the reach and benefits of knowledge and cure. Processes such as epidemiological surveillance, tele-care and the use of best evidence in clinical algorithms seem to be growingly positive. At the same time a disquieting number of studies begin to show how this potential is also an amplifier of biased policies. As a case in point, automating decisions and processes in health systems may reflect the trends towards exclusion and discrimination, or alternatively, serve as a tool for facilitating and improving access of the vulnerable population both for care and prevention.

For the purposes of this article, we will use the terms artificial intelligence (AI), algorithmic decision systems (ADS) and other related terms to describe a set of processes that substitute decisions formerly made by humans, through the use of machine data processing.

As the mainstream media is prolific in describing the virtues of all related to artificial intelligence, the article begins by a tour of a few perils that are part of these advances in healthcare and other related social

* Guest Lecturer and Researcher in the University of São Paulo and Oswaldo Cruz Foundation (Brazil).

services, approaching the process from an equal rights, equal access to healthcare point of view.

As a second step, the article reviews the many facets in which these new tools may expand the capacities to get closer to a Health for All ideal, both in developed and developing contexts.

The final section makes a summary of current attempts of creating analytical frameworks to dissect artificial intelligence systems, that may orient them to a common good or at least to do no harm.

The perils of Artificial Intelligence in healthcare

The gradual advances in artificial intelligence in many aspects of daily life are particularly noticeable in administrative processes, both in the public and private spheres. Staff hiring selection, screening of candidates for university vacancies, categorization of citizens as eligible for social programs, and even health benefits,¹ are basically done with databases and processing algorithms. Most insurance firms (health and others) use some automated process that filters the procedures that will be approved or denied.

Even though automation is praised as a panacea, reports and studies begin to surface showing that many of these systems are based on algorithms that replicate and amplify prejudices and assumptions that may be present in the minds of formulators, or invisible even to the managers who command their design.

Artificial intelligence may have a neutral algorithm, but it is usually trained with the existing data that may be biased: Black patients in the US are consistently less likely to get pain medication.² The same happens with pain treatment during labour in Black women in Brazil, as they have 40% less chances in getting anaesthesia during episiotomy.³ So if models are trained with these data, they may assume that Black people are less prone to suffer from pain (a chilling memory of slavery).

¹ Lecher C. *What happens when an algorithm cuts your health care*. The Verge. Mar. 21, 2018. (Accessed on 07.08.2020 at: <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>).

² Anderson KO, Green CR, Payne R. *Racial and ethnic disparities in pain: Causes and consequences of unequal care*. J Pain. 2009; 10 (12): 1187-1204.

³ Leal MdC. et Al. *The color of pain: racial inequities in prenatal care and childbirth in Brazil*. Cad. Saúde Pública [online]. 2017, vol. 33, suppl. 1.

In the University of Chicago, a machine learning process was used to identify patients who were most likely to be discharged early from a hospital, in order to give them special assistance to remove barriers that could prevent them from leaving the hospital when they were ready.⁴ The designers developed the algorithms using the clinical database, but in a second stage they discovered that adding the postal code of the patient residence improved the predictive capacity of the model to identify patients with shorter lengths of stay. By adding a postal code, the algorithm was indirectly classifying patients by socioeconomic status and, as poor patients have worse home conditions, they were more likely to have longer lengths of stay. In that way, the Algorithm Decision System avoided supporting the more socially at-risk population who really should be the ones that receive more help, channelling resources to provide additional support to a predominantly educated and affluent population.

In spite of the fact that the automated systems reveal and amplify the non-declared preferences of those who contract their design, some of the deleterious effects of algorithms may be hidden behind the commercial secret formulae or “secret sauce”.⁵ It has proved difficult, even for governments, to force designers to disclose their algorithms whenever the results diverge from what was expected. As financial applications are the frontline areas for the use of AI, many commercial artificial intelligence derives from the earlier robot-tax advisers and robot-traders, frequently instructed to be marginally tax avoiders or market manipulators.⁶ Algorithmic Decision Systems vendors, whenever are found to be faulty, migrate their defective systems to new agencies or areas of focus, rather than addressing the fundamental concerns they have created. After being criticised by developed countries’ governments, the AI designers move to countries with less expertise and fewer legal accountability tools.⁷

⁴ Nordling L. *Without careful implementation, artificial intelligence could widen health-care inequality*. Nature, September 25, 2019.

⁵ Richardson R, Schultz J, Southerland V. *Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems*. AI Now Institute, September 2019. (Accessed on 07.08.2020 at: <https://ainowinstitute.org/litigatingalgorithms-2019-us.html>).

⁶ Seyfert R. *Bugs, predations or manipulations? Incompatible epistemic regimes of high-frequency trading*. Economy and Society Sep 20, 2016.

⁷ Richardson. *Litigating Algorithms...*

Most frequently, artificial intelligence applications in public services have as goals downsizing governments and using technical answers to policy and political problems. An inescapable feature is to downsize human labour even if humans are cheaper, like the automated supermarket cashiers.⁸ A 2017 survey in different countries showed that 20 percent of the businesses have already used artificial intelligence to replace or avoid recruiting new workers.⁹ When austerity programs are being implemented, and especially in healthcare and social services, it is usual to employ artificial intelligence to select and control the benefits that are handed out. Usually, the objectives stated to coders are to reduce fraud and overuse. If those are the real motives, in algorithmic terms the ADS will have as a priority task to eliminate false positives in order to avoid all cases that are not fully fitting the definitions, even at the risk of having some benefits wrongly withheld. As social systems have inherently hazy boundaries, false positives are needed to avoid false negatives and vice-versa.¹⁰ The algorithm does what is asked to do and that is potentially dangerous, so be careful with what you wish for!

Digital decision-making has become commonplace in policing, marketing, credit, criminal sentencing, management and public programs. While these systems are used first in what Eubanks¹¹ call “low rights environments” where there are low expectations of political accountability and transparency (such as in programs targeted for the disadvantaged groups), they may act in synergy with the expansion of reduced-democracy systems and eventually affect us all.

Automated decision-making hides human problems from the professional elites, giving the policy makers a shield against life and death decisions, in the same way that cyber-warriors may kill civil citizens with drones from an office chair. Automation shifts the human approach in health and social services from “case” to “protocol” or “task” as defined in the algorithm. At a certain point, the human rela-

⁸ Merchant B. *Why self-checkout is and has always been the worst*. Gizmodo. March 7, 2019.

⁹ The State of AI: Artificial Intelligence in Business. Verdict 2017. (Accessed on 07.08.2020 at: <https://verdict-ai.nridigital.com/issue-one/state-of-ai-artificial-intelligence-business-data?forced>).

¹⁰ University of Chicago. Aequitas Project. (Accessed on 07.08.2020 at: <http://www.datasciencepublicpolicy.org/projects/aequitas/>).

¹¹ Eubanks V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: Picador, St Martin's Press. 2018 p. 12.

tion involving both the patient and the practitioner vanishes, and the protocol on the screen guides the whole process. The humanity of the person has become transparent; the reality is in the screen.¹²

The main feature of all artificial intelligence/algorithms applied to healthcare is the availability of hundreds of data points for each individual in large populations. Much of these data will remain in stock, with no present use but available for future uses, including hacking or selling. Even if not on purpose, there is an implicit profiling in gathering population data when there are multi-tiered health systems. Part of that profiling may consist in targeting actions for the poor. Paradoxically, it can be harmful for the middle class: the Health of the Family program in São Paulo is forbidden from entering the high-rise condominiums; therefore health authorities don't know the prevalence of public health problems, from domestic violence or child abuse to drug abuse or mosquito infestation in the swimming pools.¹³ We tend to associate these problems to poverty, due to the fact that we can compute health problems only if the poor cannot resist the state intrusion in their lives. Blanket access to personal intimacy only seems normal when something (poverty or illness) is equalled with criminality. No one may enter in a home looking for evidence of paying late a credit card, but they may enter into a favela house looking for teenage pregnant girls.¹⁴

Another omnipresent feature in big data and artificial intelligence is the non-humanly large classificatory power of human traits, which lends itself to dangerous consequences¹⁵ as it is essentially a profiling tool. Using AI to predict patterns makes profiling persistent, even intergenerational. The mark of "family history of drug abuse" will stay in the family medical records for a long time, and even go from medical records to criminal records. In this context we need to understand the "right to be forgotten" in European digital rights law.¹⁶ We need to encourage systems that forget everything that is justifiably not needed

¹² Gawande A. *Why doctors hate their computers*. The New Yorker, Nov 5, 2019.

¹³ Rosa T. *Saúde. Para onde vai a nova classe média*. CONASS. Ed. 7 April 2013.

¹⁴ Eubanks. *Automating Inequality*... p. 197.

¹⁵ Black E. *IBM and the Holocaust: The Strategic Alliance between Nazi Germany and America's Most Powerful Corporation*. Crown Books, 2012 pp. 8-12.

¹⁶ European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General*

for the future. The digital profiling has low barriers to expand and is difficult to eliminate, being relatively invisible to the public.

The algorithms populated with existing data will reflect social inequalities and they risk perpetuating systemic injustice, unless design embeds countervailing measures.¹⁷ If the designers make an algorithm for diagnosis, treatment and coverage contracted by an insurance plan, or funded by a pharmaceutical company, the resultant system will reflect and produce the outputs that were prominent in the funders' motivations. It is possible, for instance, that automating hospital services through AI-driven triage systems caters to the financial interests of hospitals (by rationing resource allocation), while failing to meet the expectations of severely ill patients in terms of access to care.¹⁸

Socially disadvantaged populations present certain negative health outcomes due to well-known social deficits. Algorithms that make their predictions based on health outcomes alone, without factoring in their social causes, can result in significant harm and increased health inequalities. For example, if poor or less educated people have performed worse after certain health interventions (due to occupational risks, no access to care or environmental factors) an algorithm can determine that people with those characteristics will always perform worse and recommend that they are not offered the intervention in the first place.

Several cases in Latin America are examples of how the use of algorithms may amplify existing inequities, even under the assumption of universal coverage. In Chile, the national health system guarantees a set of equally available procedures. In spite of the technical neutrality of these algorithms, the organizations that provide services under contracts with the Ministry of Health seem uninterested in accepting the female members, as they will experience higher costs derived from maternity events.¹⁹ An analysis of the principles of the algorithms based on the clinical guidelines used to list the diseases to be covered, showed that

Data Protection Regulation). (Accessed on 07.08.2020 at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504>).

¹⁷ Benkler Y. *Don't let the industry write the rules for AI*. Nature 569(7755): 161, 2019 05.

¹⁸ Blasimme A, Vayena E. *The Ethics of AI in Biomedical Research, Patient Care and Public Health in Oxford Handbook of Ethics of Artificial Intelligence*. Preprint; 2019 (Accessed on 07.08.2020 at: <https://doi.org/10.2139/ssrn.3368756>).

¹⁹ Universidad Diego Portales. *El derecho a la salud en el Plan Auge. Informe anual sobre derechos humanos en Chile*. 2007.

the efficiency criteria used tilted the decisions, harming the equity in the resource allocation.²⁰

The Uruguayan National pooled Fund (Fondo Nacional de Recursos -FNR) covers all the population for high cost/low frequency procedures using an algorithm linked to effectiveness and cost, therefore eliminating the bias due to the financial capacity of those affected. In spite of this, the studies produced in the last decade have discovered that in order to ensure equity, it is not enough to remove the financial barriers.²¹ As has been shown, the high-cost/low frequency procedures are the end in a pipeline of care pathway, and an algorithm that does not include social conditions will bias the use of resources towards those living in the main city and with a higher socioeconomic status.²²

Disparities in access to care and attainment of good health outcomes may become exponentially larger, and more importantly, such disparity will be less visible “because the decision will bear the authoritative objectivity often attributed to numbers and that is typically expected from automated decision-making tools.”²³

There are more long-term concerns regarding how biological big data may affect the future of research. There is evidence that data firms could offer patients or governments in poor countries to hand over databanks in return for medical care or financial reward. Such practices could create a privacy divide between rich and poor populations adding one more divide to separate different socio-economic groups. Companies offer to pay for such data, tempting low-income countries' national health systems or individual researchers to part with patient data, perhaps without thinking about the rights of those whose data they are sharing. There have been some initial efforts to address this, including the Data Sharing Principles in Developing Countries, formulated in Nairobi in 2014,²⁴ and the underlying concept inspiring the

²⁰ Zuniga-Fajuri, A. *Justicia y racionamiento sanitario en el Plan AUGE: dilemas bioéticos asociados a la distribución de recursos escasos*. Acta bioeth. [online]. 2011, vol. 17, n. 1, pp. 73-84.

²¹ González C, Triunfo P. *Inequidad en el acceso a los servicios de salud en Uruguay* in Documentos de Trabajo (working papers) 0718, Department of Economics – dECON; 2018.

²² Rodriguez A. *Cáncer colo-rectal avanzado Evaluación del tratamiento con Bevacizumab. MAC de la Farmacología Individual a la Colectiva*. November 2015. (Accessed on 07.08.2020 at: at: <http://www.farmacologia.hc.edu.uy/images/Alarico.pdf>).

²³ Blasimme. *The Ethics of AI...*

²⁴ *Data Sharing Principles in Developing Countries* (The Nairobi Data Sharing Principles) 2014.

development of these principles is that data collected through public efforts should be handled as a public good.

A pervasive way in which artificial intelligence may be influencing healthcare against the public interest, is the capacity of creating and diffusing clinical algorithms, tilting the behaviour of clinicians in an opaque way. A few years ago, the consultancy firm McKinsey issued a report called *The Road to Digital Success in Pharma*: “We envisage a world in which most care is “protocolized” – that is, in which clinical decisions on the best treatment options are suggested to physicians by an automated decision algorithm informed by advanced analytics. In this environment, winning pharmaceutical companies will be those able to influence the algorithm.”²⁵ The pharmaceutical industry seems to have taken good notice of the advice. More and more research reveals a shift in pharmaceutical marketing, which moved from catering individual prescribers towards influencing the authors of clinical guidelines that will be the basis for algorithm designers. As a second and complementary step, they are partnering with startups that embed the not-neutral clinical guidelines in algorithms that become a part of the hospital clinical management routines.^{26,27,28}

An early study in Canada²⁹ found that 59% of the authors of 37 clinical guidelines had received financial support from the pharmaceutical industry that produced the drugs mentioned in the guidelines. A cross-sectional analysis of the United States Open Payment database in 2016 showed that 86% of 125 authors of cancer guidelines had financial conflicts of interest, including 84% who accepted general payments and 47% who accepted research payments.³⁰ Another study in Japan, 2016 calculated that in six prominent oncology guidelines, of 326 eligible authors, 78.2% received payments from pharmaceutical

²⁵ Champagne D, Hung A, Leclerc O. *The Road to Digital Success in Pharma*. McKinsey Pharmaceutical and Medical Products, August 2015.

²⁶ Bulik B. *Is there a place for pharma in the emerging EHR Market?* FiercePharma News. May 11, 2015.

²⁷ Constantia Flexibles Newsletter. *Pharma and HPC Insights*. January 2018.

²⁸ McKinsey. *Real-world evidence: From activity to impact in healthcare decision making*. May 2018.

²⁹ Chudhry N, Stelfox H, Detsky A. *Relationships Between Authors of Clinical Practice Guidelines and the Pharmaceutical Industry* JAMA. 2002; 287 (5): 612-617. doi:10.1001/jama.287.5.612.

³⁰ Mitchell A, Basch E, Dusetzina S. *Financial Relationships With Industry Among National Comprehensive Cancer Network Guideline Authors*. JAMA Oncol. 2016; 2 (12): 1628-1631.

companies, with a very unclear or nonexistent disclosure of conflicts of interest.³¹ The clinical guidelines concerning the 10 top revenue drugs in the United States in 2016 were analyzed, finding that 57% of the authors had financial conflicts of interest related to the manufacturers of the drug under study and 25.6% were found to have received but not disclosed payments from companies marketing one of the 10 high-revenue medications recommended.³² A German study regarding a dermatological guideline concluded that 10 out of 15 voting members in the committee had received money from the company that produced the drug under study.³³ Dermatology guideline authors in the US received payments ranging from ten thousand to 100 000 American dollars from the industry in 40 out of 49 cases studied between 2013 and 2016.³⁴ A larger study of guidelines covering 10 disease categories in Australia published in 2019 found that 70% of them included at least one author with a potentially relevant undisclosed tie to the pharmaceutical company interested in the health condition, while writers of guidelines developed and funded by governments were less likely to have undisclosed financial ties.³⁵

Promises and hopes: how can Artificial Intelligence contribute to achieve Health for All

Artificial Intelligence holds tremendous promises for transforming the provision of healthcare services in resource-poor settings. Many of the health systems bottlenecks in such environments could be addressed and overcome using AI supported by other technological developments. The use of smartphones, combined with supporting

³¹ Saito H, Ozaki A, Sawano T et Al. *Evaluation of Pharmaceutical Company Payments and Conflict of Interest Disclosures Among Oncology Clinical Practice Guideline Authors in Japan*. JAMA Network Open. 2019; 2 (4): e192834.

³² Khan R, Scuffidi A, Rumman A et Al. *Prevalence of Financial Conflicts of Interest Among Authors of Clinical Guidelines Related to High-Revenue Medications*. JAMA Intern Med. 2018; 178 (12): 1712-1715.

³³ Schott, G; Dünnweber, C; Mühlbauer, et Al. *Does the Pharmaceutical Industry Influence Guidelines? Two Examples From Germany*. Dtsch Arztebl Int. [online] 2013 Sep; 110 (35-36): 575-583.

³⁴ Checketts J, Sims M, Vassar M. *Evaluating Industry Payments Among Dermatology Clinical Practice Guidelines Authors*. JAMA Dermatol. 2017; 153 (12): 1229-1235.

³⁵ Moynihan R, Lai A, Jarvis H, et Al. *Undisclosed financial ties between guideline writers and pharmaceutical companies: a cross-sectional study across 10 disease categories*. BMJ Open 2019; 9:e025864. doi: 10.1136/bmjopen-2018-025864.

technologies such as mobile Health, Electronic Medical Records and cloud computing^{36,37} provides ample opportunities to deliver better quality services in hard-to-reach areas, even in places where health professionals are scarce.

Algorithms and big data techniques, jointly with the explosion of social networks allow public health authorities to consolidate global trends in communicable diseases³⁸ and public health emergencies, using pieces of information such as Google hits and social media posts as early markers of outbreaks, as well as to estimate disease incidence, even when the public authorities are not communicating the cases.³⁹ The recent explosion of false claims and biased use of targeted messages, and the growing trends in state-controlled social media are unfortunately raising doubts regarding the validity of these methodologies. Taken the appropriate caveats, social media analytics seem to be an important tool to complement other initiatives in massive emergency events.⁴⁰ Even though these emergency events disrupt infrastructure systems, different artificial intelligence devices may overcome these failures.^{41, 42}

The analysis of health status both at the individual and population levels is affected by several social parameters (e.g. income, education, dietary habits, environmental factors, community context), which are not restricted to the healthcare systems boundaries. The use of large amounts of data may help to understand specific effects and interactions

³⁶ Santos A, Abreu M, Melo M et Al. *Development of telehealth services in Latin America: the current situation*. In: UNDP Health policy in emerging economies: innovations and challenges. Volume 13, Issue No. 1 June 2016 p.50-54. (Accessed on 07.08.2020 at: https://ipcg.org/pub/eng/PIF35_Health_policy_in_emerging_economies_innovations_and_challenges.pdf).

³⁷ Campanella N, Wright H, Morosini P et Al.: *Proceedings and Quality Indicators of the Primary Health Care Doctor Supporting Medical Teleconsultation System in the State of Amazonas (Brazil)*. Diversity and Equality in Health and Care. 2017; 14 (5): 227-235.

³⁸ Wahl B, Cossy-Gantner A, Germann S et Al. *Artificial intelligence (AI) and global health: how can AI contribute to health in resource poor settings?* BMJ Glob Health 2018; 3: e000798. doi:10.1136/bmjgh-2018-000798.

³⁹ Fung I, Tse Z, Fu K-W. *The use of social media in public health surveillance*. WHO Western Pacific Surveillance and Response Journal. Issue 2, June 2015 3-6.

⁴⁰ Simon T, Goldberg A, Adini B. *Socializing in emergencies—A review of the use of social media in emergency situations*. International Journal of Information Management, 35 (5) 2015, pp. 609-619.

⁴¹ Singh J, Dwivedi Y, Rana N et Al. *Event classification and location prediction from tweets during disaster*. Ann Oper Res 283. 2017 737-757.

⁴² Rahmani, D. *Designing a robust and dynamic network for the emergency blood supply chain with the risk of disruptions*. Ann Oper Res 283. 2019 613-641.

between health and various social conditions, leading to the development of more effective and efficient public health programs.^{43,44}

For achieving better use of Artificial Intelligence, we must place improving the health of everyone as the explicit goal and then build the systems toward this goal. The late Fitz Mullan demanded that "Medicine should move away from advances that put more years in the lives of the privileged, and concentrate its efforts to elevate the floor of life expectancy" for the many.⁴⁵ Systems should be designed and tested in specific steps where there's a chance for self-reflection: Is what we are doing advancing equity for everyone, or have we unintentionally worsened things?

A warning that those that commission or design algorithms need to be acutely aware of, is that software reflects the choices of the people who write it, so it is important to have fairness issues in mind. These choices reflect the priorities of the algorithm designers in three key components: outcome variables, predictive variables and validation data.⁴⁶ During the process there should be enough care to check if the algorithm is going to lead to an unfair result, as predictably it might. Which bad things unintentionally could happen and how may they be proactively avoided, given the built-in caveats in the design to benefit everyone?

Such avoidance requires careful attention to each step: picking the data, developing the formula, and then deploying the algorithm and monitoring how it is used. To introduce timely ethical concerns and tracking mechanisms concerning the social impacts of non-human-implemented actions may unleash the potential of such devices to expand the benefits of healthcare and prevention. In the same fashion that drugs have side effects, the first test for an algorithm or a drug is related to safety.⁴⁷ In March 2019, the UK National Institute for Clinical Excellence

⁴³ Shaban-Nejad A, Michalowski M, Buckeridge, D.L. *Health intelligence: how artificial intelligence transforms population and personalized health.* NPJ Digital Med 1, 53, 2018.

⁴⁴ Inter-American Development Bank: *Costa Rican Household Poverty Level Prediction.* (Accessed on 07.08.2020 at: <https://www.kaggle.com/c/costa-rican-household-poverty-prediction>).

⁴⁵ Genzlinger N. *Fitz Mullan obituary,* New York Times. (Accessed on 07.08.2020 at: <https://www.nytimes.com/2019/12/10/health/fitzhugh-mullan-foe-of-health-care-disparities-dies-at-77.html>).

⁴⁶ Eubanks. *Automating Inequality...* p. 143.

⁴⁷ Coravos A, Chen I, Gordhandas A, et Al. *We should treat algorithms like prescription drugs.* Quartz. February 14, 2019. (Accessed on 07.08.2020 at: <https://qz.com/1540594/treating-algorithms-like-prescription-drugs-could-reduce-ai-bias/>).

(NICE) published a set of evidence standards for digital health that are going to be used to check safety, effectiveness and reduction of inequalities, to be applied to new health algorithms.⁴⁸

Addressing equity in AI is not an afterthought but rather a core feature of how to implement AI in our health system. The next section deals with the initiatives that are available to be translated both in ethical and technical tools, in order to advance principles of equal access for health and care.

Is it possible to do no harm using artificial intelligence in the road towards Health for All?

The amount and speed of artificial intelligence and ADS developments in health is a major concern, as most devices and systems in place are barely tested before being offered to the market. A recent Canadian report summarized the ways to avoid that public health AI trained or tested on specific or biased data is released without previous checking using more comprehensive and diverse datasets.⁴⁹ As previously stated, the pace of innovation and the culture of secrecy^{50,51} in the industry are among the main causes of involuntary negative effects. Additionally, as previously analysed, the motivation behind many applications in delivering services or benefits seems to be related to maximize efficiency even though this may compromise equality. In both cases, the opportunity for applying social ethics to the design is usually an analysis of ex-post failures and has synergic negative effects with the lack of transparency of the algorithms. Sometimes a million-line code may have unexpected effects even for its designers. As happened with an Excel® formula,⁵² mistakes may appear randomly and only be detected reversing a large number of results.

⁴⁸ National Institute for Clinical Excellence. Evidence Standards Framework for Digital Health Technologies, March 2019.

⁴⁹ CIFAR-CIHR-IRSC: AI for Public Health Equity. Workshop Report January 25, 2019 Toronto. (Accessed on 07.08.2020 at: https://cihr-irsc.gc.ca/e/documents/ai_public_health_equity-en.pdf).

⁵⁰ Coravos. *We should treat algorithms...*

⁵¹ Richardson. *Litigating Algorithms...*

⁵² *Floating-point arithmetic may give inaccurate results in Excel.* (Accessed on 07.08.2020 at: <https://docs.microsoft.com/en-us/office/troubleshoot/excel/floating-point-arithmetic-inaccurate-result>).

In the case of health services delivery, withholding a benefit by a computer bug may be a case of life or death. As by Mateos-Garcia proposal, the tolerance to coding errors may be measured by three parameters: i. Risk: When should we leave decisions to algorithms, and how accurate do those algorithms need to be? ii. Supervision: How do we combine human and machine intelligence to achieve desired outcomes? iii. Scale: What factors enable and constrain our ability to ramp-up algorithmic decision-making?⁵³ Most artificial intelligence in healthcare falls into these parameters and should be tested using thorough methodologies. The task of testing and ensuring safety and effectiveness of algorithms in healthcare demands to work in the intersection of ethics, law (regulation), system design, computer coding and operations research, among others.

At the same time, as technology transforms itself, the questions related to ethics need likewise to be in a continuous state of change. As an example currently being subject of research in Brazil,⁵⁴ the technological changes in conceptualizing hypertension depend on changes in available technologies for the measurements, changes in hardware and software as well as the interacting fast changes of drugs and corresponding multi-billion markets. There is, therefore, a question of how to evolve principles to regulate a process that is recursively evolving in different spheres, at different paces. This process recalls Gödel's postulate, regarding the always-incomplete search for internal axiomatic consistency, raising doubts about fixed or immutable social ethical principles.

A summary of the proposals made by Virginia Eubanks about an Oath of Do No Harm in the age of Big Data and artificial intelligence,⁵⁵ include the following:

- Do not collect and keep information for the sake of it, as one day it may be used for ill purposes.
- More data points about an individual inevitably lead to profiling and commercial uses.

⁵³ Mateos-Garcia J. *To err is algorithm: Algorithmic fallibility and economic organization*. Towards Data Science, May 17, 2017. (Accessed on 07.08.2020 at: <https://towardsdatascience.com/to-err-is-algorithm-algorithmic-fallibility-and-economic-organisation-dbe18bb32abc>).

⁵⁴ Prof. Sergio Mascarenhas, Institute for Advanced Studies, University of Sao Paulo, personal communication.

⁵⁵ Eubanks. *Automating Inequality...* p. 212.

- Never create barriers between people and services. Always remove obstacles between resources and those who need them.
- Integrate systems for the needs of people, not for the interest of new or more technically elegant systems.
- Informed consent does not equal to unconditional surrender. Using data without explicit consent is ethically wrong.

Several academic groups are presently working in designing and implementing the kind of tests and codes of procedures that may unleash the potential of artificial intelligence in achieving health for all. The following is a list of some of the current efforts:

*The Montreal Declaration for a Responsible Development of Artificial Intelligence*⁵⁶

Using a critical approach to artificial intelligence, the focus of the Declaration co-constructed under the leadership of the University of Montreal, is not placed on praising the radical shift that AI is creating in human history, but on a cautious and progressive adaptation of technological innovations. Seen under this lens, its general conclusion is that the long-term vision of a society with “good AI” is still a work in progress. At the same time, and departing from optimistic or neutral views of the effects of artificial intelligence on society, the Declaration acknowledges that the calls to reduce discrimination and increase equality exist within a global context of growing inequalities. “In other words, it is difficult to isolate issues of AI ethics from issues of international justice.”

The Declaration is built on 10 principles: wellbeing, respect for autonomy, protection of privacy and intimacy, solidarity, democratic participation, equity, diversity inclusion, prudence, responsibility and sustainable development. Regarding the contributions that AI can make for society, the declaration states that its design and use must contribute to the creation of a just and equitable society; AI should promote justice and seek to eliminate discrimination, namely that of gender, age, mental and physical abilities, sexual orientation, ethnic and social origins and religious beliefs; respect of privacy and allow those who use it to access their personal data as well as the kinds of information used by the algorithm.

⁵⁶ University of Montréal: *Declaration for a Responsible Development of Artificial Intelligence*. (Accessed on 07.08.2020 at: <https://www.montrealdeclaration-responsibleai.com/>).

AI development should promote critical thinking and protect us from propaganda and manipulation. The Declaration is an AI guideline that was co-created by hundreds of citizens, experts, public actors and researchers. Due to this collective process, it makes the link between artificial intelligence and democracy: AI development should foster informed participation in public life, cooperation and democratic debate. Acknowledging the perils of the use of untested algorithms, every person involved in AI development must exercise caution by anticipating, as far as possible, the adverse consequences of AI use and by taking the appropriate measures to avoid them. As a consequence of this imperative to be cautious, the development and use of AI must not contribute to lessen the responsibility of human beings when decisions must be made. A final note states the need for a sustainable approach to the use of resources.

The challenges presented in the translation from ethical principles to regulatory norms are evident after reviewing the potential that AI has for amplifying existing injustices, the lack of transparency (even for its own designers) of the internal working of most algorithms as well as the global production and distribution of AI application. The chance of having an international compact to be used as a Rights Declaration, or as a standard to guide governments in the purchase of Algorithm Decision Systems offers an interesting avenue for collective action.⁵⁷

Further specification of how to incorporate into organizational work the principles related with the Declaration, are the subject of the initiative “Equity by Design: Science and Innovation in Health Systems – A South-North Dialogue” (University of Montreal, Health Hub).⁵⁸ Equity by Design in science and innovation in health systems (EDHS) requires that equity imperative be taken into account and embedded within the whole process of knowledge production, dissemination and use within health systems. Consequently, EDHS involves going beyond rational problem-solving or policy-making. Design thinking is defined as a “systematic innovation process that prioritizes deep empathy for

⁵⁷ Petitgand C, Regis C. *Principes éthiques et encadrement juridique de l'intelligence artificielle en santé : Exemple de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle*. Journal de Droit de la Santé et de l'Assurance Maladie. Numéro 22, 101-106, 2019.

⁵⁸ H-Pod: *Politics, organizations and law*. (Accessed on 07.08.2020 at: <https://h-pod.openum.ca/en/a-propos/presentation/>).

end-user desires, needs and challenges to fully understand a problem in hopes of developing more comprehensive and effective solutions".

*Aequitas Project*⁵⁹

The University of Chicago, through its Center for Data Science and Public Policy confronted the growing concerns on the risk of unintended bias in ADS used for social issues as criminal justice, education, public health, workforce development and social services that are affecting individuals from certain groups unfairly. Using different proposed bias metrics and fairness definitions and in spite of the lack of consensus on which definitions and metrics are the best, the Aequitas Project hopes to contribute with evaluation on real-world problems, especially in public policy.

