Response to the Public Consultation on the European Commission «White Paper On Artificial Intelligence: A European Approach to Excellence and Trust»

This document along with signatures of more than 70 practitioners and researchers is available online.

Document is available here: http://blogs.uned.es/workshopadvancingtowards/news/.

List of signatories is available here: http://blogs.uned.es/workshopadvancingtowards/list-of-signatories/.

The European Commission has opened the consultation process on the «White Paper on Artificial Intelligence: a European Approach Oriented to Excellence and Trust».

As a group of professionals and researchers from different fields related to Artificial Intelligence (hereinafter, AI) we are making a proposal to expand and improve the current approach of the White Paper, which we believe suffers from important shortcomings.

Firstly, we believe that the European commitment to the United Nations 2030 agenda should take centre stage. The White Paper's approach in this regard is technologically reductionist, with a view of public AI policies that appear to be exclusively designed to improve competitiveness, with no consideration of AI as a tool to improve people's living conditions, with particular attention to vulnerable groups, of which Europe also has its share. As the United Nations has recognized, the Sustainable Development Goals cannot be achieved without a people-focused science-based digital revolution.

Secondly, we request that the White Paper include a strategy for the universalization of the benefits of AI. In this regard, we make two important observations, the first concerning openness in the sense of the upcoming UNESCO Open Science Recommendation: open source, open data, open algorithms, etc. are enabling factors for achieving universalization of the benefits, facilitating dissemination and auditing.

In addition to promoting openness, concrete initiatives are needed to promote data ecosystems that minimize AI biases, for example by opening public-sector data sources with high scientific and social value.

The benefit obtained from AI depends on the quality of the data. Social institutions and organizations should also have access to private sector data that has a vital or significant public interest, which implies the interconnection of diverse data sources. This also requires designing services with adequate security and privacy and that optimize the usage and accessibility of the data in this interconnection context.The second observation we make about universalizing the benefits of AI is to warn about the emergence of an AI divide that would overlap with the already worrying digital divide that remains worldwide. There is a risk

that the use of AI will lead to new inequalities and create difficulties in the endeavor to improve opportunities for vulnerable groups, creating an additional barrier: that of those who will benefit from AI, and those who will not (minorities discriminated against because of social biases and stereotypes amplified by big data techniques, small businesses at a competitive disadvantage because they lack the infrastructure required by AI, etc.).

Thirdly, we believe that the current draft of the White Paper lacks a sufficiently deep analysis of ethical and methodological issues that should lead to the development of Algorithmic Justice. There is a need to bring order to the current landscape of overabundance of ethical codes, guidelines and frameworks, many of which suffer from deficiencies such as lack of scientific rigour, subjectiveness, incoherence, superficiality and redundance, thus generating confusion. We believe that the White Paper should promote R&D in the design of applied-ethics tools and in the definition of indicators of the impact of these technologies on society. R&D in certifying the explainability and justice of algorithms is also essential.

Other important aspects discussed in the document summarized here are overcoming ethnocentric and androcentric approaches, the need for AI education and literacy, the opening of dialogue mechanisms among all social actors, or the need to encourage multi-scale experimentation prior to the mass use of AI. We also warn of some of the risks of AI that are not mentioned in the White Paper and which we believe could be amplified by prioritising a vision focused exclusively on commercial competitiveness, which also risks becoming obsolete in the face of the profound social and economic changes that European society is experiencing. The White Paper refers to human rights, but only develops civil and political rights (privacy, political rights and freedoms), ignoring social, economic and cultural rights. Finally, the document summarized here also addresses the need to open a dialogue regarding the public infrastructures that are required to ensure that AI benefits society as a whole.

We therefore call for the dissemination of this document, in which we have attempted to address what we consider to be the most worrying issues, as a focal point for constructive action. We call upon AI researchers and practitioners, together with researchers and professionals from other disciplines who are involved with AI technologies, to add their signature so that many more voices can prevail upon the European Commission to integrate these key issues into the AI strategy defined in the White Paper.