It is an open source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine-learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools. As an example, the Audit Tool may check if a real-world algorithm, when applied to a sample dataset will achieve: False Positive Rate Parity – ensuring that all protected groups have the same false positive rates as the reference group; False Discovery Rate Parity – ensuring that all protected groups have equally proportional false positives within the selected set compared to the reference group; False Negative Rate Parity – ensuring that all protected groups have the same false negative rates as the reference group.

*Z-Inspection, University of Frankfurt*⁶⁰

The project for a routine inspection of automated decision systems is being developed through case studies in the Big-Data Center, University of Frankfurt. The basis of the inspection methodology is to perform two different levels of validation, namely macro-validation, centered on the values that guided the design; and micro-validation using data and checking for correspondence of the practical results and the values

⁵⁹ University of Chicago, Aequitas Project. (Accessed on 07.08.2020 at: <http://www.datasciencepublicpolicy.org/projects/aequitas/>).

⁶⁰ University of Frankfurt, Z-Inspection. (Accessed on 07.08.2020 at: <http://www.big-data.uni-frankfurt.de/z-inspection-process-assess-ethical-ai/>).

stated at the start of the process using different sociotechnical scenarios, The Z-Inspection, when applied to healthcare algorithms may minimize risks associated with AI, helping to establish trust in the system and foster ethical values and ethical actions, stimulating new kinds of innovation. It may also contribute to closing the gap between “principles” (the “what” of AI ethics) and “practices” (the “how”).

AI Now Institute Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability⁶¹

The AI Now Institute in the University of New York is a research institute examining the social implications of artificial intelligence. As part of its developments, the Institute proposed in 2018 a framework and a step-by-step guide for public agencies in the process of procurement for automated decision systems. The use of a general framework and a systematic checklist in the tendering procedure “gives both the agency and the public the opportunity to evaluate the adoption of an automated decision system before making the decision to implant it. This allows the agency and the public to identify concerns that may need to be negotiated or otherwise addressed before a contract is signed.” The Impact Assessment also includes sending alerts to those communities that may be affected once the system is in place. The proponents of the methodology expect that this type of assessment, if consistently applied, “would also benefit vendors (AI developers) that prioritize fairness, accountability, and transparency in their offering. Companies that are best equipped to help agencies and researchers study their system would have a competitive advantage over others. Cooperation would also help improve public trust, especially at a time when skepticism of the societal benefits of AI is on the rise.”

Conclusion

Artificial intelligence and its technological developments have huge potential to deliver many benefits both to individual health and to health services for the communities. They are already doing so for those who can afford to pay for the new services. They could also make an immense positive difference in improving the wellbeing of

⁶¹ AI Now Institute. *Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability* (Accessed on 07.08.2020 at: <https://airnowinstitute.org/aiareport2018.pdf>).

the less well-off members of society, but this will require to redirect the policies and procedures, acknowledging that any artificial intelligence device applied to social issues needs to actively go against inequality, otherwise it will amplify it.

There is a need to use human-centred design when developing and implementing new AI applications. It also implies considering legal and ethical questions through a human rights lens, but this is just a first step. Effective implementation will also require understanding the local, social, epidemiological, health system and political contexts. Any kind of wider scale implementation will need to be guided by a research agenda including alerts and transparency with all the communities that may be impacted by the system. Although not a panacea, artificial intelligence is one of several tools that could help in achieving the health-related targets set out in the Sustainable Development Goals, particularly those related to expanding access and quality of care in universal health systems. The leading role in any such effort will have to be played by governments through appropriate safeguards, procurement systems and regulatory initiatives. A recent UN Report calls for governments' genuine commitment to designing its "digital welfare state not as a Trojan Horse for neoliberal hostility towards welfare and regulation but as a way to ensure a decent standard of living for everyone in society".⁶² There is a need to listen to this warning and build societies that make artificial intelligence a tool for a more humane world.

⁶² UN Report of the Special rapporteur on extreme poverty and human rights. Seventy-fourth session. Item 72(b) of the provisional agenda. 2019 (Accessed on 07.08.2020 at: https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx).

AI in Medicine. Recent Progress in iPS Cell Research and Application

Shinya Yamanaka*

Introduction¹

In 2006, the research group led by Prof. Yamanaka reported that the combination of four genes of transcription factors (*Oct3/4*, *Sox2*, *c-Myc* and *Klf4*), inserted into somatic cells, converted mouse skin cells into pluripotent stem cells that were designated iPS cells (for induced pluripotent stem cells).²

A year later, the Yamanaka group demonstrated that the same four factors were able to produce human iPS cells.³ iPS cells can proliferate almost indefinitely and differentiate into multiple cell lineages, for example neurons, cardiomyocytes, muscle cells, hepatocytes, etc. The discovery performed by Prof Yamanaka allowed the development of iPS cell-based cell therapy.

The first promising results obtained using iPS cells in clinics arrived when these cells were used for age-related macular degeneration. iPS cell-derived retinal pigmented epithelial cells were transplanted into one of the patient's eyes and one year after surgery, the patient's vision in the treated eye was stabilized without rejection or tumor development. Since then clinical application using iPS cells have been used in the following fields: clinical research for age-related macular degeneration

* Center for iPS Cell Research and Application (CiRA), Kyoto University (Japan); Gladstone Institute of Cardiovascular Disease, San Francisco (USA); Takeda - CiRA Joint Program, Shonan, (Japan); Nobel Prize for Medicine 2012; Ordinary Member of the Academy.

¹ The slides by Shinya Yamanaka were presented at the General Assembly by Claudia Compagnucci (PhD, Molecular Genetics and Functional Genomics, Area of Genetic Research and Rare Diseases, Bambino Gesù Children Hospital, Rome, Italy), who wrote also this Introduction and the Conclusion, according with professor Yamanaka.

² Takahashi K, Yamanaka S. *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors*. Cell. 2006; 126(4): 663-676.

³ Takahashi K, Tanabe K, Ohnuki M, et Al. *Induction of pluripotent stem cells from adult human fibroblasts by defined factors*. Cell. 2007; 131(5): 861-872.

and cornea epithelial stem cell exhaustion, clinical trials for Parkinson's disease and ischemic cardiomyopathy. In addition to regenerative medicine, iPS cells are useful for drug development, an application that includes drug screening, toxicity studies and the elucidation of disease mechanisms using disease-specific iPS cells from patients with intractable diseases.

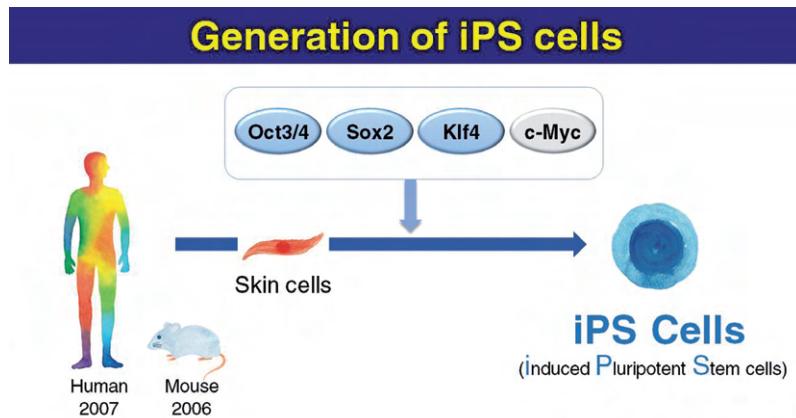
iPS cell-based science is moving forward for delivering innovative therapeutic options to the people with intractable diseases. In fact, several iPS cell Banks (including iPSCs obtained from controls and individuals with different disorders) are present worldwide and they mainly aim to contribute to the diagnosis and treatment of patients suffering from various diseases.

Over the past decade iPS cell research made great progress, but still various hurdles need to be overcome. This combination of iPS cell research with artificial intelligence (AI) provides a special opportunity for medical care and society at large.

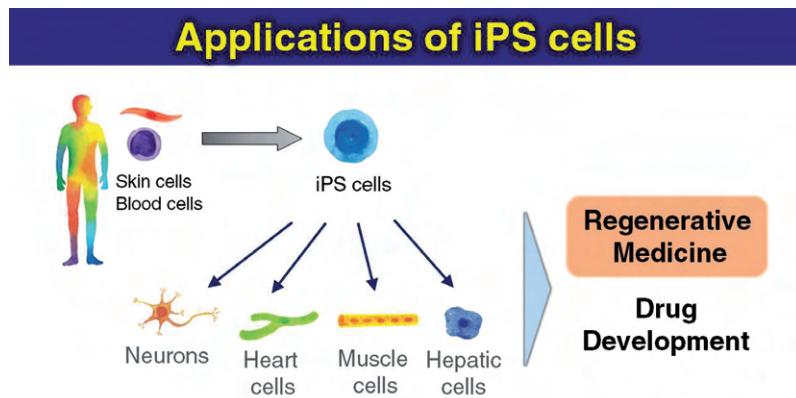
For example, different iPS cell clones have different properties in terms of their differentiation ability, proliferation ability, tumorigenicity, etc. Therefore, it is of great importance to develop a standardized methodology that can correctly predict the properties of iPS cell clones. The properties of iPS cell clones are known to be affected by multiple regulatory hierarchies including higher-order chromatin structures, epigenetic regulation, transcriptional regulation, post-transcriptional regulation and translational regulation. Thus, by applying AI techniques to iPS cell studies, Prof. Yamanaka and his colleagues are generating a trained neural network by multi- hierarchical omics-data to precisely evaluate the quality of iPS cell clones.

In addition, the combination of AI with iPS cell technology is applicable for providing "precision medicine" as AI system can learn the gene network signatures for 1,000 toxic chemicals that have different target organs such as neurons, kidneys, livers, or even carcinogens. Using this system, together with iPS cells, we expect to diagnose potential vulnerability for new drugs, new chemicals, and new foods at personal level in the future.

AI in Medicine

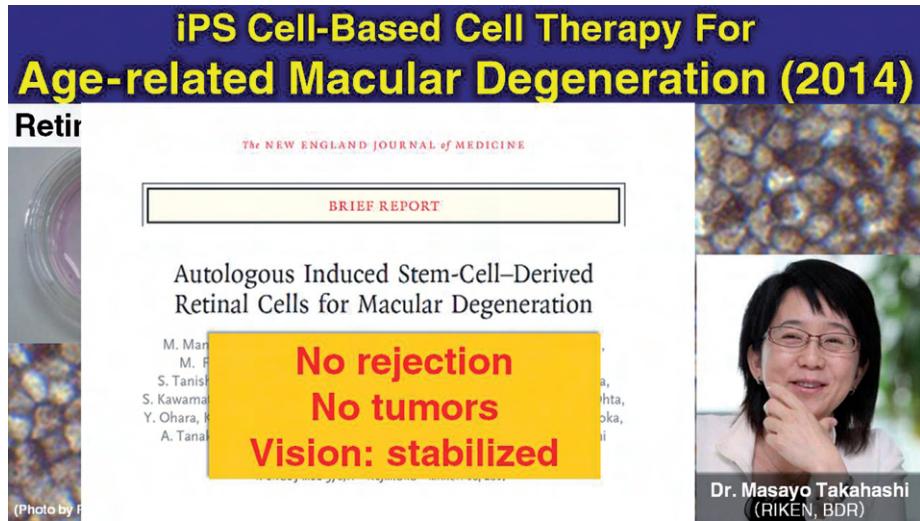


In 2006, we, through our research, were able to report that the combination of four transcription factors, Oct3/4, Sox2, c-Myc, and Klf4, converted mouse skin cells into ES cell like pluripotent stem cells, which we designated iPS cells for induced pluripotent stem cells. In 2007, we and others showed that the same four factors can make human iPS cells.



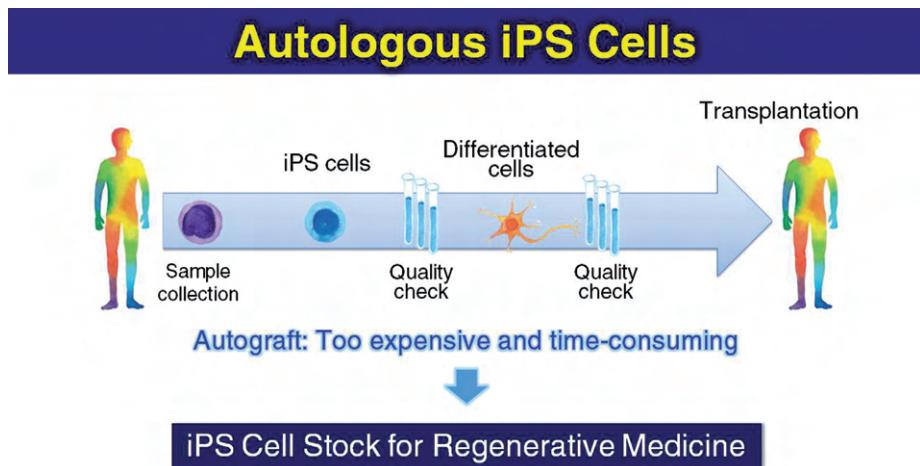
iPSCs can proliferate almost indefinitely and differentiate into multiple lineages, for example, neurons, heart cells, muscle cells, hepatic cells and so on.

Let me move on to the one of the applications of iPS cells, that is cell therapy also known as regenerative medicine.



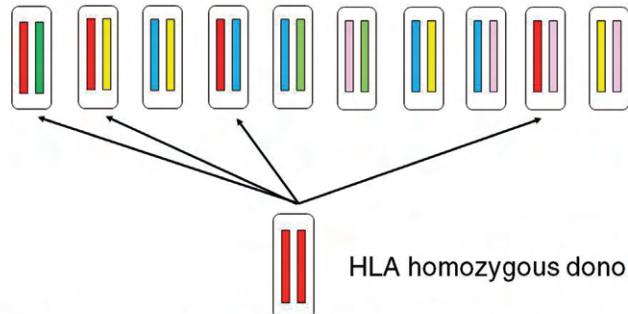
In 2014, the world's first clinical study using autologous iPSCs began for the treatment of age-related macular degeneration.

iPSC cell-derived RPE cells that passed CiRA's tests were transplanted into one of the patient's eyes in order to compare the therapy with the untreated eye. One year after the surgery, the patient's vision in the treated eye had stabilized and even showed improvement.



We are proceeding with an iPSC stock project in which clinical-grade iPSC clones are being established from "super" donors with homologous HLA (human leukocyte antigen) haplotypes.

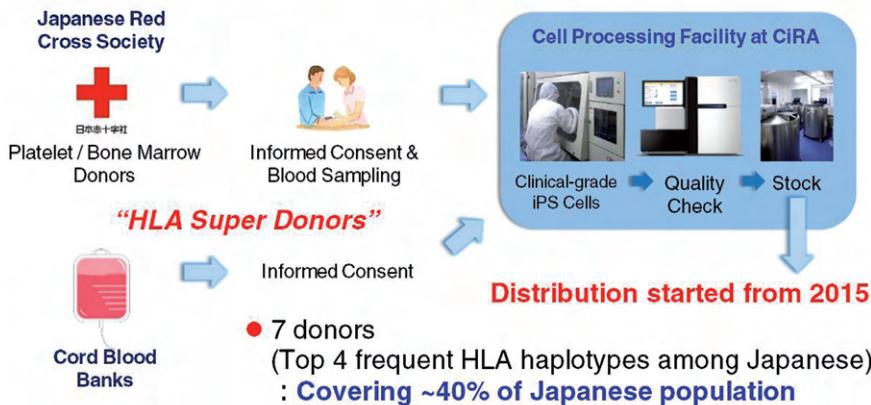
HLA Homozygous “Super” Donors



To reduce the cost & time of autologous iPSC

Homologous HLA haplotypes are associated with decreased immune response and therefore less risk of transplant rejection.

iPS Cell Stock for Regenerative Medicine



The building of an iPS cell stock for regenerative medicine involves the collection of cells from healthy donors with homozygous HLA. The aim of the stock is to hold iPS cells of guaranteed quality which can be supplied quickly to medical care institutions and research institutions in Japan and overseas when required.

In 2015, iPS cell lines generated at CiRA were available to several research institutions for further assessment. Now we distribute 7 cell lines that covered about 40% of Japanese population.

Center for iPS Cell Research and Application (CiRA)

Goal: To realize medical applications of iPS Cells



Started in April,
2010



14 Cell Processing Rooms

This is our institute, the Center for iPS Cell Research and Application, called CiRA.

CiRA was established in April, 2010. 2020 is 10th Anniversary year. Our Goal is "To realize medical application of iPS Cells".

Clinical Application Using iPS Cell Stock

Clinical Research



Masayo Takahashi Lab. (RIKEN)
Age-related Macular Degeneration

Clinical Trial



Jun Takahashi Lab. (CiRA)
Parkinson's Disease



Kohji Nishida Lab. (Osaka Univ.)
Cornea Epithelial Stem Cell Exhaustion



Yoshiki Sawa Lab. (Osaka Univ.)
Ischemic cardiomyopathy

Clinical Application Using iPS Cell Stock

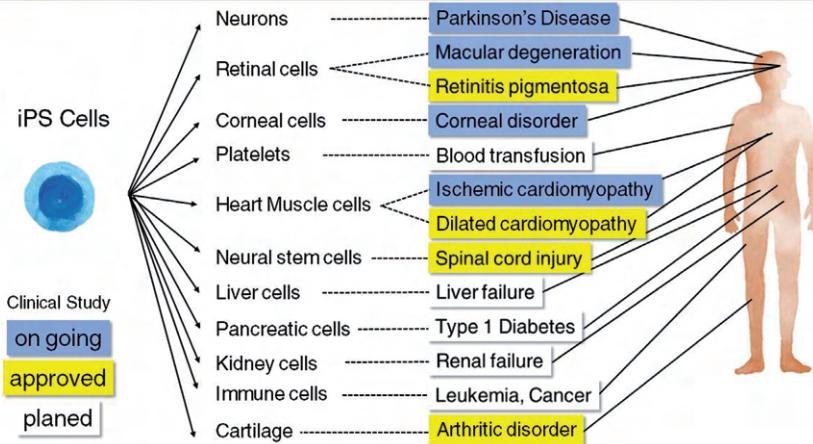
Approved by MHLW



University approved



Regenerative Medicine Using iPS Cell Stock



From iPS cells, we can prepare a large amount of human somatic cells, such as dopaminergic neurons shown as DA neurons, retinal pigment epithelial cells, cardiac cells, neural progenitor cells, and platelets. Scientists are now trying to treat patients suffering from various diseases and injuries, such as Parkinson diseases, macular degenerations, cardiac failure, spinal cord injury, and platelet deficiency. In Japan, several clinical studies are actively going on to test the efficacy and safety of these treatments.

iPS Cell Stock for Regenerative Medicine

Being distributed

- Top 4 frequent HLA haplotypes among Japanese
: Covering ~40% of Japanese population

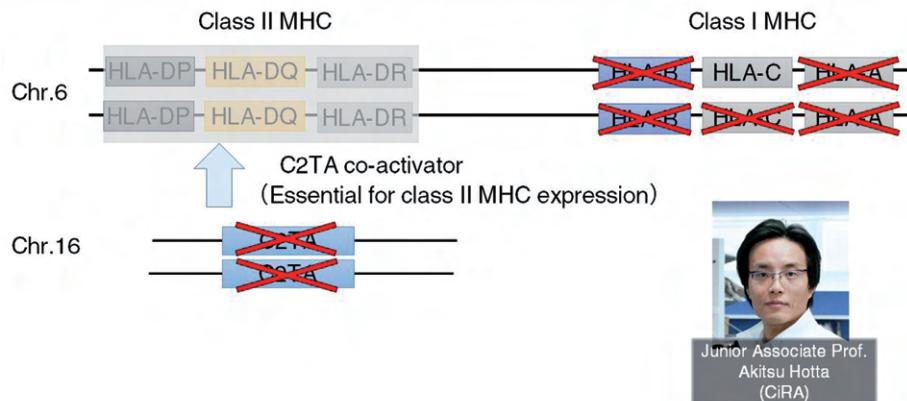
How about the remaining 60%?

- 150 haplotypes would cover ~90% of Japanese population
- >1000 haplotypes would be required to cover most of the world population

CiRA is recruiting donors who are HLA homozygous, as these donors match with a much larger number of the general population than those who are HLA heterozygous. However, donors who are HLA homozygous are rare.

7 cell lines covered about 40% of Japanese population, but how about the remaining 60%?

Alternative Approach ~ HLA-C Only



The most essential for immune matching involve HLA class I, which include the subsets HLA-A, HLA-B, and HLA-C, and HLA class II, which include the subsets HLA-DP, HLA-DQ, and HLA-DR.

We deleted both HLA-A, HLA-B and mono HLA-C alleles to retain one HLA-C allele. Like pseudo-homozygous HLA class I, these cells evaded CD8 T cells, and also evaded NK cells. More importantly, fewer lines are needed to serve a large population.

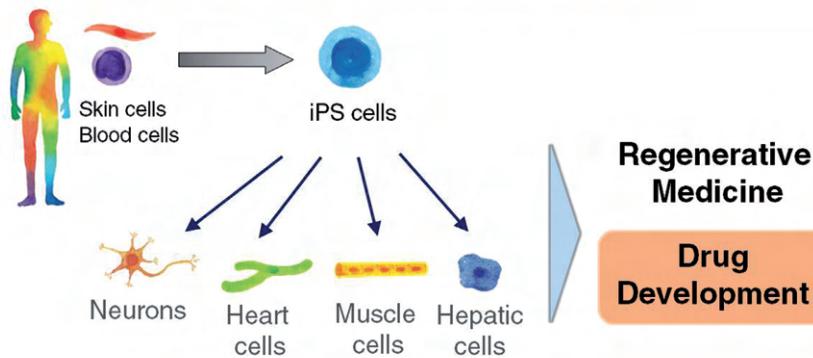
Furthermore, we deleted HLA class II.

These cells evaded CD8 killer T cells, CD4 helper T cells, and NK cells. We estimate that a core stock of 10 HLA-C retained iPS cell lines could serve the multiple populations.

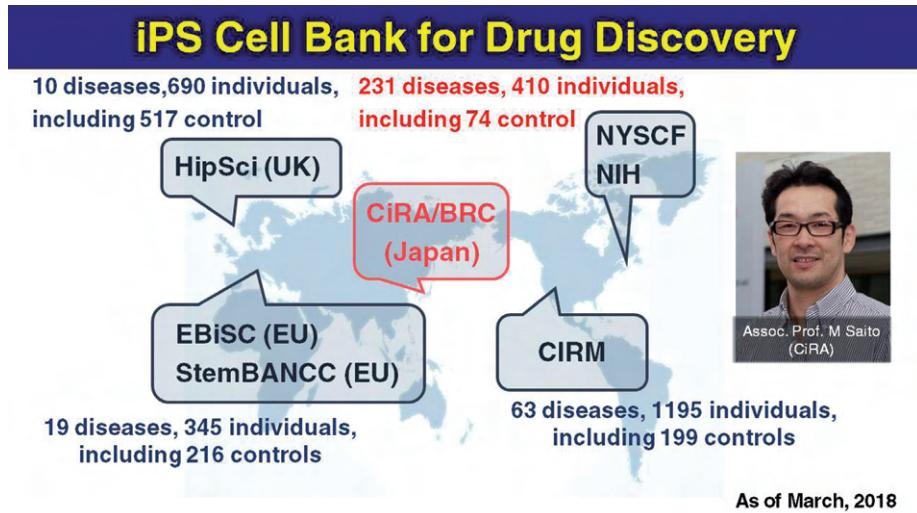
Future Plan of iPS Cell Therapy

Current	Super Donor iPS Cell Stock 4 Types: Covering ~40% of Japanese population
Alternative (2020~)	Genome-Editing iPS Cell Stock 10 lines would cover most of world population
Ultimate (2025~)	My iPS Cells

Applications of iPS cells



Let me move on to the other application of iPS cells include drug screening, toxicity studies and the elucidation of disease mechanisms using disease-specific iPSCs from patients with intractable diseases.



This slide shows the iPS Cell Bank in the world.

In Japan, our institute established iPS cell line from patients with various genetic diseases to investigate the mechanism of diseases, and to look for new treatment.

We hope our Cell Bank may contribute to the diagnosis and treatment of patients suffering from various intractable diseases.

Drug Repurposing with Patient iPSCs

Two clinical trials are ongoing at Kyoto University Hospital

Rapamycin for FOP

(Fibrodysplasia Ossificans Progressiva)



Prof. Toguchida, Assoc. Prof. Ikeya
(CiRA)

Bostinib for ALS

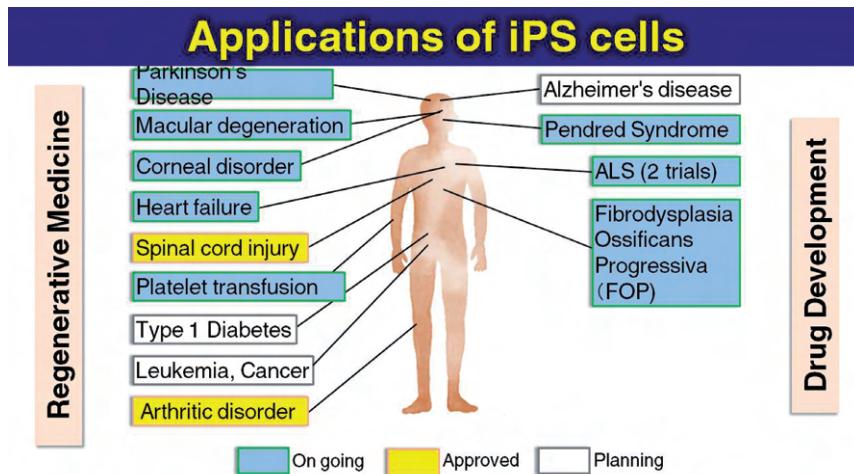
(Amyotrophic lateral sclerosis)



Prof. H Inoue
(CiRA)

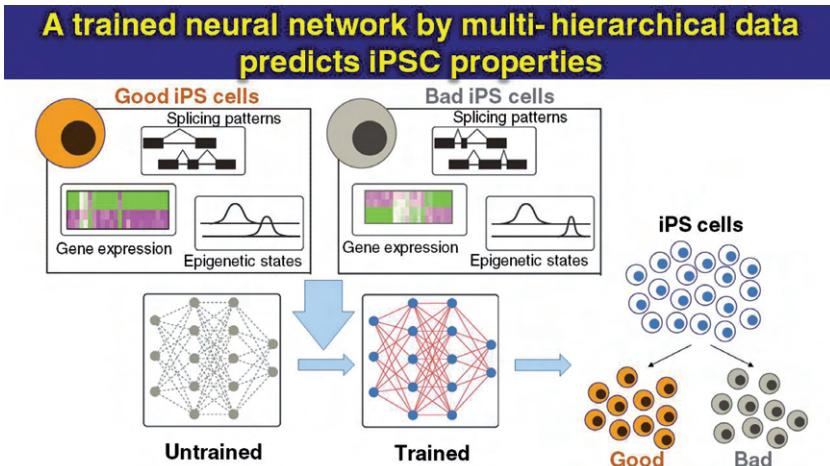
We reported a new drug screening system using iPSCs derived from fibrodysplasia ossificans progressiva (FOP) patients, revealing one drug candidate, Rapamycin; based on these findings, we have achieved to initiate a clinical trial to treat FOP patients in 2017.

Additionally, Bosutinib, a drug for leukemia was revealed to be efficacious for the treatment of amyotrophic lateral sclerosis (ALS) using a disease model established from patient-derived iPSC; based on these findings, we have achieved to initiate a clinical trial to treat ALS patients in 2019.

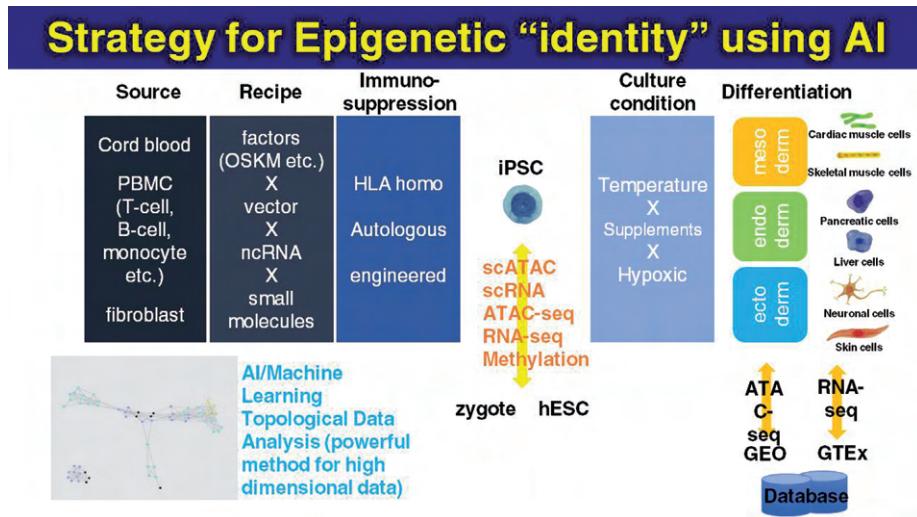


Over the past decade iPSCs research made a great progress. However, there are still various hurdles to be overcome, iPSC-based science is certainly moving forward for delivering innovative therapeutic options to the people with intractable diseases.

The combination of AI and iPSCs will have a significant impact on medical care and society at large.



Different iPS cell clones have different properties in terms of their differentiation ability, proliferation ability, tumorigenicity, and so on. Therefore, it is of great importance to develop a methodology that can predict the properties of iPS cell clones precisely. The properties of iPS cell clones are known to be affected by multiple regulatory hierarchies including higher-order chromatin structures, epigenetic regulation, transcriptional regulation, post-transcriptional regulation and translational regulation. Therefore, now, by applying AI techniques into stem cell studies, we are trying to generate trained neural network by multi-hierarchical omics-data to precisely evaluate iPS cells clones.



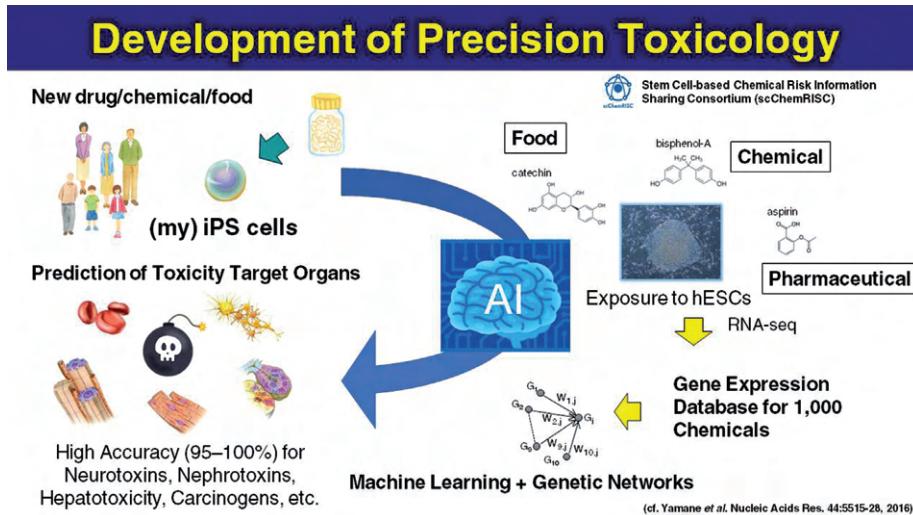
For investigating the characters of iPS cell lines, we perform many assays including genomic sequencing and gene expression analysis.