The guidelines in the European Commission's White Paper advocate individual commitment and self-regulation, considering that regulatory intervention could limit potential innovative capacity, which we consider may lead to negative medium- and long-term impacts even on European values themselves. The guidelines also defend the assumption of responsibilities focused on the correct development of algorithms but not on the social effects and impact of their use. Finally, they promote accountability for a number of commitments to technological principles (e.g. transparency) but not to upholding values (e.g. equity). All this entails a limiting and reductionist AI framework. We need to go beyond an AI framework that is restrictive, reductionist and self-regulated, and work towards a future in which not only are citizens protected against the risks of AI but in which AI is empowering and makes a decisive contribution to bringing peace, prosperity, and justice, and to enhancing social, material and spiritual well-being.

The issues involved are explored in a little more depth in the following sections.

List of Signatories

List of signatories as per 14 June 2020. Please note that the document was published on 7 June 2020, and signature collection campaign is still ongoing. Due to the closing of the public consultation, the following list might be incomplete. Please refer to the web site where an updated list is displayed.

List of signatories is available here: http://blogs.uned.es/workshopadvancingtowards/list-of-signatories/.

AI researchers and practitioners, proponents:

Ángeles Manjarrés, Universidad Nacional de Educación a Distancia, Artificial Intelligence, Spain

David Pastor-Escuredo, itdUPM – Innovation and Technology for Development Centre UPM, Artificial Intelligence and Computational Social Science, Spain

Jesús Salgado Criado, Universidad Politécnica Madrid, Business ethics, AI, Spain

Researchers and professionals from related disciplines, proponents:

Celia Fernández Aller, Universidad Politécnica de Madrid, Law and Human Rights, Spain

Simon Pickin, Universidad Complutense de Madrid, Computer Science, Spain

Arturo Fernández de Velasco, Cloud Infrastructure, Spain

Txetxu Ausín, Institute of Philosophy, CSIC, Applied Ethics, Spain

AI researchers and practitioners, signatories:

José L. Aznarte, UNED, Academia/Artificial Intelligence, Spain

Luis Manuel Sarro, UNED, Artificial Intelligence/Physics, Spain

Francisco Javier Díez, UNED, Artificial Intelligence, Spain

Cristina Puente Agueda, Universidad Pontificia Comillas, Inteligencia artificial, España

Ana Garcia-Serrano, UNED, Academia/ Artificial Intelligence, Spain

Gonzalo Génova, Universidad Carlos III de Madrid, Filosofía de la Inteligencia Artificial, España

Jónathan Heras, Universidad de La Rioja, Computer Vision, Spain

Aitor Moreno Fernández de Leceta, Artificial Intelligence Director at Ibermática, Artificial Intelligence, Spain

Ezequiel López-Rubio, Universidad de Málaga, Artificial Intelligence, Spain

Ronan Reilly, Maynooth University, Cognitive Science, Ireland

Dr Rob Miller, University College London, Artificial Intelligence and Knowledge Based Systems, United Kingdom

Carolina Mañoso Hierro, UNED, Control Systems & Robotics, Spain

Víctor Fresno Fernández, UNED, Natural Language Processing, España

Javier Carbo, University Carlos III of Madrid, Symbolic and Distributed Artificial Intelligence, Spain

José Luis Fernández Vindel, UNED, Academia / Artificial Intelligence, Spain

Josué Pagán Ortiz, Universidad Politécnica de Madrid, Artificial Intelligence, Spain

Third Sector, associations

Asociación Pro Derechos Humanos de España, Nuria García Sanz, Derechos Humanos, España

Asociación de Internautas, Ofelia Tejerina, Derechos digitales, España

Instituto de Derechos Humanos San Bartolomé de las Casas, Rafael de Asís, UC3M, Derechos Humanos, España,

Researchers and professionals from related disciplines, signatories:

Juan Manuel Goig Martínez, Catedrático Derecho UNED, Derecho constitucional, España

Ofelia Tejerina, URJC, Derecho TIC, España

Borja Bordel Sanchez, Universidad Politécnica de Madrid, Computing technologies, Spain

Richard Pancost, Universidad de Bristol, Biogeochemistry, UK

Rafael de Asís Roig, Universidad Carlos III de Madrid, Filosofía del Derecho, España

Miguel Pérez Subías, ASOCIACION DE USUARIOS DE INTERNET, SOCIEDAD CIVIL, SPAIN

Hector Silveira, Universidad de Barcelona, Filosofía del Derecho, España

Rafael Miñano Rubio, UPM, Etica y Matemáticas, España

Eloy Portillo, Universidad Politécnica de Madrid , IT Higher Education, Spain

Lucio Marcos Nontol, Seton Hall University, Ethics/Moral, USA

Margarita Robles Carrillo, Universidad de Granada, Network Engineering & Security Group (NESG), España

Mercedes Serrano, Universidad Castilla La Mancha, Derecho constitucional, España

Javier Sanz, University of Coruna, Administrative Law, Spain

Anibal Monasterio Astobiza, LI²FE (Laboratorio de Investigación e Intervención Filosófica y Ética), education, spain

Ricardo Morte Ferrer, LI²FE, Research: Philosophy and Ethics, Germany

José Carlos Lomas Huerta, Contento Curator, Contenidos Web Educación, España

Javier Palomares, IBM, IT Security Consulting, Spain

Manuel Aparicio Payá, Universidad de Murcia, Filosofía, España

Mario Toboso Martín, Spanish National Research Council, CSIC – Institute of Philosophy, Science, Technology, and Society, Spain

Belén Liedo Fernández, Institute of Philosophy – Spanish National Research Council, Philosophy, Spain

FRANCISCO GARRIDO, PROFESSOR MORAL AND POLITICAL PHILOSOPHY JAEN UNIVERSITY, EDUCATION, SPAIN

Marta Peláez Garmendia, Philosopher, Education, Spain

Patrici Calvop, Universitat Jaume I, Educación/moral phylosophy, Spain

Juan Llorens, Ministry of Economic Affairs and Digital Transformation, Engineering, Spain

Juan Garbajosa Sopeña, UPM, Software Engineering, Spain

Pablo Andrés Mazurier, UNIFI – University of Florence's Center for Cybersecurity & IR Studies, Cyberspace, International Politics and Human Rights, Italy

Pedro Garcia Repetto, Ministry of Finance, ICT, Spain

Raúl Cabestrero, UNED, Psychology, SPAIN

Gustavo Palomares Lerma, Professor Jean Monnet Chair – Dean of the Faculty of Political Science and Sociology of the National University of Distance Education UNED, university education – researcher, SPAIN

Josep Antoni Ysern Lagarda, Universidad Nacional de Educación a Distancia, Filología Catalana, Spain

Javier Callejo, UNED, Sociology, Spain

Humberto Bustince, Universidad Publica de Navarra, Academic, Spain

Ricardo Aguilar de la Torre, Junta de Andalucía, Tax Income, Spain

Jose Manuel Arizaga Álvarez, Retired, Economics, Spain

Víctor Riesgo Gómez, Investigador, Predoctoral UNED, Sociología, España

Jesús M. Pérez, University of the Basque Country (UPV/EHU), Education, Basque Country

María García Alonso, National Distance Education University (UNED), Social Anthropology, Spain

Marcos Sánchez-Élez Martín, Universidad Complutense de Madrid, Hardware Design, Spain

Francisco Cantero, Activist, New Technologies, Spain

Ian Hodkinson, Imperial College London (emeritus), Computing, UK

Helena Matute, Universidad de Deusto, Experimental & Cognitive Psychology, Spain

Anselmo Lucio Saiz, Universidad de Málaga, Comunicación Audiovisual y Publicidad, Spain