The latest advanced technologies enable us to obtain the data concerning with gene expression or epigenetic status at single cell levels.

On the other hand, the methods to generate iPS cells include many things such as what type of the cells or which kind of the factors should we use, or autologous or HLA-KO iPS cells should be better for the reduction of immunorejection.

To consider that which protocol is desired for a purpose, we need to accumulate knowledge by summarizing enormous amount of information.

A more closed connection between biologists or physicians and informaticians would be further required from now on.



We plan to be able to make “my iPS cells” to every individual person at low cost by 2025 for providing “precision medicine”. Particularly, “precision toxicology” using AI for foods, chemicals, and drugs is already ahead of other studies. We have already confirmed that ES cells are very sensitive to most of chemicals and show specific gene network signatures by chemical exposures.

Currently, many companies are supporting our activities and have established a consortium, which plans to develop AI system that can learn the gene network signatures for 1,000 toxic chemicals that have different target organs such as neurons, kidneys, livers, or even carcinogens.

Using this system, together with “my iPS cells”, we can diagnose potential vulnerability for new drugs, new chemicals, and new foods at personal level in the future.

Conclusions

In conclusion, the combination of AI system with iPS cell can leverage the strengths of each cutting-edge technologies, which provide the novel insights to change the future medical care. Borderless connection between biologists or physicians and computer scientists are mandatory to make this happen.

Third session

ARTIFICIAL INTELLIGENCE AND LAW

Policy and Governance of AI for Health: A Global Ethics Perspective

James A. Shaw *

(Chapter by James A. Shaw and Leah T. Kelley **)

Introduction

Artificial Intelligence (AI) is a general-purpose technology, and its potential applications in health care are numerous and diverse.¹ As applications of AI to various domains of health care delivery are developed and implemented,² large technology corporations with expertise in AI such as Alphabet (the parent company of Google) are positioning themselves to become central contributors to AI innovation for health.³ Collaborations between Alphabet companies and health care organizations have created public concern in Europe⁴ and North America⁵, raising public awareness regarding the ethical issues associated with collaborations between extremely large and profitable technology companies (referred to herein simply as “Big Tech”) and organizations focused on delivering health care. Drawing on the literature on surveillance capitalism and innovation systems, in this paper we provide an ethical analysis of the nature of Big Tech’s involvement in advancing AI innovation for health care.

* Research Director, Artificial Intelligence, Ethics & Health, University of Toronto Joint Centre for Bioethics; Scientist, Institute for Health System Solutions and Virtual Care, Women's College Hospital (Canada).

** Research Coordinator, Institute for Health System Solutions and Virtual Care, Women's College Hospital (Canada).

¹ Shaw J, Rudzicz F, Jamieson T, Goldfarb A. *Artificial intelligence and the implementation challenge*. Journal of Medical Internet Research, 2019; 21(7): 1-7.

² Topol E. *High-performance medicine: the convergence of human and artificial intelligence*. Nature Medicine, 2019; 25: 44-56.

³ Powles J & Hodson H. *Google DeepMind and healthcare in an age of algorithms*. Health & Technology, 2017; 7: 351-367.

⁴ *Ibid.*

⁵ Singer N & Wakabayashi D. Google to store and analyze millions of health records. New York Times, November 11th, 2019. (Accessed on 12.26.2019 at: <https://www.nytimes.com/2019/11/11/business/google-ascension-health-data.html>).

AI has immense potential to enhance the quality, effectiveness, and efficiency of health care systems around the world, and for that reason this general-purpose technology warrants extremely close attention. As with any new technology that is rapidly diffusing across a particular context of use, AI raises numerous ethical issues that require sustained analysis and practical response in terms of both organizational governance and public policy. The literature analyzing and summarizing ethical issues specifically in relation to AI for health care is already substantial, focusing on challenges such as consent for use of health data, trustworthiness of algorithmic outputs, and the potential bias embedded in the functioning of algorithms.⁶ The particular ethical issues on which we are focused in this paper have received comparatively little attention in academic literature, being those related specifically to the *nature* and *potential consequences* of collaborations between Big Tech and health care organizations.

In this paper, we focus in particular on the role of Alphabet in developing collaborations that advance their strategic presence in health-related AI innovation. We refer to Alphabet as opposed to Google, DeepMind, or Verily more specifically (all companies owned by Alphabet) to reflect these companies' common structures of ownership and accountability to Alphabet as their parent company. Our reason for focusing on Alphabet is partly a result of the high profile instances of their involvement in health care and clear documentation in the popular press of their ongoing health-related activities. However, other Big Tech companies are certainly also advancing deeper into the health care innovation domain on a foundation of AI technologies.⁷

Alphabet has been challenged publicly on the legal and ethical status of their collaborations with health care, leading for example to a lawsuit brought against Google and the University of Chicago Medical Center claiming that their 2019 collaboration was in violation of the Health Insurance Portability and Accountability Act (HIPAA).⁸ This

⁶ Jobin A, Lenca M, Vayena, E. *The global landscape of AI ethics guidelines*. Nature Machine Intelligence, 2019; 1: 389-399.

⁷ Eddy N. *Big tech poised to beat healthcare in reaping value from artificial intelligence, report says*. Healthcare IT News, 2019. (Accessed on 12.27.2019 at: <https://www.healthcareitnews.com/news/big-tech-poised-beat-healthcare-reaping-value-artificial-intelligence-report-says>).

⁸ Stoller D. Google, University of Chicago faced revamped health privacy suit. Bloomberg Law, 2019. (Accessed on 12.27.2019 at: <https://news.bloomberglaw.com/privacy-and-data-security/google-university-of-chicago-face-revamped-health-privacy-suit>).

example is illustrative of the kind of legal and ethical argumentation brought against Big Tech for its involvement in AI innovation for health care, relying largely on the obligation to obtain informed consent from patients prior to sharing data for reasons other than the direct improvement of patient care.⁹

In strategic response to challenges such as this one, Alphabet has sought to portray itself as an ethically aware company that is working on the frontiers of AI innovation. The effort to do so has included for example the brief attempt to establish an AI ethics board for Google in 2019, which was ultimately unsuccessful as a result of resistance from Google employees regarding one of the members appointed to the board.¹⁰ This and other attempts to control the discourse about AI ethics has led to accusations of “ethics washing” levied against Alphabet and other Big Tech companies. Ethics washing refers to the accusation that Big Tech uses diluted efforts to promote ethics in AI innovation as a strategy to avoid stronger regulation and deeper changes to their business models.¹¹

To gain a deeper understanding of the ethical implications of the nature of Big Tech’s involvement in AI innovation for health care, we begin with an inquiry into the path dependencies being created by Alphabet’s recent strategic positioning in health-related AI innovation. We use path dependence to refer to “historical sequences in which contingent events set into motion institutional patterns or event chains that have deterministic properties”.¹² We begin our inquiry in the recent past, examining the health-related activities of Alphabet in 2019. We outline the expectations these activities create related to the structure of collaborations between Big Tech and health care, and the broader legislative toolbox that governs such collaborations. We use the concept of path dependence to project outward to the future, describing four hypothetical futures that might be created by path dependencies currently being established by Alphabet’s activities. We then comment on the central roles of public trust and public value in enabling Big Tech’s

⁹ Cohen IG, Mello M. *Big data, big tech, and protecting patient privacy*. Journal of the American Medical Association, 2019; 322(12): 1141-1142.

¹⁰ Piper K. *Exclusive: Google cancels AI ethics board in response to outcry*. Vox, 2019. (Accessed on 12.27.2019 at: <https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board>).

¹¹ Ochigame R. *The invention of “ethical AI”*. The Intercept, 2019. (Accessed on 12.27.2019 at: <https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/>).

¹² Mahoney J. *Path dependence in historical sociology*. Theory & Society, 2000; 29: 507.

involvement in health-related AI innovation, and describe how Big Tech must go well beyond regulation in order to achieve this goal. After outlining the theoretical perspectives on which our analysis is based, our paper addresses three questions in particular.

1. What path dependencies are currently being created by the ways in which Alphabet is becoming involved in the development of AI innovations for health care?
2. What are the potential implications of these path dependencies for the development and deployment of AI for health care?
3. What are some central considerations of policy and governance in the effort to develop a model for generating AI innovations for health care that stand to maximize public benefit while minimizing unintended consequences?

Theoretical Orientation

Our paper is intended as a contribution to the anticipatory governance of AI innovation in health care. In Science and Technology Studies, anticipatory governance has been defined as having three interconnected elements: (1) a collection of practices distributed among a wide range of stakeholders in society, (2) that are oriented toward better understanding the possible consequences of new and emerging science and technology, (3) in order to establish governance mechanisms that enable their benefits while mitigating against risks.¹³ Although some authors have suggested that anticipatory governance has particular conceptual implications,¹⁴ and the concept of anticipation itself has been the subject of debate,¹⁵ we rely only on the general motivation characteristic of this approach to better anticipate emerging risks and opportunities associated with applications of technology in a particular domain. To do so we use the concept of path dependence as defined in our introduction. Specifically, we document current activities related to the role of Big Tech in AI innovation for health care, and identify potential path dependencies that might be put into motion by those activities. The goal is to more clearly describe these path dependencies in order to inform more appropriate governance decisions to address them.

¹³ Guston D. *Understanding anticipatory governance*. Social Studies of Science, 2014; 44(2): 218-242.

¹⁴ *Ibid.*

¹⁵ Nordmann A. *Responsible innovation, the art and craft of anticipation*. Journal of Responsible Innovation, 2014; 1(1): 87-98.

Our analysis is informed by two complementary theoretical concepts. The first is surveillance capitalism, which has been described and analyzed in detail by Shoshana Zuboff in her 2019 book, *The Age of Surveillance Capitalism*.¹⁶ The concept is intended to characterize the economic logic of an unprecedented version of capitalism that has become dominant globally, driven by firms such as Google and Facebook that utilize mass amounts of behavioral data to target and promote consumer behavior and amass immense profits through advertising revenues. Zuboff's concern is that this new economic logic systematically erodes opportunities for human agency that are fundamental to market democracy, and as such her issues with surveillance capitalism run much deeper than the structure of Big Tech's business models. However, for the purposes of our analysis, three features of surveillance capitalism are most relevant.

First, through the use of analytics techniques on behavioral data, firms become capable of predicting and influencing behavior on mass scales. This occurs specifically by analyzing large amounts of data representing behaviors in which people have engaged (e.g., which website links a person has followed), in order to structure the choices available to a person in the future such that they are more likely to engage in a particular behavior (e.g., follow a given link to an advertisement appearing online). In the case of health care, this influence could theoretically extend to a wide range of stakeholders, from health system funders, to clinicians, and entire patient populations. For example, the use of behavioral data in this way could influence the behavior of clinicians as they make decisions about the best way to treat patients.

Second, through control of the expertise, computing power, and relevant data, a select few firms accrue power over the various industries in which they participate. In health care, the argument from surveillance capitalism would suggest that if Big Tech gains control of health data and health care expertise (e.g., clinicians and health care leaders), then it would possess all that is required to exert immense amounts of control over the structures and processes that constitute health care. This point will be elaborated in our paper.

Third, as these firms amass such power, they further amass the ability to influence the behavior of the stakeholders listed. This creates a self-reinforcing loop, wherein the power of these select few firms over particular industries (such as health care) could grow almost indefinitely if left unchecked. The existence of law and health care policy are obvi-

¹⁶ Zuboff S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Public Affairs: New York.

ous balancing forces to this process, and as such describing this scenario is merely a matter of illustrating the logic of the argument from surveillance capitalism.

The second theoretical concept informing our analysis is the concept of an innovation system, which identifies the institutions, stakeholders, and activities involved in producing innovations in regional ecosystems.¹⁷ Detailed work has been done applying this concept to health innovation in particular, which has generated the basic but essential insight that collaboration between relevant stakeholders is essential to produce genuinely novel innovations that are of benefit both practically (i.e., in health care) and economically.¹⁸

From this literature we carry forward the insight that an appropriate balance must be struck between stakeholders, governed by relevant institutions, such that all stakeholders experience value from collaborative activities that promote innovation in the system. This perceived value can arise from any number of circumstances, yet is an essential component of a high functioning innovation system. We examine four hypothetical future scenarios based on path dependencies currently being established that illustrate different possible balances of power among stakeholders in innovation systems for the development of AI technologies for health care. We introduce these hypothetical futures not in an attempt to predict the future, but simply in an attempt to illustrate possible outcomes of path dependencies that are currently being established related to AI innovation for health care.

These two concepts (innovation systems and surveillance capitalism) enable us to examine the structures of collaboration between Big Tech and health care in finer detail. Acknowledging that collaboration is essential for a high functioning innovation system, and that Big Tech currently houses the expertise and computing power to perform analytics on virtually any kind of quantitative data, the ways in which Big Tech and health care collaborate will in many ways dictate the quality of the innovation system overall. We discuss this throughout our paper, and address the ethical implications of such collaborations for the development of AI technologies for health care.

¹⁷ Ranga M & Etzkowitz H. *Triple helix systems: An analytical framework for innovation policy and practice in the knowledge society*. Industry and Higher Education, 2015; 27(3): 237-262.

¹⁸ Consoli D & Mina A. *An evolutionary perspective on health innovation systems*. Journal of Evolutionary Economics, 2009; 19: 297-319.

What is Alphabet Doing in Health Care?

Alphabet is the corporate parent of Google, which is by far the largest of the Alphabet companies. As reported in an open letter from Google co-founder Larry Page,¹⁹ Alphabet was established in 2015 in order to enable Google's business to remain focused on "organizing the world's information" while at the same time allowing the corporation to invest in other business opportunities. Alphabet owns over 200 companies that represent investments in various industry domains, with the advertising business that is core to Google remaining by far its most profitable.²⁰ Three Alphabet companies are explicitly focused on health: Verily (focused on health care improvement), DeepMind (focused on applying AI to health care), and Calico (focused on technological strategies to combat aging-related diseases).²¹ In November of 2018, all of Alphabet's initiatives focused on health were brought into a single division under the Alphabet corporate structure, referred to as Google Health.²²

A CBInsights report describes Alphabet's activities in health care as falling into three domains.²³ The first domain is "data generation", which includes technologies focused on creating data about health-related phenomena such as fitness tracking. The second domain is "disease detection", focused on identifying signals in datasets that indicate the presence of a particular disease. The third domain is "disease/lifestyle management", focused on supporting healthy lifestyles as people live their everyday lives. A fulsome description of Alphabet companies' activities in each of these domains is beyond the scope of this paper, and readers can refer to the CBInsights report for those details. However, an overview of Alphabet's activities in 2019 is provided in Box 1, and this overview provides sufficient information to understand at least some of the important recent activities of Alphabet in relation to the development of AI technologies for health.

¹⁹ Page L. G is for Google. (Accessed on 01.07.2020 at: <https://abc.xyz/>).

²⁰ Johnston K. Top 4 Companies owned by Google. (Accessed on 01.07.2020 at: <https://www.investopedia.com/investing/companies-owned-by-google/>).

²¹ CBInsights. How Google plans to use AI to reinvent the 3 Trillion dollars US health care industry. (Accessed on 01.07.2020 at: <https://www.cbinsights.com/research/report/google-strategy-healthcare/>).

²² Li A. Google Health details its mission to "help everybody live healthier lives". (Accessed on 01.07.2020 at: <https://9to5google.com/2019/11/19/google-health-mission/>).

²³ CBInsights, *Ibid.*

From this summary of Alphabet's health-related activities, we understand there to be four distinct categories of strategic investment in health. First is investment in the surveillance of public health issues, through for example the development of a drug disposal drop-off feature in Google maps. Second is surveillance in relation to health and fitness more generally, through for example Alphabet's acquisition of Fitbit. Third is establishing access to and developing infrastructure for AI-based analytics of electronic health record data, which was observed through Google's partnerships with University of Chicago Medical Center and with Ascension Health System (Project Nightingale). Fourth and finally is Alphabet's acquisition of members of its health-related leadership team who bring with them networks and experience in key policy institutions, such as Robert Califf (former Commissioner of the United States Food and Drug Administration) and Karen DeSalvo (former National Coordinator for Health IT). We outline the potential implications of these strategic investments in the next section of our paper.

Box 1. Alphabet's Health-Related Activities in 2019 as reported by Mobi-Health News²⁴

- Purchased mobile sensing technologies tracking location and other movement data (Fitbit and Fossil Group smart watch technology)
- Patent application filed on a process for structuring health record data for predictive analytics
- Created map feature outlining drug disposal drop off locations
- Verily received \$1billion US in new venture funding
- Verily began marketing products related to screening for diabetes-related diseases
- Google-University of Chicago lawsuit related to University sharing potentially identifiable data with Google
- Google developed two apps related to accessibility for those with hearing loss ("Live Transcribe", which transcribes audio to text, and "Sound Amplifier", which boosts the volume of sound)
- Consolidated DeepMind to join Google Health
- Former FDA Commissioner hired as head of Google Health (Robert Califf)
- Former National Coordinator for Health IT hired as CEO of Google Health (Karen DeSalvo)
- Google-Ascension partnership becomes public, with much public attention

²⁴ Mobihealth News. *A look back at Alphabet's moves in 2019*. (Accessed on 01.07.2020 at: <https://www.mobihealthnews.com/news/look-back-alphabets-moves-2019>).

Three Potential Path Dependencies

The strategic investments just described point toward the establishment of potential path dependencies related to Alphabet's involvement in AI innovation for health care. Although the activity of AI innovation for health care remains far too new to establish firm conclusions about path dependencies, this does not preclude the effort to consider potential consequences of current trends. Given the strategic investments of Alphabet just summarized, we suggest there are three potential path dependencies being established that are most relevant to consider. These potential path dependencies are summarized in Box 2, and elaborated in this section.

Box 2. Three Potential Path Dependencies

- 1 - Alphabet uniquely establishes access to data across public health issues, lifestyle, and health care records, thereby having access to virtually all data related to certain health-related phenomena.
- 2 - Alphabet acquires expertise in health care, thereby obtaining the necessary knowledge to produce AI technologies that are practically relevant in health care and economically viable in the market.
- 3 - Alphabet further entrenches its influence in government decision-making by hiring senior government employees into leadership roles in the company.

The first potential path dependency arising from these investments is that Alphabet establishes unique access to the necessary data sources to become the dominant actor in the effort to integrate data across known public health issues, lifestyle patterns, and health care records. Through a corporate structure that owns a number of companies investing in these various domains, and through existing investments in collaborations involving health record data, Alphabet could solidify its access to a diverse collection of data representing virtually all aspects of an individual's or population's health. This would result in Alphabet being uniquely positioned among competitors given its access to such a variety and quantity of health-related data.

The second potential path dependency relates directly to the first. As Alphabet continues to invest in various health-related technology opportunities, it establishes ownership over the confluence of expertise required to build AI technologies that are both practically relevant in health care and economically viable from a market perspective.

The unique analytics capabilities of Alphabet companies make it very attractive for health care providers interested in AI innovation, and the further access to health-related data will presumably make employment by Alphabet companies all the more attractive to innovation-minded clinicians. Furthermore, as Alphabet establishes which approaches to collaborating with health care systems are acceptable to the public, it could become more adept at establishing agreements that provide structured access to expertise that it is not capable of owning outright.

The final potential path dependency we address in this paper is the further entrenchment of the relationships between Alphabet and the policy-making establishment of the United States Government. Hiring important figures in governmental decision-making (such as previous leaders of the government organizations responsible for coordinating and regulating technology use in health care) puts Alphabet in a position of immense lobbying power. These individuals arrive at Alphabet with a long working history and extensive professional networks, which are then accessible to Alphabet for dialogue and influence regarding the role of Alphabet in health care. Such governmental influence could put Alphabet in a position to promote an approach to policy and regulation that provides the least resistance possible to Alphabet's goals of corporate profit. These potential path dependencies could have a wide variety of actual implications for AI innovation for health care, and we outline four of these hypothetical future scenarios next.

Four Possible Futures of AI Innovation for Health Care

The potential path dependencies just summarized provide an opportunity to imagine possible futures created if these path dependencies were to take hold and become dominant in the coming years. This approach to imagining possible futures, sometimes referred to as "foresight" or "futures studies", is commonly associated with anticipatory governance as described earlier in our paper.²⁵ We take a modified approach to such foresight-informed thinking in order to engage our audience in the task of considering the potential consequences of the path dependencies we have outlined.

²⁵ Fuerth L. *Foresight and anticipatory governance*. Foresight, 2009; 11(4): 14-32.

Future #1: Growing public distrust of Big Tech and resistance from health care professionals leads to government policy that creates disincentives for Alphabet to have any role in AI for health care.

Recent experience shows that public concern about health-related AI is growing,²⁶ particularly in regards to privacy, informed consent for use of personal data, and ownership of data.²⁷ Furthermore, trust in AI systems is low among physicians, with half of physicians surveyed by one organization (N=500) reporting feeling anxious or uncomfortable using AI-based software.²⁸ Public issues with the sharing of data between health care organizations and Alphabet have led to calls for the urgency of re-thinking existing policy frameworks related to the acceptable uses of health-related data,²⁹ with some experts advocating for a model of health data regulation that contains a similar policy logic to the European General Data Protection Regulation (GDPR).³⁰

The GDPR has shifted toward consent anchored to the data itself, rather than the organization collecting the data. Therefore, any organization that collects personal data is subject to the GDPR and consent is required for any secondary use of data collected, regardless of the organization.³¹ However, the GDPR appears to consider de-identification of data a sufficient substitute for consent on the basis that it be analyzed on grounds of legitimate interest.³² If resistance from the public and health care providers regarding the sharing of health care data with Big Tech were to continue to grow, governments may be forced to exclude even de-identified health care data from being shared with Big Tech or other private interests, or at least to seriously question what counts as “grounds of legitimate interest”. Under these conditions, the obstacles to acquiring the necessary data to drive AI innovation for health care

²⁶ Powles J & Hodson H. *Google DeepMind...* pp. 351-367.

²⁷ Racine E, Boehlen W, Sample M. *Healthcare uses of artificial intelligence: Challenges and opportunities for growth*. Healthcare Management Forum, 2019; 32(5): 272-275.

²⁸ Frelick M. *AI Use in Healthcare increasing Slowly Worldwide*. Medscape, 2019. (Accessed on 01.07.2020 at: <https://www.medscape.com/viewarticle/912629>).

²⁹ Singer N & Wakabayashi D. *Google to store and analyze millions of health records*. New York Times, November 11th, 2019.

³⁰ Bari P. & O'Neill D. Rethinking patient data privacy in the era of digital health. Health Affairs Blog, December 12th, 2019. (Accessed on 01.09.2020 at: <https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/>).

³¹ Van der Auwermeulen B. *How to attribute the right of data portability in Europe: A comparative analysis of legislations*. Computer Law & Security Review, 2017; 33(1): 57-72.

³² Hintze M. *Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification and Compliance*. International Data Protection Law (Oxford University Press), 2018; 8(1): 86-101.

would be too high, and it would be unlikely that Alphabet or other Big Tech companies would continue to pursue their involvement in AI innovation for health care. In this possible future, large scale AI might not feature prominently at all in health care delivery, with only small, contained AI initiatives taking place within the regulatory boundaries and data resources contained within a given health care organization or system itself.

Future #2. Big Tech dominates AI for health care, and Alphabet increasingly exerts influence over health system planning

In a scenario where governments or other non-governmental actors are unable or unwilling to establish the policy frameworks and governance models to direct the role of Big Tech in AI innovation for health, the current activities of Alphabet in health care could feasibly continue to advance unchecked. This could hypothetically consist of the continued investment in acquiring the expertise and influence over data access and innovation in health care, along with persistent influence over government policy. In this possible future, the nature and spread of AI innovation in health care could be extraordinary, with health systems oriented toward influencing the health promotion behavior of patients and enhancing the quality of clinical decisions. However, the clear existence of the profit motive among Big Tech corporations raises important questions about the negative consequences of this potential future state.

The emphasis on monetizing the AI innovations or the underlying data assets on which they are built would open additional opportunities for revenue generation through health service delivery for Alphabet. Such opportunity could arise through the marketing of particular health-related products and services directly to consumers, which when leveraging behavioral advertising techniques, could substantially influence individual behavior.³³ These same techniques could be applied to clinical processes such that clinical decision-making is substantially influenced by AI technologies. In this possible future, through continued collection of data about the behavior of clinicians and patients, and the development of AI that intervenes in various aspects of health care and public health, Big Tech could establish immense amounts of influence over clinical processes and health system design. A key consideration

³³ Zuboff, *The Age of...*

in this scenario is that the balance of considerations would shift from focusing on whether health system decisions most strongly consider the public interest toward how strongly they consider the possibility of generating profit for Big Tech.

One additional point related to the profit-making orientation of Big Tech is the impact that expensive technological products have on the sustainability of health care systems. Lehoux et al (2016) outlined the impact that medical technologies have had on health care systems from a historical perspective, illustrating how technologies have led to specific changes in health systems (such as medical specialization) and have driven the overall costs of health care up substantially over time.³⁴ If Alphabet's role in health care were to result in AI innovations that enhance the quality or possibilities of health care, even health systems that do not have the resources to procure such technologies would presumably have some obligation to attempt to do so. Such a scenario could lead to further widening gaps in quality between health systems both within and between countries, and to further strain on growing costs in health care.

Future #3. Health systems build the infrastructure and talent necessary to use their data for analytics in house, removing the need to collaborate with Big Tech altogether. Alphabet does not become an important player in AI focused on health care

AI innovation driven by health systems presents the opposite extreme of challenges to AI driven by Big Tech. This scenario offers a certain degree of public value in that priorities for AI research and development would be set by health care providers, health systems, academic organizations, and the publics they represent. This approach would promote solutions that address direct quality and efficiency needs of health systems. However, health systems can rarely accept the risk of failed innovations due to sunk cost fallacies, resulting in sinking more resources into solutions with insufficient payoff.³⁵ This is particularly the case when such systems are publicly funded. Health systems have historically struggled to develop successful, homegrown technological

³⁴ Lehoux P, Roncarlo F, Oliveira RR, Silva HP. *Medical innovation and the sustainability of health systems: A historical perspective on technological change in health*. Health Services Management Research, 2016; 29(4): 116-125.

³⁵ Consoli. Mina. *An evolutionary perspective...*

innovation due to an inability to keep up with the commercial pace of innovation and to translate innovations into clinical practice.³⁶

These issues suggest that this possible future might lead to some substantial gains in AI innovation capabilities within health care systems and organizations, building on the important AI work already arising from within health care organizations.³⁷ However, the competing logic of public safety and the absence of the commercialization infrastructure present in Big Tech would presumably limit the spread of such innovations. In this potential future there could also be a larger role for small and medium sized technology firms, supporting the projects initiated by health care organizations or systems to develop AI innovation for health care on a more local level.

Although this scenario might prove beneficial within particular health systems or organizations, it is unlikely to address the substantial challenges of fragmented health care delivery that currently characterizes health systems around the world. The primary difference between this future and future #1 is that in the scenario we are currently describing, resources from government grants and health-related innovation would feed back into health care systems to support the development of AI innovation that is directly relevant for the needs of health care providers, organizations and systems.

Future #4: Best practices in the governance of collaborations between Big Tech and health care lead to the development of AI innovations that are in the public interest and satisfy the needs of Big Tech.

Existing reviews of health-related data privacy legislation such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States are laying the foundation for stricter controls over the sharing of health care data with industry. The logic of more general data protection policies such as the GDPR and the California Consumer Privacy Act (CCPA) shifts the focus from the rights and responsibilities of select organizations holding the data (i.e., the data custodians) toward the governance of the data themselves. This is important because it means that any organization holding the data (not just those identified as custodians) would be liable for how the data is shared, stored and used in response to the demands of the

³⁶ Jain S.H. *The health care innovation bubble*. *Healthcare*. 2017; 5(4): 231-232.

³⁷ Mamdani M, Laupacis A. *Building the digital and analytical foundations for Canada's future health care systems*. Canadian Medical Association Journal, 2018; 190(1): E1-E2

individual the data represent. Furthermore, the GDPR broadens the conventionally narrow definition of health information from only those data generated through the process of health care delivery to include all data “related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status”.³⁸

The broadening of the definition of health data and the larger scope of organizations theoretically made to be responsible for the storage, sharing and use of health data in these sorts of policies stand to potentially bolster public trust in health-related AI innovation. The review of health information policy in particular, which is currently under way in many jurisdictions around the world,³⁹ could potentially further boost public confidence in the protections of health-related data. However, protections are only one element in boosting public trust that the data will be put to uses that are in the public interest. Appropriate governance and transparency of business models on behalf of Big Tech are also an important foundation for building public trust in this regard.

In this hypothetical future, Alphabet could respond enthusiastically to the inevitable changes in the policy environment on the horizon. Alphabet could be a leader in establishing data governance mechanisms such as data trusts, such that the data would only be put to agreed-upon uses and appropriate profit sharing mechanisms would be put in place for health-industry applications of AI in particular. There is a particular responsibility on the side of Big Tech in this future to demonstrate to the public that it is interested in participating in processes of AI innovation that create public value. If such an effort is made, then collaborations between health care and Alphabet in this case could be situated as enabling improvements in health care while at the same time generating new revenues deemed to be appropriate. Such a future would see value accruing to all members of the innovation ecosystem, enabled by public trust and an adequate policy environment to ensure collaborations are structured in a way that protects the public interest.

³⁸ Commission Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with regard to the Processing of Personal Data and on the Free Movement of such Data, and Repealing Directive, 95/46/EC, 2016 O.J. (L 119) 1, 34 (General Data Protection Regulation). (Accessed on 01.07.2020 at: <http://www.privacy-regulation.eu/en/article-4-definitions-GDPR.htm>).

³⁹ Bari, O'Neill. *Rethinking patient data privacy...*

Discussion

In our discussion section we summarize what we see as the key challenge in building collaborations between Big Tech and health care that are supported in general by a wide range of stakeholders. We describe the importance of focusing collaboration on generating public value, and suggest that if Big Tech is going to be meaningfully involved in AI innovation for health care in the long term, an emphasis on the concept of public value will be essential. We then describe what we see as key changes in the evolving policy environments in which such collaboration takes place, briefly reviewing the most relevant shifts in data protection policy frameworks. We conclude by placing this effort into the context of global health ethics, outlining solidarity as an ethical principle that will ultimately drive the ethical viewpoint on collaborations between Big Tech and health care for many stakeholders around the world.

Public Trust and Public Value

Insights are emerging through public polling around the world regarding the perceptions of members of the public about AI innovation more generally. An IPSOS poll conducted in 27 countries for the World Economic Forum in 2019 found that 41% of the 20,107 respondents were worried about the use of AI.⁴⁰ 48% of respondents agreed that the use of AI by businesses should be more restricted, and responses to all questions were highly consistent across age, income, education and gender.⁴¹ Furthermore, dialogue about trust in AI technologies among health care providers is ongoing,⁴² illustrating a clear divide between the opinions of stakeholders (health care providers and the public) and the persistent activities of Big Tech seeking entryway into health care markets.

Outcry from the public in response to collaborations between Big Tech and health care has been a primary impediment to Big Tech's involvement in the health care market,⁴³ and to progress in the development of AI innovation for health care. Earning public trust therefore

⁴⁰ Boyon N. *New Global Poll: Widespread Concern about Artificial Intelligence*. 2019. (Accessed on 01.14.2020 at: <https://www.ipsos.com/en-us/wef-artificial-intelligence-press-release>).

⁴¹ *Ibid.*

⁴² Decamp M., Tilbert J. *Why we cannot trust artificial intelligence in medicine*. The Lancet Digital Health, 2019; 1(8): PE390.

⁴³ Powles, Hodgson. *Google DeepMind...*

appears to be an essential prerequisite to the establishment of positive collaborations between Big Tech and health care and the meaningful innovation that would ensue. Establishing this trust obviously requires an alternative approach to building collaborations that conveys the interests of Big Tech more clearly and creates space to clearly communicate its role in advancing *public value*.

Although the concept of public value is contested,⁴⁴ we use it here simply to refer to the enhancement of the interests of members of the public over and above the stakeholders affiliated with a given organization. We suggest that what has been missing from collaborations between Big Tech and health care are clear statements regarding the ways in which such collaborations will contribute to providing public value. If the public is to trust that these collaborations will protect their interests and contribute to public wellbeing, members of the public must be made aware that these are primary considerations when their health care data is at stake. A future where Big Tech is deeply involved in AI innovation for health care seems to rely on this important point. The intersection of public trust, public value, and collaboration between Big Tech and health care offers an important direction for future research.

Evolving Policy Environments

Health information and data protection policies are evolving, with policy frameworks such as the GDPR and CCPA being viewed as much more restrictive than past policy frameworks in relation to personal data. Three shifts in the logic of data protection policy are most relevant to our analysis here. First, as summarized earlier, definitions of health data are broadening away from merely including the data stored in health care records to include any data that could be considered to indicate a person's physical or mental health. Second, the focus of these policies shifts away from named organizations as custodians of the information toward the data themselves, such that any organization collecting, storing, sharing, or using the data is responsible for administering the necessary data protections (i.e., not just named custodians). Third and finally, more sophisticated understandings of de-identification are embedded in such policies, acknowledging the

⁴⁴ Alford J., O'Flynn J. *Making sense of public value: concepts, critiques, and emerging meanings*. International Journal of Public Administration, 2009; 32(4): 171-191.

increasingly sophisticated approaches for re-identifying data that exist using current methods.⁴⁵

These changes in contemporary approaches to data protection policy might very well have the effect of enhancing public trust that regulators are capable of protecting their data more generally. However, there is a substantial gap between (a) adequate protections over data, and (b) public trust in Big Tech that supports its role in AI innovation for health care. Floridi (2018) suggests that such regulation is important, but that "soft ethics" must fill the remaining gap "by considering what ought and ought not to be done over and above the existing regulation".⁴⁶ Regulation is fundamentally a blunt instrument, and if Big Tech is going to earn public trust it must do so by demonstrating ethical decision-making as it fills the space between regulatory requirements and strategic or operational activities. Regulation cannot force Big Tech to work toward creating public value; companies must advance that agenda on their own accord. However, without a message about creating public value, it is unclear how public trust will be earned to support Big Tech's role in AI innovation for health care.

Global Health Ethics

Surveillance capitalism must be understood as fundamentally a global phenomenon, and the reach of companies like Alphabet certainly does extend all around the world.⁴⁷ To consider the analysis we have presented in this paper from a global health ethics perspective requires that we assess the potential global implications of the deeper involvement of Big Tech in health care from a global perspective. Aligned with the possibilities described in future #2, health systems outside of the United States will feel obvious pressure to remain current with the innovations implemented there. This means there will be pressure to purchase AI innovations developed in the United States for use in health care systems elsewhere around the world, many of which have far fewer resources to support such advanced technology procurement activity. Many countries will be unable to absorb the costs, which would presumably lead to widening gaps in the capabilities of health systems

⁴⁵ Hintze M. *Viewing the GDPR through a de-identification lens: A tool for compliance, clarification, and consistency*. International Data Privacy Law, 2018; 8(1): 86-99.

⁴⁶ Floridi L. *Soft ethics and the governance of the digital*. Philosophy of Technology, 2018; 31: 3.

⁴⁷ Zuboff. *The Age of...*

between high-income and low and middle-income countries around the world. Widening gaps in health system capability could potentially have onward consequences in widening other inequalities as well.

In their seminal work on global health ethics, Benatar et al (2003) proposed that solidarity is the most essential moral value for a meaningful global health ethics.⁴⁸ They elaborate that, “without solidarity it is inevitable that we shall ignore distant indignities, violations of human rights, inequities, deprivation of freedom, undemocratic regimes, and damage to the environment”.⁴⁹ A global health ethics perspective on the role of Big Tech in AI innovation for health care would then advocate for inquiry into the ways in which an orientation toward public value might rest on a foundation of solidarity, considering what is good for the world as opposed to what is good for a more narrowly defined public (i.e., “Americans”).

The critiques of Big Tech and their role in health care arise largely from people who have solidarity as a priority and are oriented first and foremost toward the global good.⁵⁰ Ethical critiques of Big Tech will continue to be produced from this perspective, assessing the interests of Big Tech against the interests of the broader global good. Health care represents one field in which these interests play out, and if Big Tech is to win over public trust and public support for its role in health care, establishing a response to the demand for solidarity is an essential step forward. Given the circumstances of surveillance capitalism and the foundational drive to grow revenues, this represents a tall order. However, if a meaningful role for Big Tech in AI innovation for health depends on public trust, and public critiques at least in part arise from a place of solidarity, then this should amount to an important focus for Big Tech in the years to come.

Conclusions

In this paper we have outlined three potential path dependencies related to Alphabet’s involvement in AI innovation for health care, and proposed four hypothetical futures that might arise given these path dependencies. We proposed the central importance of the concept of

⁴⁸ Benatar S, Daar A, Singer P. *Global health ethics: the rationale for mutual caring*. International Affairs, 2003; 79(1): 107-138.

⁴⁹ *Ibid.*, p. 117.

⁵⁰ Sharon T. *The Googlization of health research: from disruptive innovation to disruptive ethics*. Personalized Medicine, 2016; 13(6).

public value, and suggested that if Big Tech is going to earn public support for their role in health care, then demonstrating their interest in promoting public value is essential. We concluded with a comment on global health ethics and the tension between the profit-generating motive of Big Tech and a perspective on the global good that is informed by the value of solidarity. Ultimately, we believe that Big Tech will continue to work toward advancing its role in health care on a global scale, and that establishing a way to incorporate a stronger focus on the global good will both enhance the ethical status of Big Tech's work and enable a more straightforward path to the development and deployment of AI innovation in health care.

The Secondary Use of Health Data in the Context of Big Data and the New European Legal Framework: Have We Changed the Helsinki Paradigm?

Federico de Montalvo Jääskeläinen *

Big Data offer a number of opportunities and alternatives in general, but most specifically in the field of health research. The extensive use of conventional health data and even their interlinking with non-traditional data shall boost the fight against disease, while improving prevention and diagnostic capabilities in unparalleled and unprecedented ways in the history of medicine and humanity. The results extracted from data use that took decades to obtain only a few years ago can now be revealed within months, even days, and what is more, at a very affordable cost. Algorithms enable the comparison of a large number of healthcare processes, thus offering accurate conclusions, in terms of volume, on the best treatments for every disease and the relative diagnoses.

The German Ethics Council (*Deutscher Ethikrat*) pointed out that in biomedical research, the analysis of large volumes of health data should provide a better understanding of important scientific processes and their connections. Among the most data-intensive applications are modern imaging and molecular biological procedures, such as those employed in what we call 'omics' (e.g. genomics or proteomics).¹

This concept pervades Regulation 2016/679 of the European Parliament and the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (hereinafter referred to as the EU Regulation). In fact, Recital 157 of the Regulation explicitly notes, "By coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovas-

* Associate professor of Constitutional Right, Universidad Pontificia Comillas; President of Spanish Committee of Bioethics (Spain).

¹ German Ethics Council. *Big Data and health. Data sovereignty as the shaping informational freedom.* Opinion, Executive Summary and Recommendations. 2018, p. 9.

cular disease, cancer and depression. On the basis of registries, research results can be enhanced, as they draw on a larger population. Within social science, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions. Research results obtained through registries provide solid, high-quality knowledge, which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services. In order to facilitate scientific research, personal data can be processed for scientific research purposes, subject to appropriate conditions and safeguards set out in Union or Member State law.”

Thanks to the massive use of data, results analysed collectively have a different value than results analysed individually. As Professor Vanesa Morente highlighted very clearly, the use of Big Data brings a deeper, and more significant insight, that goes beyond the obvious, like an onlooker who may spot a human face in Giuseppe Arcimboldo’s still lives only contemplating them as a whole, otherwise a mere conglomerate of separate and assorted fruits and vegetables will meet the eye. The primary purpose of Big Data entails therefore a look that not only sees, but also discovers: it is a transformative look that sees value in raw, unprocessed information.² In the medical and health contexts in general, this has an unquestioned value, because unlike other research fields, this sector attaches a particularly relevant value to the quantitative method, even though, as they say, there are no illnesses, but rather patients. In any case, in order to improve medical treatments further, the opportunity to correlate millions of healthcare processes is fundamental, in that these results shall later be contextualized and personalised.

Yet, in view of the definition given by the World Health Organization a few years ago, whereby health is “*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*”, a holistic approach to health should blur the line between health seen from the medical perspective and lifestyle. Big Data provide the technical opportunity to support such a holistic vision, as they do not confine data use to strictly or traditionally health-related data, such as

² Morente Parra V. *Big data o el arte de analizar datos masivos. Una reflexión crítica desde los derechos fundamentales*. Revista Derechos y Libertades. 2019; 41 (época II), p. 2.

clinical records, but also integrate data on a person's lifestyle, habits and even environment.

Consequently, it can be stated that clinical data are no longer a mere reminder of the healthcare process, but rather the main source of knowledge and progress in Medicine and Biology. Health data can already be considered as the true *treasure* of biomedical research, as opposed to biological samples, i.e. the *treasure* of the previous decade. In the healthcare field or in a specific clinical trial, data are not strictly of interest as documentary evidence of the most relevant facts concerning the treatment provided, treatment-related decisions or the diagnoses and conclusions reached, but for their secondary use, that is independent from the main purposes for which those data were initially provided. Patients contribute their data for a specific purpose and such data can be useful for a secondary purpose, or use, enabled by the tools offered by Big Data.

Moreover, the opportunities offered by the extensive use of health data become even more relevant in healthcare systems characterized by both essentially public management and care provision schemes (the Beveridge formula) and the recent process of digitalization of documents and clinical records, that helped introduce millions of data in a single or, at least, in easily comparable databases. Consequently, we are not exclusively referring to data extracted from research projects on humans or clinical trials, but to the secondary use of health data, which is more significant in terms of numbers and, possibly, value.

Despite the above-mentioned relevance of Big Data in general and in healthcare in particular, the European legal framework has not issued any specific regulation. It is true that the European Union has already equipped itself with a very complete regulation on personal data protection, several parts of which may apply to Big Data, however this is a new reality that may require more specific solutions. Therefore, the problem is not a dearth of general regulations, since several legal conflicts and dilemmas are legally covered by the data protection regulation, but rather of specific provisions and perhaps new principles apt to govern the innovative characteristics of Big Data.

Furthermore, this unprecedented scenario unfolds at a time when new uncertainties grow about the evolution of several diseases, which although very well known, such as cancer, offer new paradigms of

cell and protein development, as well as of new diseases, many of which are untreatable yet. The interactions among the determinants of countless diseases are highly complex. Big Data enable researchers to integrate and aggregate information from across multiple sources. The opportunity is therefore undeniable from the perspective of the protection of life: this requires that we discard an *a priori* approach that conceives health-related data processing negatively.

However, there are risks to personal rights, as there are opportunities. This is why the German Ethics Council specified that Big Data represent a major challenge to the legal system and, in particular, to constitutional law. Nonetheless, personal information goes hand in hand with these risks, even more so in areas such as healthcare, where highly sensitive data are at stake. However, Big Data will potentially multiply these risks.³ In fact, they are not only limited to the right to privacy, since information on a person's health status can affect other rights and interests, such as access to employment, credit or insurance.⁴

In any case, the new EU regulation, while not specifically addressing the particular dilemmas and conflicts of Big Data, does contain specific references to health data and, more specifically, to the requirements for their secondary use for research purposes. We may say that the Regulation opens up a new era or even a new paradigm in this field. In fact, it replaces the model based on the alternative between informed consent and anonymization, with one based on informed consent or pseudonymization that would enable a more flexible use of health data in the interest of the community and everyone's good health.

From an ethical-legal standpoint, we do not believe that we can apply the ethical-legal principles and values developed for Big Data-driven research to traditional research projects on humans, because the rights involved in research projects not focused on individuals, but on their data differ. It is no longer a matter of affecting an individual's integrity, but rather intruding in his or her private sphere.

³ German Ethics Council. *Big Data and health. Data sovereignty as the shaping informational freedom*. Opinion, Executive Summary and Recommendations. 2018, p. 10.

⁴ Martínez R. *Big data, investigación en salud y protección de datos personales ¿Un falso debate?* Revista Valenciana d'Estudis Autonòmics. 2017: 62, p. 236.

Furthermore, one might ask whether a regulatory model, based essentially on an individual's interest, still responds to citizens' desires. In this respect, some works have already shown that citizens do not oppose data sharing; on the contrary, as Haug stated, patients want their data to be shared quickly, especially to ensure that other patients may learn of any possible treatment-related adverse events. At the same time, they also want to retain some control on how the data are shared, particularly when the research purposes are essentially commercial and not so much when public health systems seek to use data to improve medical treatment or care for other patients. In fact, receiving medical care invariably involves a loss of privacy. Patients must disclose their personal information to obtain help, and that help generally derives from knowledge gained from the experiences of previous patients who disclosed their personal information. The problem is not so much in the use, but in the demand for responsible use.⁵

Obviously, this cannot mean prioritizing collective interest at the expense of individual interest, but rather seeking a balanced formula to integrate the two. This formula can be worked out when we safeguard the rights of the individuals involved by adapting one of the two requirements that the new legal model of data protection seeks to accomplish, i.e. anonymization through the new mechanism of pseudonymisation.

What is relevant in this new model is not so much an individual's prior consent to the new purpose for which data are intended or strict data anonymization. In fact, what matters is the legitimate origin of the data, the great importance of their secondary use for community health and the adoption of sufficient measures to prevent non-authorised third parties from gaining access to an individual's identity through the data, without necessarily demanding any strict anonymization. This seems to be legally achievable through what is commonly named pseudonymisation, defined by the EU Regulation as the processing of personal data in such a way that they can no longer be attributed to a specific individual without the use of additional information, as long as that such additional information is kept separately and subject to technical and organisational measures of non-attribution to an identified or identifiable individual.

⁵ Haug CJ. *Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate?* NEJM, 26th April 2016, pp. 1-2.

The advantages of pseudonymization over traditional, strict anonymization are clear from the standpoint of community health. In fact, interlinking the data to the person, even when it is extraordinarily difficult for a third party to decode them, means not only to broaden the data used in a research to include other initially insignificant data (data enhancement), but also to corroborate the results of data use with the patients' real progress (results verification), for example. And this is very relevant in today's Big Data science. Pseudonymisation is, in the end, the only guarantee against the previously mentioned misleading causalities that are one of the main risks of Big Data.

In short, against this backdrop of great opportunities in the fight against disease and in the improvement of people's health, it is important to promote new paradigms that do not present technology only as something essentially good that totally excludes human intuition and wisdom. In fact, such models should not neglect that the context has deeply evolved over the years and the enormous advantages of massive data processing must go beyond a vision exclusively based on individual interest at the expense of the common good. As it is in many other areas, true virtue seems to assert itself as the centre of gravity between the two approaches.

Furthermore, the debate must be addressed so as not to lose sight of the context. In the health protection models developed in Western Europe after the Second World War and, above all, in those based on a social-democratic formula such as the Beveridge model, it would be contradictory to maintain a position only taking into account the individual or the subjective dimension, in that those models feature essential traits of communitarianism. Going to a hospital and having a serious health problem solved thanks to public spending demands that citizens exercise their responsibility that is manifested, in the current context of technological progress, in the moral duty to share their data so that others who have not been so easily and readily treated can benefit from medical care.

(Translated from Spanish by the Pontifical Academy for Life)

UNESCO's Perspective

Cédric Wachholz*

I am deeply honored to represent UNESCO here, at the Vatican, and to be part of this important reflection on the ethics of Artificial Intelligence.

Too often in the field of technology, advancements are analyzed in terms of their potential to increase efficiency gains or profit margins. As an intergovernmental organization with a mandate to achieve peace through international cooperation, UNESCO measures technical progress differently. We assess technologies in terms of their potential to increase the well-being of individuals and to achieve sustainable development, of which economic development is just one part along with social and environmental development. As such, we are heartened to witness a growing discussion around the ethics of AI and welcome views of diverse stakeholders into the discussion.

Regardless of our backgrounds, our concern about the use of AI for the good of humankind unites us all and brings us here today.

My presentation will speak about UNESCO's approach, about how we see AI opportunities, how we are addressing challenges and I will close with a call and an invitation to you all for action.

Let us start with seizing AI opportunities, at the United Nations Educational Scientific and Cultural Organization (UNESCO), we aim at harnessing AI for development, for the SDGs in the fields of education, sciences, culture and communication and information.

How do we do that? We help in building human and institutional capacities, we advise governments on policy development, we set global standards and we are a laboratory of ideas that facilitates exchange of knowledge and innovation while also facilitating international cooperation.

* Chief of Digital Innovation and Transformation Section, UNESCO.

In the field of AI for education, there are a large number of AI applications, where we work with also with private sector companies. More specifically, we work on three aspects: on enhancing teaching and learning, on educational management and information systems, but also on 21st century skills, on competencies required in an AI-enhanced world.

For culture, I will also just give you a short example of how AI can preserve existing languages and bring them to wider audiences. A significant benefit it has for indigenous and minority communities: we have worked with universities in the EU and they have developed an AI-powered automatic speech corpus annotator for African speech corpora, which serves for quicker translation. It really works quite beautifully, you can take pictures of an object, you can annotate it in a local African language and it can be later translated.

In natural sciences: AI can improve data collection through low-cost, low-powered sensors that are being deployed in remote regions for a variety of applications. It is UNESCO's global networks on for water and development are leveraging AI applications for information collection concerning weather systems in arid areas and then use it to produce estimates of real-time precipitations worldwide. These estimates are not only used for research, but for disaster risk reduction and to inform emergency planning, reduce damages and to save lives.

In communication and information, AI is used to strengthen the flow of information and facilitate the production of news. However, it can also weaken journalistic institutions by disseminating fabricated content with harmful intent and amplifying such disinformation. Journalistic diversity is also under threat if development of AI facilitates migration of advertising to data-rich internet intermediaries and we see how a lot of newspapers are suffering from that.

I could of course not speak about AI challenges from a global perspective, without looking at the existing disparities in terms of digital and knowledge divides. Many of you might know or be aware that 49% of the world's population does not even have internet access, and only 20% have internet access in the least developed countries. This means that AI is really far away from their reality and even more so, if we look also at the percentages of individuals who have the skills to use and develop AI in a world, in which AI talent and brain drain are even increasing the existing disparities.

Now, I will not go into the privacy debate, we have seen a lot of in-depth discussion on that AI challenge and I will directly shift into the biases. There is a growing awareness that current AI systems tend to expose and amplify social inequalities and injustice. What happens is that AI is trained on biased data sets that can enhance and proliferate those biases in its outputs, leading to discriminating applications.

Gender is one of UNESCO's two global priorities and so I wish to emphasize here the gender dimension, which goes not only through algorithmic discrimination, but there's also a lack of representativeness: within the AI workforce, there are challenges like the deep-fake pornography targeted at girls and women, but also the reproduction of stereotypes, to name a few dimensions. The "black box" problem of AI is another concern: the lack of explainability, transparency and human control makes many deep learning systems function largely as black boxes. Therefore, their behavior can be difficult to interpret and explain. Accountability is a comparable problem with increasingly complex systems, which include human beings and act with multiple components, multiple data sources, therefore, it is increasingly difficult to attribute causality and also responsibility as one can often see in discussions about autonomously driving cars.

In view of time, I will not be going in much depth on the question of trust, but it is clear that often humans perceive AI as superior than their abilities and can over trust it. This can have catastrophic consequences, for example, in the judiciary area. However, one challenge that is comparable to other technologies too, is portability across borders, where AI is developed and deployed in multiple jurisdictions, and then used across international and cultural boundaries. Consequently, the pace of international tech innovation often outpaces regional and national policy responses. As my last point, I want to emphasize the importance of reflecting upon and finding solutions to problems of disinformation, echo chambers, the impact on human rights, elections and democracy challenges that have also been raised by other participants today.

Now, you might ask yourself in view of these numerous challenges: what does UNESCO do? UNESCO's Member States actually decided to have a broad approach not only to AI, but to the whole technology field, also to internet governance and other fields, which is a human rights-based approach, and it is based on openness, on access and

accessibility, as well as on a multi-stakeholder approach. Therefore, the governments in UNESCO decided that they cannot simply decide and operate in this domain, there needs to be a multi-stakeholder approach, including on the important question on what is right and wrong, if we want a human rights-based, ethical approach. If we look at the total of existing ethical frameworks, we have today more than 80 different sets of principles compiled by the Berkman Klein Centre at the University of Harvard. However, what is interesting is that these principles have been developed by different stakeholders including, governments, civil society, technical community, international organizations, and the private sector but, in terms of geographical representation, they are mostly developed in only one part or certain parts of the world. A 2019 study shows that none of the African or South American countries had developed these ethical framework, while the United States alone and Europe had 19 documents. There is a need for a unified global unified approach, which does not exist so far. There is no UN framework. In November of 2019, UNESCO was asked by its 193 Member States to develop a standard setting instrument, a Recommendation on the ethics of AI through an inclusive and multistakeholder process. It will take us two years to have consultations in all the regions and to have a multi-stakeholder and open approach. This is the call for action I wanted to close with, an invitation to all of you.

We will have the first draft of the ethics recommendation ready in mid-May and it will also be available online, that is the red box on the powerpoint you see, where everybody will be invited to join and comment and contribute to this process, which will continue for another one and a half years with a number of meetings. I will not go into the details of this process.

However, there are many frameworks, what will be the difference at the end when the Recommendation will be adopted by all Member States? It will not be simply another declaration of principles. It will be the outcome of a standard setting process in UNESCO, which recommends different actions for different stakeholder groups, so it is not just a declaration of principles, it has a concrete link to action and it will be really, truly, diversely developed with inclusive consultations for everybody in all regions. UNESCO is confident that human centered AI can harness the potential of AI when managing the challenges I mentioned earlier. To that end, we are working on AI access and capacity development, rather than retreating to technological determinism, while facing

the challenges of AI. AI is not a silver bullet for our problems or an harbinger of more. New technologies have and always will influence our societies, but ultimately only we can shape our future.

If we want AI to be a force for good, then we must foster the development of human-centered AI. Engaging in discussions about the ethics of AI as we're doing today is a commendable start, but we also need to begin to be more proactive, responding to this request by citizens, governments, civil society and the private sector for formulating regulations, laws and policies, investing in beneficial areas of research and developing talents who can work with AI.

UNESCO stands ready to participate and lead more such efforts, and I thank you for your attention.

AI Common Projects

Amir Banifatemi *

Good afternoon everyone.

Ladies and gentlemen, it is a pleasure to be here today and share with you a project that is very dear to us.

I hope I shall be able to convey to you the work that we have done with a number of partners and its impact on the world of AI, in general.

This is a question that the company I am working with has asked itself a lot: we tried to address the world's greatest challenges, whether they are disease eradication or outbreaks, or cleaning plastics in the ocean, or even going to space, or helping people be educated, or giving access to health to everyone. As those big topics are top of mind, a few years ago we asked ourselves how artificial intelligence could help humanity in addressing these challenges.

That question led us to create a grand challenge competition in 2016, in association with IBM Watson, to reward, with a large symbolic sum, teams that could demonstrate that they could develop powerful AI systems for collaborating with humans and tackling the world's greatest challenges.

We left it open to participants to define the challenges that they were trying to solve, which is quite uncommon in the world of challenges. This competition led us to have 10,000 respondents from 84 countries, which was unheard of in our experience.

Briefly, XPRIZE is a non-profit organization that launches grand multi-viewer challenges. Usually, we get 200-300 teams competing for a given challenge, whether it is capturing carbon, or finding ways to educate children without a teacher, but in this case we opened up the problem and 10,000 applicants registered, the number of teams was then narrowed down.

At the moment, we still have 5 teams competing to win that challenge, that is going to be awarded in April. Interestingly, we observed

* General Manager for Innovation and Growth, IBM Watson AI XPRIZE (USA).

that all the team's projects concerned different topics: this is the voice of people, who are trying to solve a specific problem; this is not something academic or a research project. It is coming from individuals, teams, start-ups, or groups of researchers that are trying to show what they think that they can solve with AI.

In fact, you can see that there is a distribution on health, there is a distribution on the environment, on law, on general business improvement, on science, disaster recovery, even humanity of AI and so forth.

That gives us an indication that the problems needing a solution are not really in line or in fact, they are orthogonal to what the industry is working on: we had no one talking about self-driving cars, financial automation, or cybersecurity.

These are the topics that were presented to us: some examples of interesting projects that some teams were working on refer to malaria eradication, beehive monitoring, music generation, detecting earthquakes, or helping people not be affected by hate speech in chat rooms or social media, or financial services to low-wage workers. These are all the topics that emerged from the ground up, these and 840 other emerged and competed in the challenge.

As we observed this, we tried to map those problems, and in doing so, we attempted to cluster those responses and their relative domains; we understood that, in fact, these problems are really aligned with the sustainable development goals.

These goals are the 17 UN SDG goals that you know about and this helped us invite and convene four years ago a global gathering called AI for Good Summit organized by ITU and XPRIZE, my company, with partners active in civil law.

We went from about 400 attendees to more than 3000 attendees in a matter of a few years and the goal was to bring stakeholders and academics together with investors, impact foundations to see and discuss how we could use AI for good, what good means and how we could scale this up.

This event was the origin of many projects.

Let me give you one example just to show that we had initially 700 use cases identified at this summit by the grassroots level. One example is the use of satellite imagery to predict, for instance, deforestation or other things, which are, believe it or not, important issues for many countries or the improvement of micro-insurance provision to crafts, but we have many other examples.

We learnt that the challenges to scale AI for good are the following: it is not obvious to identify real problems, we are all considering the same problems in their many localized, acute, and relevant manifestations, several people are striving to solve these, but they have no way to present those issues and bring them up.

Data as a foundation for machine learning is not accessible everywhere and accessible data is not always of quality, its privacy guaranteed, or it is not ethically sound or unbiased.

There is, as you know, a shortage of talents, so providing talents to help with those problems is not always obvious and access to frameworks and known solutions is not easy.

Therefore, if the solution has been developed, let's say, in a university in London, the outcome of that may be an article or a publication, but there is no way to implement that solution at scale and make it to the production level or bring it to people to use and to leverage.

Finally, let me borrow the term that my friend Selwick shared, there are so many principles, so many policies, but there is no set of governance and policy to support validating AI for good and scaling it up.

Faced with those lessons, after 4 years we discovered that there is really an interesting way to look at how the world of AI for good is implemented. On one side, we have a group called problem-owners, it could be an academic, it could be a city manager, it could be an advocate, a policymaker, a farmer, a mother at home endeavoring to solve a specific issue. On the other, you have engineering talents, researchers and problem-solvers, who today are knowledgeable about data and machine learning, and are trying to get faster, but there is no connection, there is a gap between these two groups.

They do not speak the same language, they do not communicate with each other so how can we help bridge that gap?

We evaluated what is needed to do that.

There is, of course, a need for data, and data-owners to make that data available; there is a need for storage and computing capabilities to help machine learning be implemented and made available and we need to connect these together; and when solutions are identified, there is a need for funding, because otherwise, it would become an on-shelf publication, or a showcase only.

However, this is not enough, we also have to think about issues of ethics and privacy, we have to think about what our governments will

support, and about policymakers who are trying to survey data, and then research will follow.

This complex web of stakeholders needs to align and work to make sure that if AI must be used for social good, for benefiting people, there is a framework to accompany, not only validate, solutions, but also to have them scaled, governed and made sustainable. The creation of AI Commons as a group came out of this observation and the glory of AI Commons is that it is a collective effort.

Think of it as Wikipedia or Linux, where we are striving to unify our efforts, to create a knowledge hub, to collaborate and solve problems together.

This is really the outcome of what we have learnt and how we can share knowledge by making it available to others.

How does it work? It works by creating three steps: one is that we started by creating a knowledge hub, a repository of cases and uses of solutions and ensured where they are applicable, where they are working, who uses them, the context for their sustainability and the policies supporting them.

That knowledge today is not pervasive, it is not available always, that is why we are attempting to create an open-source knowledge, but we intend also to find a way through that knowledge hub to help problem-owners and problems-solvers match each other, and find each other because there is always a good way to do that. Once they find each other, there is a need to collaborate and this collaboration needs to be safe. None of us is necessarily fully aware of ethical concerns, or a specialist in ethics or philosophy or privacy or law.

How can you participate safely, without the encumbrance of making mistakes or generating biases: this safe environment enables people to collaborate, so they can evaluate solutions, and then somehow get funding and scale things up without any fear of biases or ethical concerns, but with trust.

Finally, we have to promote sharing resources, because if resources are not shared, as we saw in the previous presentation, there would be a huge gap between haves and have-nots. This is not only about money or resources; it is about the speed at which we are using AI to develop solutions that will benefit the industry. As you saw in my previous slides, we had at least 10,000 applicants for this competition alone, coming from the ground up, who intended to use AI for very specific problems that are life, so how can we solve those problems?

AI Commons was created by an umbrella of initiators from different groups, mostly researchers and an activist, Francesca Rossi, who is an outstanding member and is actually here as well, I am pleased to have her. A number of groups are supporting it by providing resources, talents, knowledge and data to create this knowledge repository and make it operative. The goal that we have at AI commons, which is a collective, not an organization, or a business, is to promote the following idea: AI solutions have to integrate the diversity of people that come up with the problems, if we do not integrate them in the solution design, the solutions presented may not fit their needs.

How can we do that, how can we ensure that they are localized, they are compliant with the culture, compliant with the policy of different places? How can we scale up solutions, but at the same time preserve diversity and be inclusive, how can you be inclusive and make sure that everybody has access and the right to play that game?

For us focusing on problem-owners, on the terminal problem owners is important, that is why I think about empowerment, which is an abused and used word, but empowerment is relevant.