Juan Manuel Moreno Manso, Teacher, Philosophy (of Mind), Spain

1. THE APPROACH OF THE WHITE PAPER IS TECHNOLOGICALLY REDUCTIONIST, IN CONTRADICTION WITH THE EUROPEAN COMMITMENT TO THE UN AGENDA 2030, WHICH IS NOT GIVEN ITS DUE CENTRALITY AND, INDEED, IS HARDLY MENTIONED

● Public policies on Artificial Intelligence (henceforth AI) seem to be exclusively designed to improve the competitiveness of European companies in AI when the focus should be more on how AI helps to improve people's living conditions, contributes to the Sustainable Development Goals (henceforth SDGs) and to the fight against climate change and its consequences. We believe that the SDGs should be at the core of Europe's competitiveness strategy, especially in the digital and AI sector. There is a growing interest in the role that AI can play in achieving the SDGs on the part of international organizations, such as UN Global Pulse, the UNICEF Global Innovation Centre, the World Wide Web Foundation, and even the World Economic Forum. In addition, there are multiple initiatives promoting public-private partnerships concerning AI and the SDGs. We believe that in the medium to long term the commercial success of digital platforms will depend on their added social value that will, to a large extent, be determined by their alignment with the SDGs. Furthermore, AI can be a vehicle for international collaboration within Europe. A competitive Europe needs interconnected digital infrastructures and fabrics beyond those that can be offered by disconnected companies.

In terms of specific content on the SDGs, the White Paper makes only a timid mention of environmental sustainability, which does not appear as an important issue in any of the questions of the questionnaire. The potential and compelling purpose of technology to advance issues of rights, sustainability of communities, common welfare and social and gender equality is omitted. In addition, the current crisis has highlighted the need to use AI to improve the health response at national and European level, monitor progress of the infection and mitigate our vulnerability.

● The text does not mention (in section 4b Focusing the efforts of the research and innovation community) the promotion of lines of research in support of the SDGs, including the most pressing ones for Europe (fair and equitable data, sustainable economic development, risk prevention and disaster management). The United Nations has recognized that the SDGs cannot be achieved without a people-centred, science-based digital revolution. In particular, Europe is committed to addressing the refugee crisis and eradicating poverty through humanitarian assistance and civil protection actions, as well as through international development cooperation (see, e.g. – European Commission Fact Sheet Next steps for a sustainable European future – European action for sustainability: Questions & Answers, 10 Commission priorities for 2015-19, num. 9 A stronger global actor). This being so, the application of Artificial Intelligence to humanitarian response systems, at a global level and focused on the most affected groups, should be a priority at a structural level and in all countries, and should be supported by common cooperation systems.

Security also means, and we are now seeing this clearly, public health, access to essential medicines and treatments, safe food and water, social security (in the event of illness, accident, disability, retirement…), labour and consumer protection, access to housing, disaster prevention and environmental protection.

● AI research and development in the field of SDGs is biased towards issues relevant to the nations where most researchers and practitioners live. There are very few AI technologies

applied in countries where there is not a strong AI research base: it might be a good idea for the EU to sponsor innovation and production of AI technologies in these countries. This would reverse the current trend where AI is increasing inequalities not only within countries but between countries. The White Paper does not talk about promoting research on how to implement these technologies adapted to different cultural values and levels of wealth. AI applications reflect the needs and values of the nations and social classes that developed them. In a globalized world the lack of human rights in any part of the planet affects us all (pandemics like COVID-19, refugee crisis…).

● It would advance the commitment to the Agenda 2030 if AI were developed with the philosophy of RRI (Responsible Research and Innovation), which aims for a new model of research governance that reduces the gap between the scientific community and society, encouraging different stakeholders (educational community, scientific, business, industry, civil society organizations, politics) to work together throughout the process of research and innovation. The idea is that cooperation between different (relatively autonomous but highly interdependent) actors will make it possible to align the research process and its results with society's values, needs and expectations. Although the term RRI was coined a decade ago, it has recently gained prominence due to its inclusion in the Science with and for Society Programme promoted by the European Commission within the framework of the Horizon 2020 research strategy.