At the end of the day, you want to give agency to individuals to solve their own problems and today we have done it with the internet, we have done it with many tools, we have done it over time, historically we have examples of that.

However, when it comes to AI, there is really a window of opportunity to make sure that everyone plays the game and really participates.

We work towards achieving the common good, good is a concept that is not well defined, it is ill defined and common good is what we are trying to strive for.

What if we talked about AI for good and transitioned from AI for good to the concept of AI as common good? What if AI needed to become part of the natural resources that we give if the new world vision had to incorporate AI and if you had to pronounce AI as common good what would that mean?

These are the questions, but we do not have the answers yet, however it seems that after observing a few hundred to close to a thousand teams globally in many countries, we noticed that there is an urgent need to give access to AI and machine learning technology, and to the data available to much larger population.

I used to use a paraphrase about pictures: if we took a famous photographer's picture and an amateur's photo taken with a smartphone,

most people would not be able to tell the difference, only a highly trained eye might distinguish the differences.

Therefore, technology in the hands of many could probably become a tool for democratizing creativity and finding new pathways that we have not been necessarily aware of. The journey that we started, and we are delighted that the Pontifical Academy is one of the partners, is a journey to gather enough knowledge collectively.

Wikipedia has the most data, and is useful, how can we define together a set of knowledge that we can share collectively for everyone to use and participate in solution design and deployment?

Therefore, the goal is to empower everyone to participate in the betterment of society.

If we believe, and I do, that anyone has the power to change the world, has the power to participate, given the machine learning tools and AI as predictive tools, as tools to empower everyone to think better, to think faster and to participate in proactive modes, then, that might be a beginning.

Thank you very much.

A new RenAIssance for the future of Education

Francesco Profumo*

Introduction

Artificial Intelligence (AI) is poised to transform the modern society in a profound way. As with electricity in the last century, AI is an enabling technology that can animate everyday products and systems, endowing everything from cars to smartphones with new capabilities.¹ Many applications of AI are already part of our life – from machine translation to image recognition and music generation – and are increasingly deployed in industry, commerce, and government. Not surprisingly, autonomous vehicles as well as AI-supported medical diagnostics are areas of application that will soon be commonplace.² From the technological standpoint, these recent developments in AI are the result of a relentless acceleration of computing power (epitomized by Moore's law), unprecedented improvements in algorithms, and an exponential growth in the volume, variety, and velocity of digital data that IoT makes available for training purposes.

As AI matures, it is about to have sweeping impacts on our skills, work, and welfare.³ Digital technologies are doing for human brain-power what the steam engine and related technologies did for human muscle power during the Industrial Revolution:⁴ as a consequence, whilst the early age of automation disrupted blue-collar workers, new

* Full Professor, Polytechnic University of Turin, Department of Energy, Turin (Italy).

¹ Brookings. *How artificial intelligence is transforming the world*. 2018 (Accessed on 01.29.2020 at: <https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world>).

² European Commission. *Artificial Intelligence: A European Perspective*. 2018 (Accessed on 01.29.2020 at: <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf>).

³ McKinsey Global Institute. *Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages*. 2017 (Accessed on 01.27.2020 at: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages>).

⁴ Brynjolfsson E, McAfee A. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company: New York; 2014.

AI capabilities are rapidly changing white-collar job roles. In view of the impact of digital technologies on labor income, as pointed out by Bruegel,⁵ one of the big challenges of the twenty-first century is to redefine the nature and functioning of welfare states in the context of the fundamental changes ushered-in by automation. Along these lines, when it comes to welfare, taxing robots (more precisely, their owners) is seen by several analysts as the tool to finance the basic income – either universal or subject to conditions – and, thus, tackle income inequality.⁶

In a time when algorithms so massively affect our way of life, socio-economic, legal, and ethical implications have to be carefully addressed. Of primary interest are, in particular, the multi-faceted implications of AI in the process of human learning, which is seen as the instrument of social change par excellence. As observed by UNESCO,⁷ one of the toughest challenges posed by AI is how to rethink and rework educational programs to prepare learners for the increasing presence of AI in all aspects of human activity.

Acknowledging that the past is not necessarily a guide to the future, the present paper casts a light on the intersection between AI and education with the aim of illustrating challenges and opportunities that AI poses to the education system going forward. In more detail, three distinct yet interrelated perspectives are investigated, namely (1) AI for Education, (2) Education for AI, and (3) Education to AI.

AI for Education

A first angle to consider is that of 'What AI can do for Education today'. Access to education, at various stages, has already been transformed through e-learning, which has democratized the access to courses worldwide, changing the educational offer by public institutions and creating a new market for private actors.

⁵ Bruegel. *Digitalisation and European welfare states*. 2019 (Accessed on 01.28.2020 at: <https://bruegel.org/2019/07/digitalisation-and-european-welfare-states>).

⁶ European Commission. *What are the policy options? A systematic review of policy responses to the impacts of robotisation and automation on the labour market*. 2019 (Accessed on 01.28.2020 at: <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/what-are-policy-options-systematic-review-policy-responses-impacts-robotisation-and>).

⁷ UNESCO. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. 2019 (Accessed on 01.27.2020 at: <https://unesdoc.unesco.org/ark:/48223/pf0000366994>).

AI is playing an increasingly important role in building the future of teaching and learning. In general, AI is seen as an opportunity to optimize the provision and management of education, empowering teachers, enhancing new learning outcomes, expanding access to schools and improving educational quality provided to an increasing number of learners at all levels, especially for more vulnerable groups who have socio-economic challenges, specific educational needs or no access to formal education.

Research has found a correlation between high levels of income inequality and low levels of social mobility. Education is the main key to foster socio-economic mobility, climb out of poverty and compete in a global marketplace.

According to the UNESCO Institute for Statistics,⁸ about 258 million children, adolescents, and youth were globally out of school in 2018. The worldwide out-of-school rate was one in ten children (one in five in sub-Saharan Africa) in 2018. An estimated⁹ 617 million children and adolescents of primary and lower secondary school age – more than 55 percent of the global total – lacked minimum proficiency in reading and mathematics.

From this perspective, the latest AI technologies are enabling new opportunities to remove issues and inequalities in the education field, in order to accelerate the achievement of the Education 2030 Agenda.¹⁰

For instance, AI can remove linguistic and logistic barriers, enabling to learn from experience what type of educational approach is more effective for different students and presenting material in a form that better meets the current level of preparation.

Through the power of machine learning, AI systems are being used to overcome the one-to-many education model, to tailor and personalize learning for students, based on their ability, needs, and experience: advanced techniques including AI-powered chatbots, smart notes, flashcards, virtual facilitators and real-time feedback among others are changing the way things are done in education.

⁸ UNESCO Institute for Statistics. *Education theme*. (Accessed on 01.23.2020 at: <http://data.uis.unesco.org>).

⁹ United Nations. *The Sustainable Development Goals Report*. 2019 (Accessed on 01.23.2020 at: <https://unstats.un.org/sdgs/report/2019/goal-04/>).

¹⁰ United Nations. *Education 2030, Incheon declaration and framework for action for the implementation of Sustainable Development Goal 4*. 2016 (Accessed on 01.29.2020 at: <https://unesdoc.unesco.org/ark:/48223/pf0000245656>).

In addition to customized materials, AI systems can significantly improve tutoring with personal, conversational education assistants. These autonomous agents answer questions from students, provide assistance, assign homework and reinforce concepts that can help learners to upgrade their curricula.

Finally, educators often spend up to 50 percent of their time on non-teaching tasks: AI systems are helpful at managing these back office and administrative activities, offering teachers more time to focus on educating their students or to engage in research pursuits.

Another aspect to consider is that AI can analyze large amounts of data collected from the individual student and other students in the classroom: combining an AI-based system with student data, teachers can identify from a global database the best material and strategies to use and find the most effective teaching interactions and learning patterns.

Our society needs to promote the integration of AI and education to bring about a transformation in education. Both formal education and lifelong learning should be reconsidered in order to introduce innovative teaching methods that can help students and graduates in developing new creative skills that are becoming increasingly crucial.

As the education level of a country is strongly correlated with several indicators of wellness, it is vital that not only some countries will be able to benefit from these opportunities, but also that access to educational tools empowered by AI will be guaranteed to citizens worldwide.

The new decade provides an important window of opportunity for policy-makers to ensure that the quality of education and training systems is improved and that more people of all ages can access them.

Education for AI

A second angle to consider is up to which point modern countries are ready to meet the fast-growing demand of experts in the field.

Some studies emphasize that AI will deliver an additional economic output: a study by consulting firm Accenture¹¹ estimates AI could

¹¹ Accenture. *Why Artificial Intelligence is the future of growth*. 2016 (Accessed on 01.29.2020 at: https://www.accenture.com/t20170524T055435__w__/ca-en/_acnmedia/PDF-52/Accenture-Why-AI-is-the-Future-of-Growth.pdf).

double global economic growth rates by 2035, as a result of a strong increase in labor productivity (by up to 40 percent). PwC¹² predicts that global GDP could rise by 14 percent by 2030: this will mainly originate from substitution of labor by automation and increased innovation in products and services.

The impact of AI will be especially visible in the labor market, changing the nature of job competencies and transforming the current methods of training and future workers. Even if this scenario will bring a wide range of benefits in terms of higher economic growth, improved corporate performance and new prosperity, several scientific reports have predicted that a growing interaction with robots and smarter machines in the workplace will result in many jobs at risk of being taken over in the near future. According to the McKinsey Global Institute,¹³ about 50 percent of work activities are technically automatable by adapting current technologies and 15 percent of workforce will be displaced due to the adoption of automation by 2030.

However, considering the optimistic view of the transformation, current estimates suggest a net positive outlook (the World Economic Forum¹⁴ shows that a decline of 0,98 million jobs will be outweighed by a gain of 1,74 million new jobs by 2022), in particular for specialist roles related to understanding and leveraging the latest emerging technologies. It is thus becoming increasingly clear that AI is not a job killer, but rather a job category killer, especially in specific business areas such as transportation, retail, professional employment services, and customer service.

New technologies will also change the human aptitudes required. Less routine skills will become more important: social skills, creative skills, problem-solving, coordination with others.

The demand for experts in AI, especially in the subfield of machine learning, is growing at a very fast pace and universities worldwide are struggling to keep up with this pressing, raising demand.

¹² PricewaterhouseCoopers. *Macroeconomic impact of artificial intelligence*. 2018 (Accessed on 01.29.2020 at: <https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf>).

¹³ McKinsey Global Institute. *Skill shift automation and the future of the workforce*. 2018 (Accessed on 01.22.2020 at: <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce>).

¹⁴ World Economic Forum. *The future of jobs report*. 2018 (Accessed on 01.22.2020 at: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf).

A number of world-class universities (starting from the world's top-5 schools,¹⁵ namely Harvard University, Stanford University, University of Cambridge, MIT, UC Berkeley) have launched specific master's programs, laboratories, online courses and research groups on AI and data science.

We need to reinforce our educational system in order to adjust and improve the development of competences. The Digital Economy and Society Index¹⁶ designed by the European Commission reveals a worrying situation: 43 percent of the EU population has an insufficient level of digital skills and about 10 percent of the EU labor force has no digital skills. In addition, the number of graduates in science, technology, engineering and mathematics (STEM) is about 20 percent in EU and Information and Communication Technology specialists still account for a low proportion of the workforce (3,7 percent in the EU).¹⁷ In our country, the Italian National Institute of Statistics has estimated¹⁸ that about 80 percent of our companies with more than 10 employees, potentially the most well-structured ones, have a poor or very poor rate of ICT adoption in 2019. Furthermore, the demand for STEM jobs is rapidly increasing: according to the EU agency Cedefop prediction,¹⁹ it is expected to grow by around 8 percent by 2025, much higher than the average 3 percent for all other occupations.

Since low-income jobs are easily automated, a significant transition from the low to the high-skill domain is needed and a redesign of what people learn in schools is required. On the one hand, 65 percent of children attending primary school today will work in new job types that do not yet exist.²⁰ On the other hand, policy-makers must ensure appropriate attention to the needs of elderly people, especially older women, in order to delete the digital divide.

¹⁵ Academic Ranking of World Universities. *ARWU Report*. 2019 (Accessed on 01.31.2020 at: <http://www.shanghairanking.com/ARWU2019.html>).

¹⁶ European Commission. *Human capital. Digital inclusion and skills*. 2019 (Accessed on 01.29.2020 at: <https://ec.europa.eu/digital-single-market/en/desi>).

¹⁷ *Ibid.*

¹⁸ Istat. *Imprese e ICT*. 2019 (Accessed on 01.30.2020 at: <https://www.istat.it/it/archivio/236526>).

¹⁹ Cedefop. *Rising STEMs*. 2014 (Accessed on 01.29.2020 at: <https://www.cedefop.europa.eu/es/publications-and-resources/statistics-and-indicators/statistics-and-graphs/rising-stems>).

²⁰ World Economic Forum. *The future of jobs Report*. 2018 (Accessed on 01.22.2020 at: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf).

The problem is made worse by big investors that more and more often recruit senior researchers in the field of AI from academic institutions, offering salaries and working conditions with which public institutions struggle to compete. Few private companies worldwide, physically located in 1-2 countries (USA and China), act as magnets for talented researchers, creating a worrying unbalancing in the control and knowledge of this transformative discipline. The development of new technologies could generate relevant inequalities across countries, raising the risk of polarization between those who have access to AI and those who do not. For instance, the last AI Index Annual Report²¹ from Stanford University shows that AI is a global phenomenon but highly concentrated: about 85 percent of 2018 AI papers originate outside the U.S., 28 percent in China and 27 percent in Europe. A large, thorough, sustained effort should be done by individual states as well as Europe as a whole to contrast this, and fight against the brain drain of European (and Italian) talents currently happening.

Faced with these issues, the answer can be only a system-wide response. Primary and secondary education must focus on increasing technological understanding, businesses should take an active role in supporting their existing workforces in reskilling and upskilling, individuals should take a proactive approach to their own lifelong learning, and governments should create an enabling environment to invest in people's capabilities and assist all stakeholders in these efforts.

We need a common strategy based on introducing new academic programs and research entities that can contribute, through an interdisciplinary approach, to the capacity-building of local AI talents who are called to design, program and develop AI systems.

Education to AI

A last, but not less important, angle has to do with educating the large public to what AI is, what its uses and misuses might be, and what are the citizen's rights and duties in this respect. AI is humanity's new frontier and its first priority should be to support our fundamental human rights, reinforce social relations and maximize solidarity among peoples. The world must ensure that new technologies are used for the good of our societies and their sustainable development: in a system

²¹ Stanford University's Human-Centered Artificial Intelligence Institute (HAI). *Artificial Intelligence Index Report*. 2019 (Accessed on 01.27.2020 at: https://hai.stanford.edu/sites/g/files/sbiybj10986/f/ai_index_2019_report.pdf).

where a machine makes a decision, we want to make sure that the decision of that system is done in a way that ensures people's confidence in that decision. The more we delegate to machines, the more responsibility we must require to scientists and the more awareness we have to provide to citizens.

Recent political events worldwide clearly show how modern AI algorithms, fed with data carrying rich information about single individuals acquired in an opaque manner, could be used to heavily influence the public sentiment, with multiple implications for democracy and civilization as we know it. Propaganda, disinformation, discrimination, blackmail and other forms of interference provide a great threat to our society.

Moreover, AI brings specifically new challenges in terms of governance that are related to its interaction with human cognitive capacities. Most machine learning and AI systems analyze relevant amounts of data to identify patterns and make automatic decisions. Even if AI is often defined as a 'black box' (because we can see input data and output data but we do not really understand what exactly happens in between), we need to better know the outputs and how to interpret the machine's predictions.

According to the United Nations, we have to deal with three main challenges: information asymmetry (only developers understand how processes and algorithms are constructed and work), the worrying lack of a legal and regulatory framework (e.g., the traditional trade-off dilemma between data ownership, open access to data and data privacy protection), and the mismatch between what is useful for governments and what is favorable for people.

Many institutions and governments are concerned about the ethical implications of AI. Also, the business community suggests a new perspective: about 90 percent of organizations worldwide²² have experienced ethical issues with AI in the last 3 years. We need to immediately start thinking about how we are constructing and managing our digital infrastructure as well as how we want to design and distribute AI systems. A robust and stable legal framework, encouraged by tech giants themselves, will be needed to regulate innovation and address those

²² Capgemini Research Institute. *Conversation with leading CxOs, startups, and academics. Towards Ethical AI.* 2019 (Accessed on 01.24.2020 at: <https://www.capgemini.com/wp-content/uploads/2019/12/Report-%E2%80%93-Conversations-Towards-Ethical-AI-1.pdf>).

issues too complex or fast-changing to be adequately covered by legislation. However, the political and legal process alone will not be enough and an ethical code will be equally important. Many countries – Spain, Denmark, Germany, France, and Italy among others – have published or announced policy approaches that are described as AI strategies, but we will need greater cooperation.

Europe can seize these opportunities if it will exercise its leadership in using AI applications in a way that respects European values and principles: EU-level funding can ensure cross-fertilization of European developments in new technologies and federate investments.

The European Union has recently presented an overall strategy for the EU. Human agency and oversight, robustness and safety, privacy and data governance, transparency, diversity, nondiscrimination and fairness, societal and environmental well-being, accountability: the seven key requirements launched in 2019 by the European Commission²³ for achieving trustworthy AI are addressed to all AI stakeholders designing, developing, deploying, implementing, using or being affected by AI in the EU, including companies, researchers, public services, government agencies, institutions, civil society organizations, individuals, workers, and consumers.

The core principle of the EU guidelines is that the EU must develop a 'human-centric' approach to AI, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union.

Policy-makers should encourage human-centric approach discussion on an ethical AI and they have to work for involving all relevant parties: a non-discriminatory, equitable and transparent use of new technologies is fundamental in our increasingly computerized society.

Together, we can be drivers of good innovation and must find the best solutions to ensure that the development of AI is an opportunity for humanity.

Concluding remarks

The far-reaching impacts that AI exerts on our society at large make clear the need for a holistic approach when it comes to the intersection

²³ European Commission. *Ethics guidelines for trustworthy AI*. 2019 (Accessed on 01.24.2020 at: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>).

between AI and education. In fact, the capacity of AI to make an increasingly large spectrum of human skills obsolete calls for an evolution of the overall education system. Although we have a long road ahead if we are to reimagine the future of education, a number of strategic directions may be defined by drawing on the challenges and opportunities previously discussed.

In a world of accelerating technological and economic change ushered-in by AI, a future of rapidly changing jobs is foreseen. To keep pace with such a fast-evolving landscape, a lifelong learning model has to be established. This will allow to make change the new normal and to put people in the position to build knowledge and skills at all stages of their lives, both for personal fulfillment and to benefit society.²⁴ To supplement the formal education, people will go back to school several times in the course of life, thus preventing their know-how from becoming obsolete. Easy access to skill upgrading meant to match the labor market's needs will be possible by means of flexible adult education systems. They should address the increasing need for digital literacy as part of a systematic individual competence development that may take advantage of the opportunities made available by personalized learning powered by AI.

To turn such a model into reality, it is vital – from pre-school to higher education – to teach students how to learn. This means providing equal opportunities for everyone to acquire basic skills, a qualifying education and a solid foundation for lifelong learning.²⁵ To ensure that no one is left behind, the resulting education model should combine world-class experiences that foster talent with differentiated pathways meant to accommodate weak learners.

In addition, such an education model has to be framed around creativity, one of the few areas on which humans will continue to outperform machines. AI, in fact, is incredibly smart and is poised to evolve exponentially, but it will never match human creativity. Even if traditional educational programs often tended to squash creativity before it could fully develop due to memorization and analytical learn-

²⁴ Singularity University. *The Exponential Guide to the Future of Learning*. 2019 (Accessed on 01.27.2020 at: <https://su.org/resources/exponential-guides/the-exponential-guide-to-the-future-of-learning/>).

²⁵ European Commission. *Lifelong Learning Strategy*. 2018 (Accessed on 01.30.2020 at: https://eacea.ec.europa.eu/national-policies/eurydice/content/lifelong-learning-strategy-22_en).

ing, creativity should become the linchpin of the next-generation learning process as it is the distinctive way we express ourselves as humans. Gazing forward, AI has what it takes to become an amplifier of human creativity. The education system will thus cultivate a new breed of AI-savvy creatives able to harness the machines to get the job done and free-up time and resources for creative thinking.

In a time when society faces ever more complex problems that require systems thinking, it becomes crucial to produce professionals who have the skills to work with people from a diverse set of disciplines. While traditional education had the responsibility to produce individuals who are well on their way to become experts in their field of interest, nowadays integrating concepts from different disciplines to generate new ways of thinking is a pre-requisite for professional relevance. This holds true also for mastering AI and getting the most out of it. The comprehension of AI, in fact, requires T-shaped profiles equipped with robust STEM competencies (e.g., coding, statistics, algorithmic design) as well as complementary non-technical competencies (e.g., ethics, law, human-machine interaction).

Finally, in a labor market made unpredictable by the relentless wave of AI-driven automation, the prominent source of competitive advantage for a worker does no longer reside in mere hard skills but rather in soft skills. As firms seek to deploy their talent pool more effectively, social and emotional skills – such as leadership, critical thinking, initiative-taking, communication, and negotiation – are painstakingly sought-for in the workplace as they distinguish resilient professionals inclined to retraining and redeployment.²⁶ As a result, educators may need to consider redesigning and establishing new metrics to measure skills in a broader sense. They can also look to teach soft skills such as problem-solving or collaboration in a way that is less subject related, for example through making presentations in class, providing detailed critiques on written assignments, and encouraging deeper thinking that explores questions of why and how.

²⁶ McKinsey Global Institute. *Skill shift automation and the future of the workforce*. 2018 (Accessed on 01.30.2020 at: <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce>).

Fourth session

THE ROME CALL: "RENAISSANCE. A HUMAN-CENTRIC
ARTIFICIAL INTELLIGENCE"

The Rome Call: “RenAIssance. A human-centric artificial intelligence”



Introduction

Promoted in the two-year period 2019–2020 by the Pontifical Academy for Life (PAV), the ethical debate on the impact that new technologies have on human life has produced, in addition to the two workshops dedicated to robotics (2019) and artificial intelligence (2020), **The Rome Call**, a document inviting the different actors of the digital world to endorse six essential principles for an ethics of artificial intelligence, with a special focus on legal and educational issues.

The ceremony celebrating the first signatures of the Rome Call took place at the end of the **2020 workshop on Friday 28th February** at the Auditorium della Conciliazione, during an event titled *renAissance*, before PAV academics and over one-thousand sector experts and enthusiasts.

Welcomed by Archbishop **Vincenzo Paglia**, President of PAV, who illustrated the meaning of the document and the commitment of the Academy on this topic, the first signatories of the Rome Call addressed the audience from the stage: **Brad Smith**, President of Microsoft, **John Kelly III**, Executive Vice President of IBM and **Qu Dongyu**, Director-General of FAO. In their speeches, they explained their endorsement of the Call, each emphasizing different aspects of the document. To their reflections was added that of the President of the European Parliament, **David Sassoli**. The signing ceremony was also attended by the Italian Minister for Technological Innovation and Digitization, **Hon. Paola Pisano**, representing the first national government joining the Rome Call, to be followed by many others.

The event should have ended with a private audience with Pope Francis, which was suspended due to a slight indisposition of the Holy Father. Pope Francis nevertheless sent the speech prepared for the occasion, which was delivered by Archbishop Paglia.

In addition to the signing of the document, also the ceremony for the **2020 Doctoral dissertation Award on artificial intelligence ethics**, promoted by the Academy with the contribution of Microsoft, was held. Two winners were awarded the prize on equal terms by the President of Microsoft: the Belgian Tijs Vandemeulebroucke for his dissertation

entitled "The use of Socially Assistive Robots in the care for older adults: A socio-historical ethical analysis" and the American Dominique J Monlezun for his work "Personalist refinement of human rights and social contract bioethics: applications for artificial intelligence".

The text of the Rome Call (also available at www.romecall.international) and the speeches made by all the Speakers are presented in this section of the volume.

The attached DVD offers the full recording of the Rome Call: "RenAIssance. A human-centric artificial intelligence", along with a photo gallery, an interview to Prof. Luciano Floridi (Professor of Philosophy and Ethics of Information, Oxford University, UK) and some extra contents.

Vincenzo Paglia *

Dear friends, we are experiencing an actual “epochal change”, as often recalled by Pope Francis. And we realize it now more than ever. That is the reason why we have gathered here today.

The humankind has known times of deep change that have had a profound impact on the way humanity lives (cf. the wheel, the steam engine, electrical energy). Today, we are confronted with a technological innovation that appears particularly “disruptive” (*disruptive innovation*), both for its continuously increasing speed, and pervasiveness, as it deeply permeates the lives of people and the society alike.

I am referring to “emerging and convergent technologies” that have a deep involvement in the living matter, as they act on the molecular bases of the human body, to the point that they even challenge the notion itself of human life, transforming the way in which we interpret and even modify it. In his letter *Humana communitas*, addressed to the Pontifical Academy for Life, Pope Francis urged us to broaden our horizons, and to understand the sense of the expression “human life” more deeply.

What we call “AI” falls exactly in this framework. We are aware that AI machines can help carry out a wide range of tasks that can improve human life. After all, humankind has the responsibility to make our lives better. However, we know well the risks that we run. It is undoubtedly reductive to think of these technologies only as tools to perform certain functions more quickly and efficiently. In fact, they somehow help us to “enhance” life, but in a way, they change the way in which we live in the world, the way we perceive the reality and ourselves, up to posing radical questions on the identity of humans. We often do not even realize that we interact with automatic systems or disclose information about our personal identity on the Net. In this way, we produce a severe asymmetry between those who extract data (for their own interest – data owners) and those who provide them (unknowingly – i.e., us.) It is well known by now that AI machines

* President of the Pontifical Academy for Life.

can generate actual schemes that control and guide mental and relational habits. It is all the more evident that the "human" dimension is influenced in such a way as to "succumb" to smart machines, much more than the other way around. It is a very serious threat. We cannot stand and watch.

We might say that for the first time in history, man can destroy himself. It happened with nuclear fission in the 1940s. At the time, governments felt the obligation to sign pacts of non-proliferation of nuclear weapons. There emerged, then, another perspective: ecology. The unrestrained exploitation of the planet for profit-making purposes led us to the brink of an ecological disaster. Despite subsequent reconsiderations, governments met in Paris to sign a global agreement on climate. Dear friends, after the atomic bomb, and the environmental crisis, we are now on a third front that directly challenges the human dimension. There is a growing awareness that we can destroy our "common home", however, what has not clearly emerged yet is that we risk destroying its inhabitants, i.e. the specific qualities of human beings and the human family itself.

In this scenario, we must question ourselves on our responsibilities to prevent tragic outcomes. How can we prevent man from being technologized, and promote the humanization of technology, instead? How can we avoid being in thrall to "algo-cracy", i.e. the power of algorithms? Isn't it necessary to develop a vision of the society and of the future of the planet that sees man as its unconstrained protagonist? I think that it is first necessary to assign a dogmatic and authoritarian role to both political management and technocratic liberalism. This is only possible if ethics is given a role not only when the product is "thoroughly complete", when there is nothing else to do than (try to) rule its use, but also along its research itinerary. In other terms, it is not sufficient to limit our attention to controlling individual devices, leaving end-users with the task of using them practically, according to abstract and general reflections on the respect of individual rights and dignity. In fact, experience teaches us that this ethical response is useless when everything is already decided. Certainly, the principles enshrined in the Social Doctrine of the Church – such as dignity, justice, subsidiarity, solidarity – are inalienable. However, the complexity of the contemporary technological world demands a dialogue among these sectors for them to be actually incisive. There is a need for an ethics that reflects on the criteria underlying the design itself of algorithms and the responsibilities of those who work in the individual stages of their production.

In my view, ethics is called to accompany the whole development cycle of technological devices, up to choosing projects to invest resources on. This is possible in case of an inter-disciplinary model of ethics in which various skills collaborate on all the development phases of technological devices (research, design, production, distribution, individual and collective use). The goal is to grant an expert and shared control over the processes leading to integrated relationships between human beings and machines in the new era opened by AI.

Let us be careful, though. This task cannot be performed by any component of society by itself. Dialogue and collaboration among all the stakeholders involved are fundamental, if our common interest is to be the safeguard of the "common home" and the peaceful coexistence of the whole "human family". Currently, in no sector it is possible to overlook a global vision, even more so in the technological field.

That is the sense of today's meeting. As we face today's enormous challenges, we are urged to have the moral impulse to draw a future of peace for the planet. We may learn from history and make an analogy with what happened in the aftermath of World War Two. After the tragic experience of totalitarian regimes that had crushed the dignity of people and their expressions – on the strength of a moral impulse – men drafted the universal declaration of human rights. Dear friends, we should not wait for similar disruptive events to materialize. We can and must anticipate them and guard against them before they happen. As a human family, we must mobilise beforehand, to reaffirm and promote the right of all people to inhabit the Earth with their diversity and in peace.

In light of these reflections, the Pontifical Academy for Life started a collaborative endeavour with a number of interlocutors: academics, members of civil society and companies producing these technologies. The Call that will be signed today is the first step. Faced with the changes produced by technological innovation, the Academy feels called to respond, according to its specific nature. We intend to promote reflections and initiatives that contribute to support a technological innovation that can be a factor of authentic human development for the promotion of the common good. We intend to pursue the perspective of a humanism for the digital era. Besides the portion of the international academic world that already joined the Pontifical Academy for Life in this process, we collaborate with two companies, IBM and Microsoft, who agree on some fundamental perspectives with the Academy. In addition, the involvement of outstanding public institutions, such as the

European Parliament, represented by its President, and an international body, such as the FAO, represented by its Secretary General, gives us hope for the future. In fact, the Call's spirit is not restricted to the partners that are here today, while excluding others. Quite the opposite, our intention is to give rise to a movement that extends and involves other stakeholders: public institutions, NGOs, industries and groups to produce guidelines for the development and use of AI technologies. From this perspective, we may say that the first signature of this Call is not an accomplishment, but rather the beginning of an engagement that appears all the more urgent and important than anything that we have done so far. Each of us here is called to feel involved and offer their contribution.

Brad Smith *

Good morning. It's my pleasure to be here. It's my honor to represent Microsoft and to be here with these other such distinguished speakers.

I want to talk a little bit this morning about the Rome Call for AI Ethics. I wanted to say a little bit about why it matters, about why it matters, in fact, so much.

I think it matters in part because, I must state the obvious, Artificial Intelligence is going to change the world. And before we go too far forward with it, we need to think hard about the kind of impact we want to have.

I think there's a few places that are better to have this conversation than here in Rome. And I think it's especially fitting that this has been framed the way it is, to talk about it in terms of the Renaissance.

Because, after all, the Renaissance changed the world. Indeed, for many of us around the world, we might think of the Renaissance first in terms of art but, as you can see, the Renaissance in so many ways changed almost everything. It led over the course of two centuries to inventions that we still take for granted today.

It changed the way we work with each other.

It changed the economy and the way the world works. It did more than that. It changed the way we were given the opportunity to think about the world.

When we think about Gutenberg and his printing press and we think about the fact that, of course, the first thing it was used to do was to print a Bible.

A Bible that gave new power to humanity to read about religion. And, of course, it wasn't just Gutenberg, it was a series of inventors including Manutius who took Gutenberg's invention. And, like so many aspects of technology, he made it better. With new typefaces he started to make things like a very heavy Gutenberg Bible smaller and more portable. In many ways, they invented the world that we live in today.

* President of Microsoft.

We continue to change that world as we democratize information and we put it in the hands of more people around the world. Of course, that was the Renaissance. The Renaissance was followed, starting in the middle of the 1700, with in many ways the era in which we still live today: waves of Industrial Revolution.

When you look back at the history of technology what you fundamentally find is that every era has many inventions, but so many of them are grounded in one invention that enables so much else.

First, it was the steam engine. And then it was electricity. And then it was the combustion engine and then the microprocessor.

I think, in many ways, we should think about Artificial Intelligence as not just the newest step but another one of this defining fundamental enabling technologies.

In many ways, I think it's right to think about the Artificial Intelligence as the combustion engine of the 21st century.

The combustion engine more than any single invention remade the world a century ago. Between 1910 and 1940 it, of course, made possible for people to drive and to have a car. But, on farms, that led to the invention of the tractor. For the movement of goods, it made it possible to have a truck.

It brought people literally closer together around the world because it made it possible to invent an airplane. It changed the face of war because it made it possible to invent a tank.

It touched every aspect of humanity. And I think it in a similar way the era we are now entering, an era that you can think of in the year 2020, we can recognize that over the next three decades between now and the middle of this century, Artificial Intelligence is likely to have a similarly and sweeping impact.

In so many ways, it not only will change the world, it will impact the world and that just in ways that are good, although the good will be powerful. In so many respects, Artificial Intelligence will enable a new generation of opportunity.

It probably will become the most powerful tool in the world. It will become a tool that will enable us to do things that we only dream of doing today.

It should enable us over the course of the next three decades to find the cure for cancer.

We desperately need Artificial Intelligence to do what we believe it can do. And that is help us ensure the sustainability of the Planet. In

every field of human knowledge much like we learnt from the Renaissance itself, Artificial Intelligence will fuel the quest for knowledge.

But what we also know from the history of technology, as Abp. Paglia mentioned, is that every tool poses new challenges as well.

Artificial Intelligence would generate a new generation of challenges.

It probably will become the Planet's most formidable weapon. We're already seeing this around the world.

We're seeing AI used to fuel more powerful cyberattacks, whether it is by criminals or by certain nation states. We're seeing Artificial Intelligence put to work in various parts of the world to fuel mass surveillance.

Mass surveillance that poses new questions and challenges for the rights of people.

We're seeing Artificial Intelligence unleash a new generation in terms of the automation of more jobs.

We will see new jobs created, but we will see many existing jobs changed or even displaced. So, like every tool it is and will be a weapon. And because Artificial Intelligence will be the world's most powerful tool almost by definition, it will also become the Planet's most formidable weapon.

That's why I believe it's so important that we also think about this as a clarion call about the need for a new generation for ethics.

As Abp. Paglia mentioned and as Pope Francis has stated, this is an important moment for the history and the future of humanity.

That's why today is so important.

That's why I believe the Rome Call for AI Ethics is so important.

Why does this matter?

It matters for many reasons. But I think the first thing to reflect upon is that the Rome Call is about creating a broad community.

I come from a company that invents technology, a company that works with a company like IBM, that represents a broad community and industry that is creating technology. And yet I would be the first to say that this technology is too important to be left solely to the people who will create it. Because it will affect all of us, we need to work together in a community that represents all of us.

I think it's of fundamental importance that the Catholic Church is a voice for humanity in thinking about the future of technology.

The world needs the Catholic Church to be a voice for humanity.

The world needs the Catholic Church to be a voice that will welcome as it is everyone else to this broad table and this big conversation.

This is a conversation that will need to bring together the world's great religions. Organizations like the United Nations, organizations like the European Parliament, the great NGOs, and representatives of civil society, as well as those of us across the business community and the technology sector.

That is the first reason that I believe the Rome Call for AI Ethics matters so much. But it also matters because of what it says. I think that the Rome Call starts with something of fundamental importance.

It starts with people. And, as it says, what really defines us as people is that we are endowed with reason and conscience.

The future of humanity requires that both of these come together. And, as the Rome Call recognizes, what we need to do is exercise our reason and conscience to safeguard the rights and the freedoms of individuals.

The future of humanity needs to be protected one person at a time. With a recognition that each of us is an individual and has something unique and special. I think the Rome Call is also important because it connects two fundamentally important concepts of our time.

We need both to serve and protect human beings and the environment in which they live.

The planet on which we live.

This is not a phrase that we would have seen put together in this way if we were meeting a decade ago. But as we think about the decade ahead, it has become fundamentally clear that the protection of humanity requires not only the protection of human beings but the planet as well.

Ultimately, as the Rome Call makes clear, we need to think about digital rights in an AI era. The rights of people in an era of Artificial Intelligence and, as the Rome Call states, there are six fundamental rights that are important to the future of humanity and this planet.

We need to ensure that Artificial Intelligence is designed and deployed in ways that are inclusive of all people. That are impartial and fair, that are reliable and safe, that are guided by the need for people to be secure to have their privacy protected. And, perhaps, most broadly, to be transparent so that we as people can understand how machines are making decisions and to ensure that machines remain responsible, accountable if you will, to humanity.

These are the new rights of our time and the Rome Call underscores the importance of our working together to advance them.

Another aspect of the Rome Call that I think is of great importance is that it recognizes that the protection of people and the advancement of these rights needs to start with education.

It will require a new approach to education to reach a new generation of people. In some ways, it's most obvious to think about this as reaching the next generation, the youngest generation of people, but the truth is we are entering a new era in which we will all need to learn new skills. That's what the Rome Call makes clear. This isn't just about skills relating to new technology. It isn't just about data science. It isn't just about computer science.

It is fundamentally about the liberal arts. It is about ethics. It is about humanity as well.

Ultimately, it calls for a new generation of learning. And that too requires a broad conversation.

When I step back and I think about all of this, I'm constantly reminded of an important speech that was given almost six decades ago.

It was a speech given in the midst of a new focus, on a new generation of technology and what it would mean for people. It was a speech given in 1962 at a university, at Rice University in Houston, Texas.

It was a speech given by the President of the United States John F. Kennedy. The speech is remembered to this day because it was a speech in which he called on the people of the United States to reach the moon before the end of the decade.

That is what people remember the speech for. But there is another line in the speech that I think speaks even more powerfully to us today. Because during the speech President Kennedy uttered the words, "Technology has no conscience." When you think about it, technology never has had a conscience. Think about all of those inventions from the Renaissance.

Think about the Gutenberg Bible. Think about the combustion engine. Think about what it did both to bring people together and to build the tank.

Technology has no conscience. Its use will never be defined by the products themselves. Technology has no conscience, but people do. And we do. That is what matters. That is what the Rome Call for AI Ethics requires that we keep in mind.

In the history of humanity, I might argue that in some ways we are unique. Why?

We are the first generation of people in the history of people to create machines that can make decisions that previously could only be

made by people. The future of humanity rests on our ability to get this right.

If we come together and make decisions wisely and effectively, then future generations will benefit and if we fall short or fail, every generation that follows will pay the price. That is why this moment is so important.

That is why the Rome Call for AI Ethics is vital. That is why the voice of the Catholic Church and all of the voices of humanity need to be heard. Technology has no conscience, but people do, and we must.

We must call not just on our ability to reason but on our conscience. That is what will help shape the future.

Thank you very much.

John E. Kelly III *

Good morning. It's a pleasure to be here this morning to represent both the IBM Corporation, which is a corporation and a technology company that has survived and prospered for over 100 years and has produced much of the technology that we know as computing and introduced it in a responsible way.

I'd also like to congratulate the Academy and Monsignor for a wonderful piece of work, and I'm privileged to join all my colleagues from industry and government as well in this important effort.

I'll also say on a personal note, having spent nearly half a century of my life and career in technology and also spent my entire life as a Catholic, these two worlds for me have grown in parallel; and at this moment, this historic moment, for me personally it's bringing these two worlds together at a very, very critical point in time. And once again, I thank the Academy for this opportunity.

I believe to understand where we are at this point in time and where we're going and the responsibility, we need to understand a little bit about how we got here. So, let me just quickly, quickly review 100 years of technology. Let's start 75 years ago. This is a machine that IBM built with Harvard University called the Mark 1.

The Mark 1 was the first computer at scale, and when I say "at scale," it was 80 feet long and weighed five tons. It was used for the most advanced science and research of the time. It had amazing performance: it could add two numbers in less than one second. It could multiply two numbers in about six seconds; and, it could divide two numbers in about 12 seconds. At the time it was an astonishing tool, 75 years ago.

* Executive Vice President, IBM.

Fifty years ago, as Brad mentioned, we put a human being safely on the moon and returned that human to earth. I had the privilege not only of growing up and seeing this live in 1969, but last year I had the privilege of meeting the man in the upper right-hand corner, Gene Kranz. Gene Kranz was the mission control director. He was the man in Houston that determined the flight pattern, determined when the astronauts would land and how to get them safely back.

And when I talked to him about his interaction with the IBM computers that put the man on the moon and got him back, he said to me, he said, John, neither humans nor the machines could have put a man on the moon and got them back safely. It took humans – astronauts, my control center people and your computers – to make that happen. Man and machine did something astonishing, and we changed our view of humanity when we looked back from the moon at the earth and saw our place in space.

Fast forward 25 years ago. We worked with the Vatican Library to digitize and make available all of the knowledge in the extensive Vatican Library, once again democratizing the information and causing us as humans to think about our place from a religious standpoint and our place in society. Again, technology and human history coming together and democratizing knowledge in a very unique way.

Fast forward to just nine years ago. Ten years ago, we would not be talking about artificial intelligence. It was the last thing on people's minds. But on that day in February, the Watson artificial intelligence machine, as the Father mentioned, beat two human beings in natural language and open domain questioning at rapid speed. And the world was stunned.

And from that moment on, artificial intelligence exploded around the world. Every company pursued it, every human had it top of mind. It was a very special moment in time, and I'll have more words to say about that in a moment.

Fast forward to just two years ago. We built the two largest supercomputers on the planet, and they remain the largest today. These computers...now, remember that Mark 1 system could do one calculation per second. These machines can do 200 billion calculations a million times a second, and they're built to be the largest, most intelligent sys-

tems in the world. In just a few years we have come to a point where the capability of these systems is simply unprecedented.

Now, let's take a different look at it, and let me carve it into three timeframes because I think we've seen three fundamental eras of computing and something has changed fundamentally. The first era, the tabulating era, we used simple things like punch cards, and every punch in a card that we as humans put in was fed in the computer and it represented a piece of data or an instruction. And the machine had no knowledge, nothing. It did exactly what we punched in that card, and it basically did arithmetic.

We think we're sophisticated today, but really, we've been in this generation of programmable computers and instead of punching cards, we write software, and every instruction that that computer executes is defined in that software. In the tabulating era, in the programmable era, a computer could do nothing that we did not explicitly tell it to do, and it had no awareness of what it was doing.

We now have entered an entirely new era of computing, and that's why we're here today. We've entered the era of artificial intelligence where machines no longer require explicit instructions to perform every procedure. They learn, they think, and they execute largely based on what they have learned.

This, ladies and gentlemen, is a fundamentally different era of computing; and as was discussed, has fundamentally different implications on the world.

Let me say a little more about this era of artificial intelligence. And I could go back into the fifties when scientists were experimenting with artificial intelligence, but I would argue that this was truly the advent of artificial intelligence. In 1997, in our labs in IBM, we built a computer called Deep Blue...you'll notice everything in IBM is called blue. And while we had a human being moving the chess pieces, every move came from this computer, Deep Blue.

The computer went on to beat Garry Kasparov, the grandmaster of chess. Garry Kasparov had never lost a computer game to a human or a machine, ever in his lifetime. When we won that match, I remember speaking to Garry right after. And needless to say, he was a little bit

upset, but he said to me, he said John, I've played many games in parallel, but that game was strange. He said, it was like I was playing multiple people in the same game. And I said to him, Garry, that's because I trained the machine with multiple grand masters. So, he was not able to psychologically figure out who he was playing in that game.

I'll also tell you that I met Garry last summer, and Garry said to, he said, John, we did it wrong. It shouldn't have been me versus the machine; it should have been me and the machine playing either a human or a machine. He said, you give me that machine today and I can beat any machine or any human. I thought that was really insightful.

Fast forward to this famous Jeopardy! match. Once again, the computer system dominated a man versus machine in a open domain question and answering; and as I said before, it woke the world. And if you read the fine print from Ken Jennings, he says, I welcome my computer overlords – sort of a joke, not too funny.

That, ladies and gentlemen, was the last time IBM and I did a man versus machine. We realized that that was not the right paradigm. It's not man versus machine; it's man and machine doing things that neither one can do on their own.

We thought deeply about this, and I would maintain that these machines are not simply something that we build and it learns and it acts on its own. It's not something that really has a consciousness. But these machines – these artificial intelligence machines – are simply a reflection of us as humans.

When I build an artificial intelligence machine, it's as dumb as that podium: it only learns based on the data and the training we give it. So, what would we expect of a machine that takes the digital exhaust from society, is taught by humans? Wouldn't we expected to develop the biases and thinking of us as humans? It has no ability to generate its own hypotheses. So, when I think of artificial intelligence machines, to me, it's like looking in the mirror and seeing ourselves through the eyes of this machine.

I would also argue, state that this is all about choices, and choices are based on our ethics. And a great example of this was immediately

following that demonstration of the Watson AI machine, we had a choice to make in IBM. We could either put this machine strictly to work in financial services sectors and other places to make money with this machine...I mean, think about it. This machine was faster and had abilities far beyond what most humans could do in narrow areas.

But we said, no, we're not going to go that way. We determined the first use of this artificial intelligence machine was in the area of healthcare, and so we immediately went to work training the machine at places like Memorial Sloan-Kettering in the area of cancer. And as of today, over 150,000 cancer patients have benefitted from the training and exposure to the Watson AI machine. And by the way, 80 percent of them are not in the United States or Europe, they're in places like China and India and South America where the need is the highest.

And we didn't stop there. Yesterday I was at Bambino Gesù Hospital, pediatric hospital here in Rome. And for those of you not familiar with Bambino Gesù, it is one of the largest children's hospitals in the world, owned by the Vatican, and it treats children from all over the world with severe cancer, rare diseases.

We formed a partnership yesterday with Bambino Gesù to take Watson, our AI machine, and apply it to two use cases. One is childhood brain cancer, which is one of the most horrific, horrific diseases facing mankind; and, the second is rare diseases. So, think about rare diseases. There are about 8,000 known rare genetic diseases. They affect about eight percent of the human population globally.

But any given physician or doctor any place in the world will rarely see any of those 8,000, or they'll occasionally see one, and frankly, not know what to do. So, IBM in partnership with Bambino Gesù is going to collect the data from the hospital and others and produce one of the largest data sets and apply artificial intelligence and then democratize the knowledge of rare diseases in children. And we look forward to a tremendous partnership with Bambino Gesù.

So, back on this theory that it's man and machine. As I said, we're not going to play games anymore. This is no longer man versus machine. And around the world, ourselves and our colleagues in the tech industry are hard at work, whether it's in agriculture, healthcare, energy exploration, bringing the best that humans bring with the best

of these machines to do things and reach decisions in timeframes that matter, and always – always – make better decisions for outcomes for humanity. And frankly, I haven't seen a single industry or a single institution that can't benefit from this technology.

I also agree that one of the secrets here is education. Artificial intelligence is a very advanced technology, but we cannot allow it to only go to the have's and create another set of have nots. We have to bring artificial intelligence globally, and we have to bring it to those most in need.

We've been very privileged to bring on board students to learn artificial intelligence globally with a particular focus on Africa where the opportunities are immense. We have brought our artificial intelligence education systems, software systems to Africa to teach the students on the ground so that we can be inclusive in advances in artificial intelligence.

Now, I'd like to conclude by, again, thanking the Academy. These highlights of the Call for Ethics I think are spot on and the IBM Company is committed not only for ourselves to follow these great guidelines but to also be a spokesperson, spokescompany around the world and globally and advocate for these among our tech partners and among the governments of the world. If we can be transparent, inclusive, secure systems, et cetera we can do quite well.

I'd like to end with this quote from Pope Paul VI. The human heart absolutely must become freer, better and more religious as machines, weapons and the instruments people have at their disposition become more powerful. I think that says it all. In a sense, it's another way of saying that this isn't about what the machines will do and not do; this is about what we as humans choose to do with the things that we create. Thank you very much.

David Sassoli *

Good morning to everyone.

First, let me extend my warmest greetings and gratitude to His Excellency, Abp. Vincenzo Paglia, for his kind invitation, and to the members of the Pontifical Academy for Life, the outstanding speakers and all of you who have promoted and driven this important meeting.

We live in a time of major changes and great challenges. The onset of new communication media, the digital revolution and, in general, what we call artificial intelligence, offer us extraordinary opportunities, but at the same time, they are radically modifying the way we act and the way we are.

Up to the present day, research and innovation have enabled our societies to progress and achieve great goals, while fostering citizens' well-being. Over the past few years, technological progress has accelerated rapidly, so much so that it has undergone an actual revolution that has had and is still having a very strong impact both economically and socially.

Robotics and artificial intelligence offer new possibilities to address the challenges that our society will face over the next decades: optimised energy networks, sustainable production, precision agriculture, governance of the economy and finance, and a moderate use of resources.

However, benefits and challenges can go hand in hand and know no boundaries.

We use artificial intelligence in transportation, where a simple app on our smartphones can predict traffic conditions or even operate driverless vehicles; we use it to trace our water and energy consumption or, in the healthcare field, to analyse huge quantities of data that help us treat some pathologies and identify the best practices to prevent them.

As anyone may understand, artificial intelligence can enable the concentration of enormous power to the detriment of the most vulnerable.

* President of the European Parliament.

Moreover, algorithms consider human beings as simple data in rationalization, efficiency-building and revenue-generating processes.

The progress of robotics and artificial intelligence, together with globalized digital communication and the potential of networks, pose probing questions that demand a deep reflection and especially the capacity to read changes with farsightedness and a great sense of responsibility.

Ladies and Gentlemen,

What we may define the “fourth industrial revolution” – after the steam, electricity, and automation revolutions – implies that technology itself has predictive capacities of human activities, thanks to data and the algorithms that follow. This revolution has overturned our development models, and needs to be accompanied by guidance to expand and not diminish our rights of social, political, economic and technological citizenship.

A new world imposes new rhythms.

Moreover, the new era arises from technical advances that are unifying and enable science to touch deeply the values that we often take for granted.

Digitalization and automation are profoundly changing the way we live and relate to society. A series of questions emerges and the related answers impose great responsibility.

How can we prevent the digital revolution from bringing more inequalities in terms of rights?

How can we enable the world of education to be up to address these challenges? How can we grant young people fair access to productive processes?

How can we compensate the job losses due to the onset of new technologies with new job opportunities?

A few years ago, writer Isaac Asimov said that a “robot may not injure a human being or, through inaction, allow a human being to come to harm”.

Today more than ever, we need to formulate policies that help us reap the fruits of technological progress and grant at the same time the compliance with those social norms that represent for us indispensable achievements.

In our contemporary society, artificial intelligence that comes from the combination and breakdown of a set of algorithms, is now an integral part of our daily lives; however, it is appropriate to recall that intelligence cannot be separated from human thinking and conscience, albeit the concept of relativity is intrinsic in the scientific experience.

The attempt to understand the effects of major transformations can help prevent an overwhelming, uncontrolled individualism that is inclined to cause a great sense of solitude, as well as new forms of marginalization.

Faced with the unknown variables that this time of great transformation presents, profound transparency is essential. To this end, the active involvement of public opinion, national parliaments, our universities, and the labour and business world is fundamental to encourage an accurate reflection on public regulation. It is a decisive challenge for the European Union to formulate and adopt common standards on the new technological boundaries and measure their impact on the respect of fundamental rights.

In this historical moment, the European regulatory work – on themes ranging from data to privacy – is worth supporting, in that it produces a positive response in the processes of globalization. After all, what happens here causes an immediate reaction outside the European space.

If the European Union has always supported policies in favour of research and innovation, the European Parliament then has the duty to protect its citizens even more from the impact that the new technologies may have.

During the past legislature, the Parliament asked the European Commission to update and integrate the Union's juridical framework with clear ethical principles that take the human factor in due consideration and do not underestimate it, in fact, our citizens must have the possibility to control their data, protect their privacy and discern the information that they receive.

The European Commission has welcomed this request and recently presented a White Paper on Artificial Intelligence – including its social implications – that triggered a European-wide debate. Together with the White Paper, the Commission also published a strategy promoting the access to non-personal data for big, small and medium-sized enterprises, granting the respect of the private sphere.

It will be up to the European Parliament to analyse these texts with great attention over the next few months. We shall provide our contri-

bution to reflect on how Europe could become a world leader in this transformation, while remaining a global model in the protection of the rights and dignity of people.

The digital revolution is deeply changing our lifestyles, the way we produce and consume. We need rules combining technological progress, business development and the protection of workers, people, and democracy. In a scenario in which uncertainty still seems to prevail, it is necessary to sustain career reorientation policies, investing increasingly in life-long learning.

In this sense, Europe can be instrumental in a world that has no rules, but needs to find new ones.

This is our challenge: the extent to which we allow these technologies to develop and influence our lives.

Without any regulation, artificial intelligence can be a risk that may jeopardise not only the protection of personal data, but also expand the digital gap in terms of access and knowledge.

A reliable artificial intelligence system should not undermine fundamental rights and for this, it is necessary to create preventive impact assessments, and promote a human-centred approach, because humans are the only ones who can knowingly guide the actions and decisions made by an artificial system.

We need more data scientists, more engineers and more philosophers to understand the long-term effects of these systems. It is necessary to strive to integrate research and innovation with the humanist tradition that is at the basis of the rights enshrined in the EU Charter of Fundamental Rights.

Moreover, it is our duty to ensure that the technology developed is safe, that responsibilities are clear and that humans can always control decisions.

As prof. Benanti said: "If we want machines to support man and the common good, without replacing the human being, then algorithms must include ethical, not only numerical values."

For these reasons, we need to network, forge alliances and generate new methodologies. We must strive to seize the opportunities offered by the Fourth Industrial Revolution in the best possible way, knowing that if we do not guide it, then algorithms shall guide us.

To do that, we must strive for artificial intelligence to develop within an appropriate legal framework that can – as Mons. Vincenzo Paglia

stated – “accompany the whole technological development cycle: from the choice of research lines to the design, production, distribution and end-users.”

The opportunities science provides can lead to a process of unification of life that is matched by knowledge unification. Humans tend to converge on scientific data. Leonardo da Vinci believed that the scientific experience has the capacity to settle any dispute among men... we want disputes to be resolved also in view of a shared defence of the fundamental rights and freedom of the person.

Everybody's contribution is essential to this effort, and certainly Christian personalism can still provide a decisive contribution.

Thank you.

Dongyu Qu *

Excellencies, Ladies and Gentlemen:

I would like to thank the Pontifical Academy for Life for convening this meeting on the highly relevant topic of artificial intelligence.

I am very pleased to contribute to the interesting exchange by adding the perspective of food production, food security, and food systems as a whole into this discussion on AI.

Since I was also a scientist for more than 30 years – in BT and not IT – I consider AI to be a cross-cutting area between IT and BT, because when you talk of the AI, actually, it is just the new hybrid between IT and BT.

My speech will touch upon four points: How should AI be? In which areas of agriculture and food can we apply it? What is required for the digital transformation of agriculture? How can we implement this?

Machine learning made its way out of research labs and into our daily lives. The previous speeches already mentioned the very beautiful story on how we enjoy life based on innovations that promise to help us tackle humanity's greatest challenges.

The significant impact upon the functioning of our societies and economies is evident.

Artificial Intelligence needs to be transparent, inclusive, socially beneficial and accountable. That's why two speakers, Mr. Brad Smith and Mr. John Kelly already mentioned the principles that we should focus on and follow.

And we need to ensure the human-centric approach in designing and implementing artificial intelligence today and in the future.

From a Food System transformation perspective, we look at digitalization, big data and AI as sources of hope – as part of a solution.

* Director-General of FAO.

AI is a practical tool for advancing scientific and professional solutions to help farmers, foresters, fisher folk across the globe.

Because Artificial Intelligence is about recognizing patterns, making sense of messiness and presenting powerful analysis. Agriculture in rural areas is always a complicated system so we need big data. We need computing to model, to analyze and to have a practical solution which is a comprehensive solution.

Analyzing big data and using new technologies in our work is a real break-through: satellite imaging, remote sensors, mobile and blockchain applications, you name it.

FAO already uses many of these tools in projects to optimize food chains, manage water resources, fight against pests and diseases, monitor forests, identify species, increase preparedness of farmers when disasters strike and in many other activities.

A digital agriculture means a more efficient, sustainable agriculture. One that is much better able to improve rural livelihoods.

Turning agriculture and rural areas digital means revolutionary changes in biological and environmental factors, in agricultural processes such as production, operations, processing, management, marketing, and in rural governance.

This ambitious vision requires adequate Big Data Centers and Cloud Platforms.

Databases containing information on arable land, on important agricultural germplasm, on shared rural assets and on farmers and agricultural businesses need to be set up.

The digital transformation of production operations will also need to be accelerated. This includes remote sensing of all aspects of planting, building digital livestock facilities, establishing smart aquaculture systems and digitizing the seed industry.

The digital transformation of related management services is also needed: To build a digital service program for rural agriculture, including smart environmental monitoring and digital governance.

The necessary infrastructure that all of the above is based on will need to be established. Broadband connections, roads, bridges, and with all this data service provider.

All of the above also needs human capacity: Talent support, rural vocational training, and enhancing digital skills. Appropriate perfor-

mance reviews, incentive systems will open new possibilities for youth and rural population.

And how do we coordinate all this at a global level? As Director-General of the Food and Agriculture Organization of the UN, we will look at this point from a multi-lateral perspective.

The UN system has an important global role to play in balancing technological progress with social progress.

In that respect, we need responsible innovation and a better understanding of the implications and the potential benefits of AI on the world.

This includes looking at gaps between the developed and developing countries.

And the digital gap is already a reality that needs to be addressed: 6 billion people are without broadband today, 4 billion without internet, 2 billion without mobile phones and even 400 million people are without any digital signal – so they are already completely ignored by the digital world.

Additionally, there are significant gaps between men and women, young and old and clearly rich and poor.

But there is also a gap in promoting dialogue, creating synergies and enhancing awareness for issues specific to digital agriculture.

So how can we close this 'international gap' between the key actors? And how can we bring about a fundamental shift in agriculture towards digitalization?

At the Global Forum for Agriculture last year, the international community tasked the FAO with designing and presenting a Digital Platform to address the issue of a digital agriculture.

Earlier this year in Berlin, 76 Ministers endorsed FAO's proposed International Platform for Digital Food and Agriculture.

And today I come here and I am really thanking the support from private sector like Microsoft and IBM and others – not only focusing on the market issues but also focusing on food and agriculture issues and the environment of course.

The Digital Platform will strive to engage all actors, players and stakeholders within the agri-food system, and will activate cross-sectorial and cross-competence experts to consolidate, enhance and diffuse the state of digitalization in the sector with a strategic approach.

The Platform will help governments to identify the potential of digitalization, to enable stakeholders to access and benefit from digital technologies and it will facilitate dialogue, raise awareness and build trust in digital technologies.

The task ahead of us is a big one.

We are convinced that transforming our food systems to feed the world will be achieved with a digital agriculture and SDG – Sustainable Development Goals – of 2030 Agenda is our common consensus endorsed by the General Assembly of the UN.

As President John F. Kennedy said, also heard from Mr. Smith, technology has no conscience and people who master technology have a conscience, have a passion. Like Holy Father here, they always have a deep passion and love and heart for people.

And food is a basic human right, regardless of the fact that you are poor or rich, and that is a fundamental demand to survive.

FAO is willing to play its part as a facilitator, as a knowledge Organization, to help you use technology in the right way to help people in developing regions for their livelihood improvement.

Thank you, thank you very much.

© Copyright - Food and Agriculture Organization of the United Nations

Rome Call for AI Ethics

INTRODUCTION

"Artificial intelligence" (AI) is bringing about profound changes in the lives of human beings, and it will continue to do so. AI offers enormous potential when it comes to improving social coexistence and personal well-being, augmenting human capabilities and enabling or facilitating many tasks that can be carried out more efficiently and effectively. However, these results are by no means guaranteed. The transformations currently underway are not just quantitative. Above all, they are qualitative, because they affect the way these tasks are carried out and the way in which we perceive reality and human nature itself, so much so that they can influence our mental and interpersonal habits. New technology must be researched and produced in accordance with criteria that ensure it truly serves the entire "human family" (Preamble, Univ. Dec. Human Rights), respecting the inherent dignity of each of its members and all natural environments, and taking into account the needs of those who are most vulnerable. The aim is not only to ensure that no one is excluded, but also to expand those areas of freedom that could be threatened by algorithmic conditioning.

Given the innovative and complex nature of the questions posed by digital transformation, it is essential for all the stakeholders involved to work together and for all the needs affected by AI to be represented. This Call is a step forward with a view to growing with a common understanding and searching for a language and solutions we can share. Based on this, we can acknowledge and accept responsibilities that take into account the entire process of technological innovation, from design through to distribution and use, encouraging real commitment in a range of practical scenarios. In the long term, the values and principles that we are able to instill in AI will help to establish a framework that regulates and acts as a point of reference for digital ethics, guiding our actions and promoting the use of technology to benefit humanity and the environment.

Now more than ever, we must guarantee an outlook in which AI is developed with a focus not on technology, but rather for the good of humanity and of the environment, of our common and shared home and of its human inhabitants, who are inextricably connected. In other words, a vision in which

human beings and nature are at the heart of how digital innovation is developed, supported rather than gradually replaced by technologies that behave like rational actors but are in no way human. It is time to begin preparing for more technological future in which machines will have a more important role in the lives of human beings, but also a future in which it is clear that technological progress affirms the brilliance of the human race and remains dependent on its ethical integrity.

ETHICS

All human beings are born free and equal in dignity and rights. They are endowed with reason and conscience and should act towards one another in a spirit of fellowship (cf. Art. 1, Univ. Dec. Human Rights). This fundamental condition of freedom and dignity must also be protected and guaranteed when producing and using AI systems. This must be done by safeguarding the rights and the freedom of individuals so that they are not discriminated against by algorithms due to their “race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status” (Art. 2, Univ. Dec. Human Rights).

AI systems must be conceived, designed and implemented to serve and protect human beings and the environment in which they live. This fundamental outlook must translate into a commitment to create living conditions (both social and personal) that allow both groups and individual members to strive to fully express themselves where possible.

In order for technological advancement to align with true progress for the human race and respect for the planet, it must meet three requirements. It must include every human being, discriminating against no one; it must have the good of humankind and the good of every human being at its heart; finally, it must be mindful of the complex reality of our ecosystem and be characterised by the way in which it cares for and protects the planet (our “common and shared home”) with a highly sustainable approach, which also includes the use of artificial intelligence in ensuring sustainable food systems in the future. Furthermore, each person must be aware when he or she is interacting with a machine.

AI-based technology must never be used to exploit people in any way, especially those who are most vulnerable. Instead, it must be used to help people develop their abilities (empowerment/enablement) and to support the planet.

EDUCATION

Transforming the world through the innovation of AI means undertaking to build a future for and with younger generations. This undertaking must be reflected in a commitment to education, developing specific curricula that span different disciplines in the humanities, science and technology, and taking responsibility for educating younger generations. This commitment means working to improve the quality of education that young people receive; this must be delivered via methods that are accessible to all, that do not discriminate and that can offer equality of opportunity and treatment. Universal access to education must be achieved through principles of solidarity and fairness.

Access to lifelong learning must be guaranteed also for the elderly, who must be offered the opportunity to access offline services during the digital and technological transition. Moreover, these technologies can prove enormously useful in helping people with disabilities to learn and become more independent: inclusive education therefore also means using AI to support and integrate each and every person, offering help and opportunities for social participation (e.g. remote working for those with limited mobility, technological support for those with cognitive disabilities, etc.).

The impact of the transformations brought about by AI in society, work and education has made it essential to overhaul school curricula in order to make the educational motto "no one left behind" a reality. In the education sector, reforms are needed in order to establish high and objective standards that can improve individual results. These standards should not be limited to the development of digital skills but should focus instead on making sure that each person can fully express their capabilities and on working for the good of the community, even when there is no personal benefit to be gained from this.

As we design and plan for the society of tomorrow, the use of AI must follow forms of action that are socially oriented, creative, connective, productive, responsible, and capable of having a positive impact on the personal and social life of younger generations. The social and ethical impact of AI must be also at the core of educational activities of AI.

The main aim of this education must be to raise awareness of the opportunities and also the possible critical issues posed by AI from the perspective of social inclusion and individual respect.

RIGHTS

The development of AI in the service of humankind and the planet must be reflected in regulations and principles that protect people – particularly the weak and the underprivileged – and natural environments. The ethical commitment of all the stakeholders involved is a crucial starting point; to make this future a reality, values, principles, and in some cases, legal regulations, are absolutely indispensable in order to support, structure and guide this process.

To develop and implement AI systems that benefit humanity and the planet while acting as tools to build and maintain international peace, the development of AI must go hand in hand with robust digital security measures.

In order for AI to act as a tool for the good of humanity and the planet, we must put the topic of protecting human rights in the digital era at the heart of public debate. The time has come to question whether new forms of automation and algorithmic activity necessitate the development of stronger responsibilities. In particular, it will be essential to consider some form of “duty of explanation”: we must think about making not only the decision-making criteria of AI-based algorithmic agents understandable, but also their purpose and objectives. These devices must be able to offer individuals information on the logic behind the algorithms used to make decisions. This will increase transparency, traceability and responsibility, making the computer-aided decision-making process more valid.

New forms of regulation must be encouraged to promote transparency and compliance with ethical principles, especially for advanced technologies that have a higher risk of impacting human rights, such as facial recognition.

To achieve these objectives, we must set out from the very beginning of each algorithm’s development with an “algor-ethical” vision, i.e. an approach of ethics by design. Designing and planning AI systems that we can trust involves seeking a consensus among political decision-makers, UN system agencies and other intergovernmental organisations, researchers, the world of academia and representatives of non-governmental organizations regarding the ethical principles that should be built into these technologies. For this reason, the sponsors of the call express their desire to work together, in this context and at a national and international level, to promote “algor-ethics”, namely the ethical use of AI as defined by the following principles:

1. **Transparency:** *in principle, AI systems must be explainable;*
2. **Inclusion:** *the needs of all human beings must be taken into consideration so that everyone can benefit and all individuals can be offered the best possible conditions to express themselves and develop;*
3. **Responsibility:** *those who design and deploy the use of AI must proceed with responsibility and transparency;*
4. **Impartiality:** *do not create or act according to bias, thus safeguarding fairness and human dignity;*
5. **Reliability:** *AI systems must be able to work reliably;*
6. **Security:** *and privacy: AI systems must work securely and respect the privacy of users.*

These principles are fundamental elements of good innovation.

Rome, February 28th, 2020



ABSTRACTS OF POSTERS

Validation of Artificial Intelligence in Medical Diagnosis, utilizing models traditionally used in the Financial Industry

Adrian Attard Trevisan (*Aberystwyth University - United Kingdom*)

Artificial intelligence (AI) and machine learning (ML) have promising prospects in the healthcare sector where it is projected to take up some of health workers' responsibilities and optimize work processes. As of now, AI and ML have found their use in anomaly detection, predictive modeling, and scoring systems. Some of the algorithms that are only emerging in the healthcare sector are already widely used in finance. The question arises as to how compatible these algorithms with the current needs of the healthcare system and what possible problems may occur when validating them. Scoring systems in medicine give rise to a reasonable doubt concerning their ethicism and precision. The validation of predictive modeling and anomaly detection largely used in finance may be challenged in the light of new scientific findings that require ongoing readjustment. Lastly, the healthcare sector suffers from the lack of cohesive shared databases, which would slow down validation and implementation of the new algorithms.

General Views of Bioethicists in Bulgaria about Artificial Intelligence in Medicine

Silviya Aleksandrova-Yankulovska (*Medical University-Pleven - Bulgaria*)

This work aims at presenting some aspects of the views of members of Bulgarian Association for Bioethics and Clinical Ethics (BABCE) on AI in medicine. A long-term goal is to continue discussion within BABCE in view of studying further public attitude towards AI and to stimulate debate at a national level.

Methodology: Focus group among members of BABCE. Questionnaire of 14 questions and qualitative analysis were applied.

Results: All participants in the discussion considered themselves familiar with AI and make a distinction with robotics. The difference with the robotics does matter in the ethical debate. However, most of the participants consider that the national debate on the application of

AI and robotics in medicine is insufficient. Suitable areas for application of AI included: personalized medicine, gene sequencing, imaging, disaster medicine, intensive care, diagnostics, out-patient care assistance, space medicine. Some of the foreseen ethical problems were: issues of control over the technology, confidentiality of patient's data, conflicts with patient's autonomy, trust, resource allocation issues, dehumanization, responsibility issues. Terminal care, pediatrics and psychiatry were pointed as areas where AI shall not be applied. There was a shared opinion that the application of AI in medicine must be controlled by the professional organizations, interdisciplinary ethics committees, patients' organizations, the public.

Conclusion: Development of medicine challenges health professionals, patients as well as bioethicists to develop together a framework for effective and safe application of the technology in line with the public values. In some countries, like Bulgaria, technologies come a little bit later than in Western Europe that shall be seen as an advantage for the ethical debate and public preparedness for welcoming or rejecting the new technology.

BABCE members: Antonia Grigorova, Makreta Draganova, Atanas Anov, Martin Mirchev, Anelia Koteva, Maria Radeva, Viktoria Atanassova, Lubev Veskov, Nikolai Yordanov, Albena Kerekovska, Desislava Bakova, Neviana Feschieva.

"If they asked you to jump off a cliff?": AI and clinical decision-making

Helen Smith (*Centre for Ethics in Medicine, University of Bristol - United Kingdom*)

Technologists have developed artificially intelligent (AI) powered systems to aid clinical decision-making; some have been deployed into healthcare. It is not always known how those systems make their decisions (known as the black box problem). My Ph.D. research has analysed the legal basis of this scenario as it relates to the clinician and the technologist; I am currently testing how the outcomes of my legal analysis can be challenged ethically. An AI being a black box is problematic due to the professional requirement for a clinician to be accountable for the patient's care.¹ If the clinician cannot explain their

¹ General Medical Council. *Good medical practice*. London, 2013. (Accessed on 09.07.2020 at: <https://www.gmc-uk.org/-/media/documents/good-medical-practice---eng>

decision making their practice is not adequately accountable. There is evidence that technologists are using this to their advantage by deploying a system whilst stating that their system "does not make decisions on what a doctor should do".² My legal analysis showed that novus actus interveniens is a problem: the clinician performs a new intervening act if they choose to use the system's outputs for a patient.

If a technologist's system's output is harmful, the clinician's action of using that output could be found as the cause of that harm, thus the technologist is deemed not liable and the clinician could be pursued for a negligence claim. Through deployment of black box systems, technologists may influence the decision making of clinicians, but without thorough prior consideration, we are allowing technologists to intimately interfere with the clinical decision-making process without ensuring that they have the opportunity to take responsibility for their contribution.

Ethically, I am concerned for the clinical professionals potentially holding singular responsibility for the consequences of black box system use; I am currently considering how the technologist could share legal and ethical responsibility if their system has influenced the clinician and therefore contributed to harms caused.

Artificial Intelligence, Offender Rehabilitation & Restorative Justice

Ana Catarina Alves Pereira (*Leuven Institute of Criminology, KU Leuven - Belgium*)

The application of a penal punishment as a reaction to crime is grounded on the anthropological view of the human being as a moral agent capable of choice and, thereby, a subject responsible for his actions. However, a conflicting, deterministic anthropological view can be found at the base of the "dominant rehabilitation model in the correctional domain, the Risk-Need-Responsivity Model" (RNR),

lish-1215_pdf-51527435.pdf); Health & Care Professions Council. *Standards of Conduct, Performance and Ethics*. London, 2016. (Accessed on 09.07.2020 at: <https://www.hcpc-uk.org/publications/standards/index.asp?id=38>); Nursing and Midwifery Council. *The Code for Nurses and Midwives*. London, 2018. (Accessed on 09.07.2020 at: <https://www.nmc.org.uk/standards/code/read-the-code-online/>).

² Hengstler M, Enkel E & Duelli S. *Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices*. Technological Forecasting & Social Change, 105; 2016: 105-120.

which "sees the offender as a bearer of risks and as a passive object of the intervention, just as the machine to be repaired is viewed by the engineer".³ Under the rationales of the RNR model, risk assessment tools are amongst the most common applications of Artificial Intelligence technology to criminal justice according to the 2018 Global Meeting on the Opportunities and Risks of AI and Robotics for Law Enforcement. These risk assessment tools, currently already heavily used in western correctional and probation services, calculate, based on the detection and weighing of static (e.g. criminal history) and dynamic risk factors, the individual's recidivism risk or probability, for crime in general and/or for specific types of crime, such as, for example, sexual crime. In turn, this risk evaluation is used for purposes of tailoring the 'treatment' necessary to modify dynamic risk factors presented by the individual, or answer the individual's criminogenic needs, in prison or in probation, influence parole decision-making and monitoring the individual after re-entry into the community. We propose to conclude our poster with the presentation of the alternative Good Lives Model, a rehabilitation model that does not preclude risk management but places a crucial emphasis on human agency. We explore how the GLM can contribute to a more restorative criminal justice, as defended by His Holiness Pope Francis at the 2019 World Congress of the International Association of Penal Law.

Ontological Plasticity and the Challenge to Anthropocentrism: Invoking Ethical Parity in Material Relations

Denis Larrivee (*Loyola University Chicago - USA; University of Navarra Medical School - Spain*)

Tacitly acknowledged in neuroscientific and technological research is an ethical imperative prioritizing value in the human being for whom the understanding or advance is intended to benefit. Termed anthropocentrism, such prioritization places human beings at the apex of organismal life and grounds ethical, bioethical, and neuroethical praxis, thereby promoting human flourishing while simultaneously restricting harmful intervention in the human being. Anthropocentrism, however, has been

³ Walgrave L, Ward T & Zinsstag E. *When restorative justice meets the Good Lives Model: Contributing to a criminology of trust*. 2019: 3.

challenged a) ethically, for its perceived placement of value in the human being alone and b) philosophically, in certain metaphysical approaches on the nature of being, philosophy of science accounts that predicate human properties in networks of entities rather than in human entities alone, and mechanist conceptions of human nature. Together, these challenges replace anthropocentrism with a value architecture that is more inclusive and technocratic, neither delimited nor determined by property attribution. The trend toward horizontality undertaken in ethical parity models, however, poses a multidimensional challenge to an ethics prioritizing the human being, a challenge mediated at the level of the ethical subject, i.e., in the siting of value contingency, in its theory of ethics, i.e., in how ethics is normatively anchored, and in ethical praxis. In consequence, it modifies ethical mediation as an intentionalized moral enactment, which is framed by a referential ontology. Conversely, philosophy of science inferences drawn from neuroscience suggest that ontological qualifications are fundamental properties of living systems, distinguishing them from technical devices and artificial biological systems. These latter findings thus offer ground for anthropocentric models, situating them in 'meta' physical principles governing the assembly of neural organization. This poster will review the multidimensional changes entailed in ethical parity models and contrast these with a modified anthropocentric model of ethical stewardship, which is premised on meta principles governing the emergence of ontological hierarchy.

Human-Centric Algorithms in Healthcare 4.0: The Agenda of Campus Bio-Medico for a Good Polyclinic

Laura Corti, Luca Capone, Paolo Soda, Marta Bertolaso (*Campus Bio-Medico University, Rome - Italy*)

Healthcare 4.0 would bring the following improvements: strengthen prevention processes, improve health systems' sustainability, make better care services for chronic patients and aged patients. One of the main issues is that there can be no sustainability without solidarity in the care processes at every level and across levels. Technology can help, but we need a human-centric approach in which human being is at the centre of progress that is, investing on the awareness of all the players in the care processes of the entwined but integrated dynamics that hold the integral development of any living system and its development (personal, functional and cultural). Therefore, it is necessary to develop

new technologies able to involve the patient actively in the clinical process in a different way.

Developing human-centric algorithms moreover means that the AI system has to be equally user-friendly for the stakeholders, safe on privacy, transparent and connected with the healthcare system. The case study, we have considered, is Campus Bio-Medico University, that works with an ecosystem of research units, focused on the integration of Artificial Intelligence in the biomedical context.

The CESA (Center of Healthcare of the Elderly), the University Hospital and the future Dea are great examples of the application of the human-centric paradigm.

Fit for Purpose? The GDPR and the European Governance of Health-Related AI Technologies

Luca Marelli (*Marie Skłodowska-Curie Fellow, Centre for Sociological Research, KU Leuven - Belgium*)

In spite of their promise for research and care, the rise of artificial intelligence (AI) technologies and advanced big data analytics within the health domain is fraught with significant ethical, societal, and legal concerns. Prominent among these are challenges related to large-scale processing of (sensitive) personal data, which call for the establishment of ethically sound and socially robust data governance mechanisms. In the European Union, the introduction of the General Data Protection Regulation (GDPR) in 2018 served as the cornerstone of its newly unfolding data governance regime. Informed by principles and values such as privacy, accountability, transparency, and fairness, the GDPR is premised on the objective to effectively balance the protection of European citizens and the promotion of a thriving European Digital Single Market and data economy.

Still, shortcomings of this regulatory effort have been noted by recent ethical, socio-political, legal, and policy scholarship. Focusing on the deployment of health-related AI technologies and big data practices with the European digital health ecosystem, this poster charts the main lines of tension emerging between the current GDPR-based data governance regime and the broader societal shifts coming along with the expansion of AI in health research and care.

Central aspects of the GDPR – i.e. key underlying data protection principles and regulatory categories, the reliance on the “notice-and-

consent” model, the (narrow) remit of the Regulation vis-à-vis harms and discriminatory practices related to personal data processing – are misaligned with the surge in big data practices and AI technologies. This throws into doubt whether the Regulation is fully fit for the purpose of governing current developments in this field. Failing to address these criticalities with adequate policy responses poses obstacles to reaping the societal benefits of AI-based innovation, and it diminishes safeguards for the individual citizens of European nations and the European community at large.

ARTificial Intelligence

Caroline Lawitschka, Philip König (*University of Vienna - Austria*)

Software engineers are coming up with new and gradually more sophisticated programs to generate art, be it musical or visual. AIVA is an AI-composer that produces music for movies, games and even its own record, which was released in 2017. GAN is another AI that can produce visual artworks based on art-historical currents such as impressionism or expressionism. Even more sophisticated is CAN, which can not only recognize different art styles but based on a database, can generate new styles and forms of art. Anticipating a more mainstream approach to art generating AI gives rise to a multitude of philosophical questions: How will such art affect our understanding of art as a category? How will it change the artistic landscape in terms of exhibitions, collaborations and such?

Ethical Problems of Using Artificial Intelligence in Medicine

Elena Vvedenskaia (*Pirogov Russian National Research Medical University, Moscow - Russia*)

AI systems are in demand by doctors when solving various tasks: assessing the probability of complications of diseases; collecting patient data; helping to make diagnoses and prescribe treatment; analyzing data of seriously ill patients in real time. Medical care through AI systems is more focused on disease prevention, contributing to improved public health. Despite the advantages of using AI in medicine, there

are negative consequences for patients and doctors. Thus, the use of these technologies for the sake of effective treatment leads to the problem of violating the right of patients to privacy and maintaining the confidentiality of personal data, to the disclosure of medical secrets, which threatens the loss of privacy. Data from the e-card used for artificial intelligence training may be available to the insurance company, which will increase the price of the medical policy and life insurance if the patient does not lead a "healthy" lifestyle and does not follow all the doctor's recommendations for treatment. The employer may refuse to employ an applicant if it has information about the presence of chronic diseases and / or genetic predisposition to certain types of diseases.

There is a real threat of discrimination against people based on physical and genetic characteristics. Questions also arise: who is the true owner of medical data, and who can manage it to what extent - the patient, doctor, clinic, insurance company, employer, or computing service? It should be noted that a doctor cannot rely on "smart algorithms" completely. Cognitive systems have problems with the quality and volume of medical information. When using the algorithm in medicine, there is a probability of a diagnostic error that can occur at the first two stages of detection and perception of symptoms: recognition of the leading manifestations and identification of the decisive signs of the disease.

Recent Results and Activities in Trustworthy Artificial Intelligence

Francesca Alessandra Lisi (*Università degli Studi di Bari "Aldo Moro" - Italy*)

The growing number of successful AI applications raises several new issues, notably the need to increase the degree of trust in AI technologies. According to the Guidelines presented by the High-Level Expert Group on Artificial Intelligence, trustworthy AI should be: (1) lawful, i.e. compliant with all applicable laws and regulations; (2) ethical, i.e. not violating ethical principles and values; (3) robust, from both a technical and social perspective. Ethics come into play in many AI applications. For instance, the problem of evaluating the ethical behaviour of AI-based chatbots in customer service has been addressed by

Dyoub et al.⁴ Here, the proposed approach combines two logic-based AI techniques, Answer Set Programming (ASP) and Inductive Logic Programming (ILP), for defining the detailed ethical rules that cover real-world situations from interactions with customers over time. ASP is appropriate for representing and reasoning with ethical rules because it can deal with norms and exceptions, whereas ILP can automatically generate those ethical rules that are difficult to encode manually. Diversity, non-discrimination and fairness are also among the requirements covered in HLEGAII, 2019. Algorithmic biases must be avoided, as they could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination, e.g based on gender or race. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle. With reference to gender, a number of initiatives have been recently undertaken, among which the ACM WomENcourage 2019 workshop "Gendering ICT"⁵ addressed the twofold problem of including the gender dimension in computer science/engineering and increasing the presence of women in the field. The workshop also stressed the importance of paying more attention to how data are collected, processed and organized in machine learning applications.

Components of the Digital Technological Revolution: Algorithm, Artificial Intelligence and Digital Communication, and its Impact between Young Mexicans

Fernando Huerta Vilchis, Íñigo Fernández Fernández (*Universidad Panamericana, Mexico City - Mexico*)

In our proposal we want to present some ideas that support the relevance of what we call the "Components of the Digital Technological Revolution", which are three: algorithms, artificial intelligence and digital communication. Thanks to the growing dominance of digital technology, these elements operate closely together and have converted the organizations that efficiently manage them in social entities with an enormous potential, which forces us to reflect whether the Digital

⁴ Dyoub A, Costantini S, Lisi F.A. *Towards Ethical Machines Via Logic Programming*. ICLP Technical Communications, 2019.

⁵ Accessed on 09.07.2020 at <http://www.di.uniba.it/~lisi/genderingICT/>.

Technological Revolution is accompanied by an ethical sense for those who operate it and use it.

We understand that although these three elements maintain a continuous interaction, it is that of communication the one that has special importance because thanks to it the contents of the other two can reach human beings and be used.

Communicating is not a mere act of transmitting the results produced by the algorithm or the "decisions made" by artificial intelligence, on the contrary, it is an act of generosity that involves sharing with others in order to achieve the common good.

In this exercise, we wish to show the contributions and impact of the Components of the Digital Technological Revolution in the day-to-day life of Mexican society through its use in digital communication, especially in social networks to subsequently carry out an assessment of their employment in which we establish whether ethics guides the use of this technology among young Mexicans.

The Dark Side of Consumer-Smart Object Relationship: a Non-User Perspective

Luigi Monsurro, Ilaria Querci, Paolo Peverini, Simona Romani
(Sapienza Università di Roma - Italy)

Smart Objects, such as Fitbit devices or Amazon Echo, promise to become an essential presence in consumer life and routines. Due to their capabilities, such as the ability to talk, to "understand" the consumer through data and to customize their services, these devices can be recognized as a social entity and also play different kinds of social roles. However, the diffusion of Smart Object is not meeting the expectation. The resistance to technologies, indeed, is not a novel phenomenon: many frameworks in the literature examine the barriers that a consumer can have toward technological devices, even in the smart technology domain. However, these models, since they do not consider the possibility that the Smart Object can interpret a social role, may be inadequate to understand the resistance toward these devices fully. Pivoting on Smart Object social roles, instead, the relational approach, already used in the marketing literature, can be an appropriate tool to understand the non-user resistance toward these innovative devices with anthropomorphic features. Using ZMET interviews involving non-users, four types of fear emerged, each one connected with a social role played by the

Smart Object: Fear of Being Controlled (the Smart Object as a Stalker); Fear of Being Dominated (the Smart Object as a Captor); Fear of Being Subordinated (the Smart Object as a Master); Fear of Losing Self-Control (the Smart Object as a Seducer). On the one hand, this work offers interesting insights about a new and unexplored barrier that has to be further examined: the relational barrier. On the other hand, applying the relational approach toward non-users, new kinds of social roles of the Smart Object, uncovered by the previous literature, emerged.

Sociological View of Medicine of the Future

Natalia V. Prisyazhnaya (*Institute of social sciences of Federal State Autonomous Educational Institution of Higher Education - Russia*)

The emergence and expansion of Internet space, the existence of virtual reality, the development of artificial intelligence, robotic medicine, the use of neural networks, Big Data arrays in health care – poses a number of challenges to modern society and medicine of the future, giving, on the one hand, very large – unprecedented before – opportunities for the development and introduction of new technologies into medicine (As well as for their scientific and practical study), and on the other hand, actualizes the need for self-determination in the new reality of members of society. The introduction of new technological solutions into the practice of health care defines new requirements to the level of professional training of medical specialists. At the same time, trends in medical education are determined by the processes of digitalization of the industry and the global challenges of mankind. In turn, the expected consequence of digitalization of medical education and health care will be the transformation of the social role of the doctor in the short term. According to the results of the research carried out by the Institute of Social Sciences of Sechenovsky University "Medicine of the Future in the Representations of Medical Specialists of the Senior Level",⁶ medical specialists of the senior level highlight a number of trends in the development of medicine of the future, among which:

- Wide introduction of new technologies into the practice of medical activity (artificial intelligence, robotics, genomic interventions, distribution of bio- and neuroimplants, medical gadgets, etc.);

⁶ Essay analysis, n = 204, 2018-2009, Moscow.

- Acquisition of new knowledge to ensure the recovery of most known diseases, increase of life expectancy (up to immortality);
- Changing the role of the doctor (displacing traditional specialties and levelling the value of the doctor's knowledge, reviewing the list of doctor's competences necessary for work);
- Changing the patient's "consciousness"
- and, above all, involvement in a healthy lifestyle, acceptance of cyberorganization processes as a norm, spread of transhumanism.

Thus, it is obvious that the medicine of the future should integrate the social phenomena of digitalization and Informatization of society that already exist in the present and become a technologized and digital area of population health management.

AI: Four Questions for the Great Challenge of the 21st Century

Álvaro José García-Tejedor, Vicente García Plá (*CEIEC Research Institute - Universidad Francisco de Vitoria - Spain*)

Understanding how the human mind works is one of the frontiers of present-day science. This interest led last century to the emergence of Artificial Intelligence, whose objective is to understand the high-level cognitive processes that characterize us as human beings as well as their implementation in computational systems. The advances already made are so important and extend transversally in such a large number of other disciplines that it is necessary to analyze what the implications of this overwhelming intrusion of which we are only partly aware are.

Beyond scientific-technical approaches, AI interrogates us with four questions that force us to rethink the plausible scenario of the advances in this area and their influence on man and society:

- Can a machine think? This raises the epistemological question: What is consciousness? Can a machine really think or only partially imitate a human-like way of responding and acting?
- Is a thinking machine human? The underlying anthropological question is: What are the attributes of the person that are unique and specific? What would the relationship with people and their integration into society be like if a machine develops self-consciousness?
- Can a thinking machine be bad/good? This leads us to the ethical question: Can moral/ethical answers be expected in the actions of an

AI? And in human actions in front of machines? What is the impact of AI on man/society?

- Do we want a machine like that? We face the question of meaning: Is this search the fruit of the desire to contribute to the common good? Does it respond to the individual's interest in demonstrating technical superiority?

These questions need to be addressed in a transdisciplinary way and from a deep knowledge in different research fields. To this end, we propose the elaboration of a "Cyberanthropology Dictionary (or Lexicon)" to unify language and terms, laying the foundations for dialogue between different disciplines.

A Taxonomy of Artificial Intelligence Opacity

Manuel Schneider, Agata Ferretti, Alessandro Blasimme (*ETH Zürich - Switzerland*)

An ethical concern that is often raised with artificial intelligence is the opaqueness of its inner workings. This point is particularly relevant for systems incorporating machine learning in which the machine 'learns' on its own how to best solve a given task and encodes the knowledge necessary to solve that task in the system. The learned knowledge representation is usually not in a form understandable by humans and the 'decisions' of the system are hard to comprehend. For that reason, AI systems, especially when machine learning is used, are often considered to be black boxes.

However, researchers demonstrated that for certain types of applications part of the AI system's learned decision logic can be understood. This indicates that the inner workings of an AI system might not be as opaque as they seem and, further, that a system's degree of opacity depends on how one defines opacity. Therefore, we analysed different mechanisms that contribute to the notion of opacity. We distinguish three types of opacity: i) lack of disclosure, ii) epistemic opacity, and iii) explanatory opacity.

We show that opacity can be the result of both technical and human factors. Such a framework can inform the discussion on opacity and help to determine strategies on how to reduce it.

Artificial Intelligence and Sensitive Thought

Giovanni Amendola (*University of Calabria - Italy*)

In today's landscape, we are witnessing a technological development incomparable to any artifice created in the past by man, so much so that we can talk about a new and further global revolution. This revolution finds its theoretical foundation, in addition to the advancements of mathematical, physical and natural sciences, on computer science and, in particular, on the logical-mathematical notion of algorithm and calculability developed particularly by Turing and in parallel by Church. Despite the theoretical limits of calculability, strictly connected with Gödel's undecidability theorems, the paradigm underlying this scientific approach has been oriented towards the achievement of tasks typically considered pertinent to the human being, initiating a new science, which finds application in almost all areas of human knowledge, precisely Artificial Intelligence.

We will try to show how Artificial Intelligence, founded on calculability, can be conceived as a sort of extension of a well-determined form of human thought, that defined by Heidegger as "calculating thought", whose roots have been recognized by some in rationality of clear and distinct ideas, which have played a decisive role in the methodological framework of modern sciences and beyond.

Although this perspective appeared as dominant in the nineteenth and first half of the twentieth century and continues to deeply connote the socio-political characteristics of western societies, in its economic-financial and techno-bureaucratic apparatuses, a different thought emerges from different perspectives and beyond the calculation. It is a rationality that is no longer aseptic and cold, but sensitive, at the height of human experience, made up of sufferings and joys, of anxieties and hopes, of a search for meaning, love and justice. Finally, we believe that it is possible to find the traces of such a "sensitive thought" within the Jewish-Christian revelation, where human intelligence carries the signs of the divine Logos.

CA17124 DigForASP: A European cooperative action for AI Applications in Police and Digital Investigations

Raffaele Olivieri, Stefania Costantini, Francesca Lisi, Jesus Medina Moreno (*Cost Action CA17124, Universita dell'Aquila - Italy*)

In the frame of Police Investigations, in particular to Digital Investigations and Digital Forensics cases, data collection on "crime scene" needs further elaboration for the contextualization in the real case. The "Evidence Analysis" phase has the aim to provide objective data and suitable elaboration of these data can help the Investigators in the formulation of possible investigative hypotheses, which could later be presented as proofs of evidence in courts. Investigations with a high amount of heterogeneous data represent a huge problem for the human mind in the search for events, connections, facts or demonstrate alternative solutions. However, many investigative problems can be formalized and expressed with a mathematical approach and solved with reasonable efficiency using Artificial Intelligence and Automatic Reasoning. COST Action CA17124, called DigForASP ("DIGITAL FORensics: analysis tests through intelligent systems and practices"), financed by the European Union with the funds for "European cooperation in science and technology, Horizon 2020", was born for the exploration, study the delicate issue of the application of Artificial Intelligence and Automated Reasoning to the investigative world, through the creation of a multidisciplinary scientific network. DigForASP, with activities in the period September 2018 - September 2022, has aims to help the human operator (Law Enforcement, Lawyers, Public Prosecutors, Judges, social scientists, criminologists) in the analysis of investigative data as well as the formulation of hypotheses for the resolution of complex cases, through Artificial Intelligence techniques available to guarantee ethic, reliability and verifiability.

Artificial Intelligence & Pluralistic Global Bioethics: Thomistic-Aristotelian Personalist Refinement of the United Nations' Social Contract View of Rights-duties in AI-genetic Engineered Nanotechnology

Dominique J. Monlezun, Claudia Sotomayor, Colleen M. Gallagher, Alberto Garcia (*Artificial Intelligence & Advanced Analytics Center, Cardiac Catheterization Laboratory, Department, New Orleans - USA*)

Introduction: Artificial intelligence (AI)-guided genetic engineered nanotechnology and robotics (AI-GNR) is widely recognized as the technological revolution posing the greatest transformative potential to humanity; it has already demonstrated its technical capacity to permanently alter the biology and physics governing the global human family. Yet there are no substantive and pluralistic ethical or legal analyses for AI-GNR—despite its real and imminent apocalyptic potential. This analysis therefore seeks to provide the first substantive and comprehensive global bioethical, legal, and health analysis of AI-GNR by providing the first known defense of the world's only global bioethics utilized by every nation on our planet.

Methods/Results: This study historically and philosophically defines the Thomistic-Aristotelian personalist foundation of the rights and duties-based social contract framework of the United Nations (UN) as articulated in the 1948 Universal Declaration of Human Rights (UDHR) which formed the basis of all subsequent UN instruments (including the 2005 Universal Declaration of Bioethics & Human Rights [UDBHR]) and thus modern international law, which serves as the single most influential ethical and legal body on state-level legislation of technology that includes AI-GNR. This study demonstrates the superior philosophical strengths (in metaphysical, formal logic, and ethical terms) of this personalism compared to the dominant competing modern ethics, in addition to its unique advantage of facilitating convergence of pluralistic belief systems to common ethical conclusions. It then applies this approach with a historic level of concrete specification to AI-GNR in its ethical, legal, and health aspects.

Discussion: AI-GNR is already re-shaping humanity at a level, speed, and permanence never before seen. This study provides the first known definition and defense of a global bioethics that can unite the world in a common philosophical language already animating an ongoing political mission of enduring peace, and thus may help save humanity from AI-GNR's worst cataclysmic capacity.

Ethical Problem of the Trademark Registration for “NEON Artificial Human”

Youngjin Jin (*The Catholic University of Korea, Seul - South Korea*)

STAR Labs have developed and launched what they call an “artificial human”. According to the STAR Labs, this “artificial human” resembles actual human beings and has the ability to sympathize with a real person via real-time conversations. STAR Labs named it “NEON Artificial Human” and applied for its trademark registration. However, permitting this trademark registration involves an ethical problem because “artificial” means that it was created by human technology, indicating that the artificial human is a human being created using human technology. Creating an artificial human and granting its trademark registration would establish that humans can also be commercialized, thereby undermining human dignity. Thus, I examine the following four points. First, consumers experience the reality as well as the virtual world while using STAR Labs products, which can cause confusion regarding human identity. The trademark registration for “NEON Artificial Human” can further aggravate this confusion. The term “Artificial Human” stands out more to consumers than the ambiguous word “NEON”. Second, the research and the pursuit of profit by companies in relation to artificial intelligence (AI) must be premised on minimal AI ethics for the global human community. If the research aims to create another species of humans as the STAR Labs CEO say, it must not be researched based on the AI ethics of each company. But before that, a crucial question must be answered in advance: Can we really allow humans to create other humans? Third, the most important thing is the corporate will and effort to comply with the AI ethics. The case of He Jiankui, who created the first human genetically edited babies overshadowing “On Human Gene Editing: International Summit Statement”, shows that the same can happen in AI research and development. Fourth, the paradigm must be changed to actively accept AI ethics for trademark examination as well. We humans have not yet answered the question “Does humankind really want to create a new human species, albeit an ‘artificial’ one?” Therefore, we must not grant the exclusive rights of trademarks that suggest or imply the creation of a new human species as an AI technology-related product.

Human-in-the-loop Artificial Intelligence

Fabio Massimo Zanzotto (*University of Rome Tor Vergata - Italy*)

Little by little, newspapers are revealing the bright future that Artificial Intelligence (AI) is building. Intelligent machines will help everywhere. However, this bright future may have a possible dark side: a dramatic job market contraction before its unpredictable transformation. Hence, in a near future, large numbers of job seekers may need financial support while catching up with these novel unpredictable jobs. This possible job market crisis has an antidote inside. In fact, the rise of AI is sustained by the biggest knowledge theft of the recent years. Many learning AI machines are extracting knowledge from unaware skilled or unskilled workers by analyzing their interactions. By passionately doing their jobs, many of these workers are shooting themselves in the feet. In this paper, we propose Human-in-the-loop Artificial Intelligence (HitAI) as a fairer paradigm for AI systems. Recognizing that any AI system has humans in the loop, HitAI will reward these aware and unaware knowledge producers with a different scheme: decisions of AI systems generating revenues will repay the legitimate owners of the knowledge used for taking those decisions. As modern Merry Men, HitAI researchers should fight for a fairer Robin Hood Artificial Intelligence that gives back what it steals.

Does Artificial Intelligence Have a Purpose?

Juan Jesús Gutierrez (*Universidad Católica de Ávila - Universidad Pontificia Comillas - Spain*)

Following the philosopher Hans Jonas, on our poster, we will ask whether the machines with 'intelligence' have a finality. Going into the concept of the finality we will understand this as a purpose, as that which gives direction to the action, behavior aimed at achieving an objective.

However, to reach this, movement (effectors) and sensation (ability to perceive, receptors) are not enough, but the willingness is necessary. That is, "behavior according to purposes demands the presence of purposes". Could one explain, as the cybernetics intended, the behavior towards a purpose without a purpose? It is appropriate to address the confusion between 'making a purpose' and 'having a purpose'. The

separability, Jonas will say, between the purposes and its realization allows the latter to be delegated and distributed among many people without them even knowing the objective in question.

For Jonas the human being, the animal and, in general, living organisms are needy and indigent, thus creating a close union between need (metabolism) and impulse (to prolong existence). It is the emotions, and not only the data of the receivers, what creates goals and purposes. Thus, it is an interest that drives and makes the animal move. The animal is not only perception and movement but also feeling, which connects the previous two and is already present in the undifferentiated and pre animal phase in the continuous metabolic exchange.

The gradual difference of the human being will be that they want to; that is, they have intentionality and suffers when they fail. Nevertheless, neither 'suffering' nor 'joy', nor 'success' nor 'failure', nor 'satisfaction' nor 'frustration' follow to the modus operandi of a machine [...].

Therefore, on our poster, we will approach the question of purpose, will, emotions or moral acts in Artificial Intelligence.

Artificial Intelligence and the Future of Nursing Profession

Sung Hee Ahn (*The Catholic University of Korea, Seoul - South Korea*)

Artificial intelligence (AI) is already around us. The main functions of AI in healthcare are in learning situations, planning simulations for practice, problem-solving tools, and even speech recognition. AI technologies are being developed to improve patient management and outcomes.

This paper examines AI's nursing applications and their positive and negative aspects to provide future prospects for nursing professionals. Examples of AI applications for improving nursing education with intelligent systems are protocols and guidelines, automatic diagnosing and decision support tools, temporal reasoning and planning, natural language and terminology, image and signal processing.

Examples for nursing practice with AI are electronic health records, voice electronic nursing record systems, triage nurse, virtual nurse platform for intervention, automated guided vehicles. Positive outcomes of AI in nursing practice as follows: AI could help nurses with paperwork and leave them more time for patients. Negative outcomes are as fol-

lows: overreliance on AI technologies may depersonalize nurse-patient interactions and erode rapport, accountability toward AI. Decision making for patients via AI algorithm will be a chance for regardless of the patients' desire, technological literacy, and economic means, and violation of patients' autonomy, privacy, and confidentiality in inpatient data sharing. However, when it goes wrong, the question arises: "Who should be responsible, and can we trust AI?"

Therefore, nurses need to understand how AI can be most helpful for patients, skilled nursing education, and future practice. An essential step to this is to examine the personalist bioethical issues in nursing education and training with AI and deep ethical learning about AI application. Through this process, AI in nursing will be a system that supports advanced technology and high touch in nursing.

The Advent of Artificial Intelligence in Arts or the Creativity of Artifacts

Maria Addolorata Mangione, Alberto Carrara (*Pontifical Athenaeum Regina Apostolorum, Rome - Holy*)

The humanoid Ai-Da, fruit of art gallery director Aidan Meller's idea, represents the "first ultra-realistic humanoid robot in the world". One of the main objectives expressed by those who designed it is stir up a debate on the concept of life and on the future of humanity itself. In the poster, after the analysis of such a project, we will proceed to some brief reflections on the concept of life, followed by an examination of the concept of creativity, to evaluate its possible application to an artificial intelligence. In order to avoid moving away from the truth concerning man, we consider it fundamental not to limit ourselves to a biological-organic conception of life: life is a similar concept, and dwelling on a single meaning means losing sight of the psychic and spiritual dimensions, which together with the somatic dimension constitute the human being. The lack of a suitable distinction between artificial intelligence and the human mind is an expression of a rationality that has as its fruit a mechanistic model of nature; which holds it possible to automate exquisitely human activities, such as creating a work of art. All this proves misleading, as it leads to the neglect of other areas of application of artificial intelligence, which would favor a human-sized technological revolution.

Accessible Numbers: Artificial Intelligence and Cultural Inclusion

Luca Baraldi (*Energy Way - Education Unit, Modena - Italy*)

There cannot be a reflection on Artificial Intelligence ethics, regardless of an analysis of the impact beyond the mere technical implications. The impact of AI should, first of all, generate a profound substantial question, on the way in which human being is evolving and how the interconnected knowledge is transforming humanity. It is no longer possible to interpret social reality, in all its manifestations, regardless of constant interaction with technology.

Artificial Intelligence has become an integral part of the ecosystems of production, diffusion and circulation of knowledge and information. On the one hand, it is inevitable to recognize and accept the importance of AI for the advancement of knowledge. On the other hand, it is necessary to promote a dimension of education that allows each person to understand the uniqueness of human thought and to rediscover the substantial value of free will.

In the international context there is constant talk about new humanism, digital humanism and new anthropocentrism, but it is essential to prepare methodologies and tools that allow every level of global society to understand the profound value of these concepts. The application of AI in everyday life is transforming cognitive processes and epistemological dynamics, resizing the value of experience, radically changing the dimension of interaction, delegitimizing the central role of doubt as a tool for stimulating knowledge and discovery.

The potential for AI support to cognitive processes involves the risk of feeding dynamics that replace thought, rather than assisting it, resulting in less autonomy in the exercise of critical thinking and in enhancing imagination as a cognitive experience. Encouraging the exercise of critical thinking means, first of all, creating differentiated tools that allow each person, at every educational and cultural level, to understand what AI is and what its real impact could be on everyday life. We have the responsibility to promote conceptual accessibility tools, to find a way in which a deeper understanding of the AI phenomenon might benefit the design of accessible communication strategies and measures for cultural inclusion.

In Tech we Trust...but we need Human as a Right

Elisa Spiller (*Candidate University of Padova - Italy*)

The poster will address some issues related to the use of artificial intelligence for automatic decision-making purposes, with a specific focus on those processes that have a significant impact on people fundamental rights. The research takes as starting human dignity, exploring the importance of this principle in contemporary person-based constitutional law theory. This assumption will be key element to analyze to the relationship between two other principles that represent the two face of the current technological revolution. On the one hand, there is the principle of digital by default: a strategy based on the presumption that technology may positively contribute to the efficiency of decision-making procedures so that to make it a new right. On the other, instead, there are the issues concerning the so-called non-exclusivity principle: an assumption that aims to guarantee human supervision on automatic processes, ensuring the right to challenge data-driven decisions before a human expert operator. On these premises, the poster exposes a study on the recent case-law about AI in the EU and national case-law. In particular, the aim is to see how these decisions are fostering a right-friendly approach in the use of data-intensive technologies, even setting some preliminary legal limitations. The analysis principally converges on three main points. As first, it focuses on the relevant constitutional case-law that, over time, have set limits to the use of automatisms in the application of the law. Then it examines the different opinions concerning the principle of non-exclusivity, focusing on the reasons why should be desirable the automation of just nondiscretionary decisions. Eventually, it addresses the issues related to fairness and transparency of decision making, exploring the possible technical and legal solutions that might ensure the interests of the people involved. "In tech we trust... but we need human as a right", therefore, hopes to contribute in the ongoing interdisciplinary debate on these topics, sharing common concerns emerging in the regulation of AI. Building on the principles of constitutional law tradition and human rights literacy, the aim is to foster an appropriate translation of the related values in the design and the use of these technologies, promoting an anthropocentric approach.

IN MEMORY

Remembering Cardinal Elio Sgreccia

Antonio G. Spagnolo*

It is hard to summarise the long life and extensive work of Cardinal Elio Sgreccia in a few minutes, as they extended over a timespan of about 45 years since he arrived at the Rome branch of the Catholic University of the Sacred Heart in 1974, coming from the Marche region where he was born on 6th June 1928 and ordained in 1952.

As a spiritual assistant at the Faculty of Medicine and Surgery and at the adjoining University Hospital Agostino Gemelli, he met students and nursing staff, and it is in this unique environment that he had the opportunity to personally come into contact, "at the patient's bedside", with the sick and their illnesses, with their families and all the problems that the diseases brings with them. This played a decisive role



* Director, Institute of Bioethics and Medical Humanities at Catholic University of the Sacred Heart, Rome (Italy).

in fostering his interest in the study of bioethics first, and then in the pastoral care of life.

Cardinal Sgreccia himself described his life in his last book, an autobiography published a few months before his death: *Contro Vento. Una Vita per la Bioetica*, a lucid account of his life, I dare say 'from conception to natural death'. A spiritual testament in which he pointed at the most significant events in "his" life for bioethics.

From an academic point of view, his first adventure was the establishment in 1985 of the Centre of Bioethics at the Faculty of Medicine and Surgery, of which he was appointed Director. Among his papers I found the speech he had prepared that year to inaugurate the Centre. He wrote: "Being born, walking, thinking and learning to speak all at the same time is impossible in the development of the human being: it is on the contrary what happened to the Bioethics Centre of the Catholic University inaugurated in Rome on 20th June 1985. The news that the Catholic University has set up a Centre for the study of bioethics, the only one at university level in Italy, and was the first to have included in its Statute the teaching of this subject [bioethics], meant that the Centre itself was forced to act immediately and make its voice heard on crucial issues... In Article 3, the Statute of the Centre sets out a specific task for itself, that of being faithful to the Magisterium of the Church and to scientific rigour (...). The Centre, inspired by the principles of the Catholic University of the Sacred Heart, undertakes to make constant reference to the criteria of scientific soundness, fidelity to the Catholic vision of life and attention to the problems posed by scientific progress and social evolution".

The title of the inaugural conference, also covered by L'Osservatore Romano, described the ethical dimension as a bridge capable of connecting economy and health. According to Cardinal Sgreccia, ethics acts in human beings as the mysterious and vital force that makes a plant grow by unifying all its parts from root to leaf, from the sap that rises to the light with which it synthesizes chlorophyll. For this reason, also the cost-effectiveness and fairness of healthcare spending will have to take into account above all a personalist and ethical notion of health. Because he came from a family of farmers, Cardinal Sgreccia often used comparisons with nature and the Earth, and the topic of the environment and agriculture was always present in his lectures and writings.

The establishment of the Centre was followed in 1992 by the foundation of the University Institute of Bioethics and its academic organi-

zation. Cardinal Sgreccia passed the national competition becoming Full Professor of Forensic Medicine (one of the scientific disciplines in which bioethics is included in Italy), he became the holder of the first Chair of Bioethics in Italy, and the Faculty of Medicine and Surgery of the Catholic University appointed him Director of the first University Institute of Bioethics.

He was one of the pioneers of Bioethics in Italy, and authored the first manual on Bioethics, which he wrote over the summer of 1986 in his home in the Marche countryside. The manual was translated into several languages and is internationally known.

The structuring of his thought and methodology in the field of bioethics – based on ontologically founded personalism – led him to critically rethink the methodology of North American bioethics. He developed a unique definition of bioethics as “A discipline with a rational epistemological status, open to theology as a superrational science, the last instance and horizon of meaning. Bioethics, beginning with the description of the scientific, biological and medical data, examines rationally the lawfulness of man’s intervention on man”. It is based on these founding principles that he proposed his “triangular method” to analyse ethical problems in the medical field to discern the right decision. This methodology aimed at finding an answer different from that of “principlism” proposed by the North American method, which was based on the lack of an ethical theory of reference that in Cardinal Sgreccia’s view was that of personalism. However, he always sought dialogue with this methodology and, on the occasion of the tenth anniversary of the Bioethics Centre, invited to an international conference on the roots of bioethics precisely the pioneers of its North American counterpart: Edmund Pellegrino, Tom Beauchamp, Warren Reich ... with whom he established a fruitful dialogue that continued over time.

Again in the academic sector, since 1975 he was first co-editor, together with Prof. Angelo Fiori, then editor in chief of the magazine *Medicina e Morale*, an international journal of bioethics that, after many years of existence, started being indexed in Scopus, just in the days of Cardinal Sgreccia’s death.

On the ecclesial side, almost at the same time as the establishment of the Institute of Bioethics, on 5th November 1992 he was appointed Titular Bishop of Zama Minor and Secretary of the Pontifical Council for the Family, receiving episcopal ordination from the hands of John Paul II on 6th January of the following year.

Cardinal Sgreccia held office at the Pontifical Council for the Family until early 1996, when he devoted himself full-time to the office of Vice-President of the Pontifical Academy for Life. In June 1994, in fact, he had been appointed to this position alongside Prof. Jérôme Lejeune, the first President of the Vatican institution. He held the same office also with Lejeune's successor, Juan de Dios Vial Correa, until he was himself appointed President of the Pontifical Academy by St. John Paul II on 3rd January 2005. In this capacity, his activity was characterized above all by the publication of several documents in a series which collected the proceedings of the annual scientific congresses celebrated in conjunction with the general Assemblies of the Academy itself.

St. John Paul II's *Evangelium Vitae* (EV) was a constant reference in the activity of Cardinal Sgreccia, both in the academic world and in the Pontifical Academy for Life. According to EV no. 98, in fact, "a specific contribution [to bioethics] will also have to come from Universities, particularly Catholic Universities, and from Centres, Institutes and Committees of Bioethics". And earlier, at no. 27, the Encyclical stresses that "The emergence and ever more widespread development of bioethics is promoting more reflection and dialogue – between believers and non-believers, as well as between followers of different religions – on ethical problems, including fundamental issues pertaining to human life".

Within the Pontifical Academy for Life, on the topic of the fundamental ethical problems affecting human life, Cardinal Sgreccia's contribution has been instrumental in clarifying some issues that are particularly relevant to public moral conscience: among them, organ donation, stem cells, conscientious objection, vegetative state.

He tirelessly visited developing countries to bring the message of the Gospel and the principles of personalist bioethics to their people through the organization of training courses and activities locally.

He left the office of President of the Pontifical Academy for Life on 17th June 2008, during his 80th year of life, becoming President Emeritus. In 2010, he was elevated to the rank of Cardinal by Pope Benedict XVI. On 7th June 2019, Pope Francis presided over the rite of the *Ultima Commendatio*. Cardinal Sgreccia's studies and pastoral activities were deeply marked by three Popes, to whom he has always expressed his fidelity.

As mentioned above, he described his life journey in his last book, published a few months ago, which contains his spiritual testament. I wish to recall here his lucid message, which interpreted the events of own life as what pushed his journey forward, in the same way as

a boat is propelled by wind, to harness which we constantly need to “adjust” our sails, even when it blows against us.

The wind that we harness is an extraordinarily powerful legacy that asks us to set our small sails so as start along an adventurous “journey” towards a quest for truth, for an intercultural dialogue based on reason, open to transcendence. Hence his motto that I wish to recall in conclusion, a motto that Card Sgreccia wrote himself as an inscription on the numerous copies of his book that he gave to ‘his many children’: “The best is always to come and it is always possible”.

In his approach, the integration between the Christian doctrine and scientific subjects always played a very important role, showing that the Church can legitimately speak about science and do so in a knowledgeable manner, with rigour, intelligence and faith in humanity. This also represents the foundation of the dialogue that the Pontifical Academy for Life has developed in the course of the subsequent years, on how to give rise to that new ‘civilization of love’ on which John Paul II insisted so much, and which passes through all the phases of human life, especially during the first formative years within the family, the true womb capable of welcoming, guarding, protecting and developing a new mature and conscious life. In the words of HE Mons. Vincenzo Paglia, “Cardinal Sgreccia has been the protagonist and the brave and wise life and soul behind the Pontifical Academy for Life, supporting and promoting study activities and the protection of human life against the challenges raised by technological and biomedical progress”. And the Pontifical Academy “continues along the path traced with foresight by Cardinal Sgreccia, with the same bravery – as Pope Francis said in his Letter *Humanae Communitas* – in reading the signs of the times and providing an answer to the need for meaning expressed by our contemporaries”. For this reason, it is essential to participate in the discussion with all the actors, so that the development and use of the resources that science and technology make available to us may be oriented towards the promotion of the dignity of the human person and the highest universal good.

(Translated from Italian by the Pontifical Academy for Life)





Finished printing in February 2021
by Stabilimento Tipografico «Pliniana»
Viale F. Nardi, 12 – 06016 Selci-Lama (PG)
st.pliniana@libero.it
www.pliniana.it

PUBLICATIONS OF THE PONTIFICAL ACADEMY FOR LIFE

PROCEEDINGS

Evangelium Vitae di Sua Santità Giovanni Paolo II.
Enciclica e Commenti apparsi su "L'Osservatore Romano"
A cura della Pontificia Academia pro Vita
Città del Vaticano: Libreria Editrice Vaticana, pp. 347, 1995.

La Causa della Vita

Atti della Seconda Assemblea Generale della Pontificia Academia pro Vita (20-22 Novembre 1995)

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 179, 1996.

Comentario Interdisciplinar a la *Evangelium Vitae*

Obra dirigida por Ramón Lucas Lucas
Madrid: Biblioteca de Autores Cristianos, pp. 811, 1996.

Commento Interdisciplinare alla *Evangelium Vitae*

A cura di Elio Sgreccia e Ramón Lucas Lucas
Città del Vaticano: Libreria Editrice Vaticana, pp. 823, 1997.

Evangelium Vitae e Diritto

Atti del Convegno Internazionale (23-25 Maggio 1996)
A cura di Alfonso López Trujillo, Julián Herranz, Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 629, 1997.

Identità e Statuto dell'Embrione Umano

Autori vari: Carrasco de Paula, Colombo, Cozzoli, Eusebi, Laffitte, Leone, Lucas Lucas, Melina, Palazzani, Pessina, Serra
Città del Vaticano: Libreria Editrice Vaticana, pp. 303, 1998.

Identity and Statute of Human Embryo

Proceedings of the Third General Assembly of the Pontifical Academy for Life (14-16 February, 1997)
Edited by Juan de Dios Vial Correa and Elio Sgreccia
Vatican City: Libreria Editrice Vaticana, pp. 458, 1998.

Human Genome, Human Person and the Society of the Future

Proceedings of the Fourth General Assembly of the Pontifical Academy for Life (23-25 February, 1998)

Edited by Juan de Dios Vial Correa and Elio Sgreccia
Vatican City: Libreria Editrice Vaticana, pp. 509, 1999.

Biotecnologie Animali e Vegetali: Nuove Frontiere e Nuove Responsabilità
Autori vari: Ancora, Benvenuto, Bertoni, Buonuomo, Honings, Lauria, Lucchini, Marsan, Mele, Pessina, Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 197, 1999.

The Dignity of the Dying Person

Proceedings of the Fifth General Assembly of the Pontifical Academy for Life (24-27 February, 1999)

Edited by Juan de Dios Vial Correa and Elio Sgreccia
Vatican City: Libreria Editrice Vaticana, pp. 479, 2000.

Evangelium Vitae: Five Years of Confrontation with the Society.

Proceedings of the Sixth General Assembly of the Pontifical Academy for Life (11-14 February, 2000)

Edited by Juan de Dios Vial Correa and Elio Sgreccia
Vatican City: Libreria Editrice Vaticana, pp. 548, 2001.

La cultura della vita: fondamenti e dimensioni

Atti della Settima Assemblea Generale della Pontificia Academia pro Vita (1-4 marzo, 2001)

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 332, 2002.
(other languages: English)

Natura e dignità della persona umana a fondamento del diritto alla vita. Le sfide del contesto culturale contemporaneo

Atti dell'Ottava Assemblea Generale della Pontificia Academia pro Vita (25-27 febbraio, 2002)

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 271, 2003.
(other languages: English)

Etica della ricerca biomedica. Per una visione cristiana

Atti della Nona Assemblea Generale della Pontificia Academia pro Vita (24-26 febbraio, 2003)

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 406, 2004.
(other languages: English)

La dignità della procreazione umana e le tecnologie riproduttive: aspetti antropologici ed etici

Atti della Decima Assemblea Generale della Pontificia Academia pro Vita (20-22 febbraio, 2004)

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 304, 2005.
(other languages: English)

Qualità della vita ed etica della salute

Atti dell'Undicesima Assemblea Generale della Pontificia Academia pro Vita (21-23 febbraio, 2005)

A cura di Elio Sgreccia e Ignacio Carrasco De Paula
Città del Vaticano: Libreria Editrice Vaticana, pp. 272, 2006.
(other languages: English)

L'embrione umano nella fase del preimpianto. Aspetti scientifici e considerazioni bioetiche *Atti della Dodicesima Assemblea Generale della Pontificia Academia pro Vita (27 febbraio - 1 marzo 2007)*

A cura di Elio Sgreccia e Jean Laffitte
Città del Vaticano: Libreria Editrice Vaticana, pp. 352, 2007.
(other languages: English, Spanish)

La coscienza cristiana a sostegno del diritto alla vita

Atti della Tredicesima Assemblea Generale della Pontificia Academia pro Vita (23-25 febbraio 2007)

A cura di Elio Sgreccia e Jean Laffitte
Città del Vaticano: Libreria Editrice Vaticana, pp. 262, 2008.
(other languages: English, French, Spanish)

Accanto al malato inguaribile e al morente: orientamenti etici ed operativi

Atti della Quattordicesima Assemblea Generale della Pontificia Academia pro Vita (25-26 febbraio 2008)

A cura di Elio Sgreccia e Jean Laffitte
Città del Vaticano: Libreria Editrice Vaticana, pp. 312, 2009.
(other languages: English, French, Spanish)

Le nuove frontiere della genetica e il rischio dell'eugenetica

Atti della Quindicesima Assemblea Generale della Pontificia Academia pro Vita (20-21 febbraio 2009)

A cura di Jean Laffitte e Ignacio Carrasco de Paula
Città del Vaticano: Libreria Editrice Vaticana, pp. 312, 2010.
(other languages: English, French, Spanish)

Bioetica e legge naturale

Atti della Sedicesima Assemblea Generale della Pontificia Academia pro Vita (11-13 febbraio 2010)

A cura della Pontificia Accademia per la Vita
Città del Vaticano: Lateran University Press, pp. 174, 2010.

NEW SERIES OF PUBLICATIONS OF THE PONTIFICAL ACADEMY FOR LIFE

The Management of Infertility Today

Proceedings of a Workshop sponsored by the Pontifical Academy for Life (February 24, 2012)

International Journal of Gynecology and Obstetrics, Vol. 123, S. 2, Dec. 2013.

Faith and Human Life

Proceedings of the XIX General Assembly of Members of the Pontifical Academy for Life (February 22, 2013)

Edited by Ignacio Carrasco de Paula and Renzo Pegoraro

Città del Vaticano: Pontificia Accademia per la Vita, pp. 167, 2013.

Ageing and Disability

Proceedings of the XX General Assembly of Members of the Pontifical Academy for Life (February 21, 2014)

Edited by Ignacio Carrasco de Paula and Renzo Pegoraro

Città del Vaticano: Pontificia Accademia per la Vita, pp. 289, 2014.

Fede e Vita Umana

Atti della XIX Assemblea Generale (21-23 Febbraio 2013)

A cura di Ignacio Carrasco de Paula e Renzo Pegoraro

Città del Vaticano: Pontificia Accademia per la Vita, pp. 172, 2015.

Assisting the Elderly and Palliative Care

Proceedings of the XXI General Assembly of Members of the Pontifical Academy for Life (March 5-7, 2015)

Edited by Ignacio Carrasco de Paula and Renzo Pegoraro

Vatican City: Pontifical Academy for Life, pp. 295, 2015.

Virtues in the Ethics of Life

Proceedings of the XXII General Assembly of Members of the Pontifical Academy for Life (March 3-5, 2016)

Edited by Ignacio Carrasco de Paula, Vincenzo Paglia and Renzo Pegoraro

Vatican City: Pontifical Academy for Life, pp. 252, 2017.

Accompanying Life. New Responsibilities in the Technological Era

Proceedings of the XXIII General Assembly of Members of the Pontifical Academy for Life (October 3-5, 2017)

Edited by Vincenzo Paglia and Renzo Pegoraro

Vatican City: Pontifical Academy for Life, pp. 200, 2018.

Equal Beginnings. But then? A Global Responsibility

Proceedings of the XXIV General Assembly of Members of the Pontifical Academy for Life (June 25-27, 2018)

Edited by Vincenzo Paglia and Renzo Pegoraro
Vatican City: Pontifical Academy for Life, pp. 244, 2019.

Robo-Ethics. Humans, Machines and Health

Proceedings of the XXV General Assembly of Members of the Pontifical Academy for Life (February 25-27, 2019)

Edited by Vincenzo Paglia and Renzo Pegoraro
Vatican City: Pontifical Academy for Life, pp. 300, 2020.

STUDIES AND DOCUMENTS OF THE PONTIFICAL ACADEMY FOR LIFE

Riflessioni sulla clonazione

A cura della Pontificia Academia pro Vita
Città del Vaticano: Libreria Editrice Vaticana, pp. 21, 1997.
(other languages: English, French, Spanish, Portuguese, German, Russian, Arabic)

Dichiarazione sulla produzione e sull'uso scientifico e terapeutico delle cellule staminali embrionali umane

A cura della Pontificia Academia pro Vita
Città del Vaticano: Libreria Editrice Vaticana, pp. 19, 2000.
(other languages: English, French, Spanish, German, Portuguese)

La prospettiva degli xenotraimenti: aspetti scientifici e considerazioni etiche

A cura di Juan de Dios Vial Correa e Elio Sgreccia
Città del Vaticano: Libreria Editrice Vaticana, pp. 55, 2001.
(other languages: English, French, Spanish)

Il divieto della clonazione nel dibattito internazionale. Aspetti scientifici, etici e giuridici

A cura della Pontificia Academia pro Vita
Città del Vaticano: Libreria Editrice Vaticana, pp. 51, 2003.

L'embrione umano nella fase del preimpianto. Aspetti scientifici e considerazioni bioetiche

A cura della Pontificia Academia pro Vita
Città del Vaticano: Libreria Editrice Vaticana, pp. 45, 2006.
(other languages: English, French, Spanish)

NEW SERIES OF "STUDIES AND DOCUMENTS"

Le banche di cordone ombelicale

A cura della Pontificia Accademia pro Vita

Città del Vaticano: Pontificia Accademia per la Vita, pp. 92, 2013.

(*other languages: English*)

Post-abortion trauma. Possible psychological and existential aftermaths

Edited by the Pontifical Academy for Life

Città del Vaticano: Pontificia Accademia per la Vita, pp. 210, 2014.

Enciclica "Evangelium Vitae". Vent'anni dopo

A cura della Pontificia Accademia pro Vita

Città del Vaticano: Pontificia Accademia per la Vita, pp. 128, 2016.

White Book for Global Palliative Care Advocacy. Recommendations from the PAL-LIFE expert advisory group of the Pontifical Academy for Life, Vatican City

Edited by the Pontifical Academy for Life

Città del Vaticano: Pontificia Accademia per la Vita, pp. 104, 2019.

(*other languages: German, Italian, Portuguese*)

OTHER TEXTS

I Papi e la Pontificia Accademia per la Vita – The Popes and the Pontifical Academy for Life

Città del Vaticano: Libreria Editrice Vaticana, pp. 240, 2020.

Curare la vita. Etica e Tecnologie

Di Fabrizio Mastrofini e Nicola Valenti

Bologna: Edizioni Dehoniane, pp. 105, 2020.