The RRI is a radical move toward openness and socialization of techno-scientific processes that is embodied in four principles of governance: anticipation, reflexivity, deliberation and accountability. This openness and socialization is embodied in the idea of public engagement that is essential for researchers in terms of more ethical solutions to difficult questions that facilitate acceptability and trust in research.

The RRI agenda comprises 6 policies to be taken into account throughout the research and innovation process:

a) Citizen participation, to encourage multiple stakeholders to be involved in the research process from conception through development and to delivery of results.

b) Gender equality, to promote gender balance in teams.

c) Scientific education to improve educational processes and promote scientific vocations among the youngest.

d) Ethics to promote scientific integrity, in order to prevent and avoid unacceptable research practices

e) Open access to scientific information, in order to improve collaboration between interest groups and open dialogue with society

f) Governance arrangements, to provide tools to promote shared responsibility among stakeholders and institutions.

● Another aspect that is missing concerns the social and ethical impact of AI. As AI permeates into more and more areas of society and it should increasingly be able to measure its own impact: AI should be self-assessing. Clear metrics are needed in this regard to avoid bias and undesired indirect impacts.

Social and ethical impact assessment is difficult to carry out if there are no common tools (ethical and technological) and dedicated research because individual companies do not have the capability to perform it. Furthermore, fragmented development of AI impact assessment by different companies can increase biases and, at the same time, make them uncompetitive against the platforms of the digital giants.

● Finally, to avoid introducing bias, AI needs unbiased data. The creation of an ecosystem of useful data representative of all population groups for use in public policy should be a priority. People-centred data is the first step for a people-centred AI.

2. THE APPROACH IS ALSO ETHNOCENTRIC AND ANDROCENTRIC

● Although the need to avoid ethnic and gender discrimination and biases in Data Science projects is mentioned (guaranteeing the representation of all groups in the data), there is no mention of the need to design applications adapted to different cultural sensitivities and values (culture-aware systems) or the need to enhance the representation of all groups in the research, design and development of AI applications (through mechanisms such as community-based development methodologies, citizen science, etc). We emphasize the multi-cultural strength of Europe and the opportunity that this represents for AI, as well as the synergies that can exist to build a more cohesive Europe if an AI sensitive to people and to these cultures is developed. This is only possible with the strengthening of pan-european research involving both the public and private sector.

● The White Paper does speak of the need to promote women's vocations in AI technologies, but not to encourage female entrepreneurship in AI. In the computer industry only 1-2% of new companies receiving venture capital funds are run by women, even though women-led companies get a 200% return on investment. This is the area where gender biases in the technology industry are most apparent. This is a missed opportunity to develop a more inclusive and innovative AI.

3. UNIVERSALIZING THE BENEFITS OF AI

● There is no mention of any position on the «open science», in the sense of the upcoming UNESCO Open Science Recommendation , or «open AI» paradigm (in particular, open source and open algorithms to facilitate the dissemination of the benefits of AI). A paradigm of knowledge sharing, replicability and openness is needed. It is important, as we have mentioned, to have AI assessment frameworks that can include auditing to ensure that systems perform their function properly. In addition, there is an increasing amount of research on the impact of IA on society, which is a topic that Europe needs to address in a multi-disciplinary way as a matter of priority. Technology is not neutral and the adoption of AI is not within the reach of all, so it will not diminish inequalities on its own unless specific policy lines are developed with these objectives in mind. Recent studies show this to be the case. There is a risk of AI creating new inequalities as well as greater difficulties in the process of improving opportunities for vulnerable groups, thus creating an additional barrier between those who can benefit from AI and those who cannot.

● We may need to ask how AI could help disadvantaged sectors of the population (or even harm them if no measures are put in place to avoid this outcome). The risk is of creating a new AI gap, which would overlap with the already worrying digital divide that exists globally today.

● It would be highly desirable to mention the promotion of a social economy in the AI sector.

4. ON THE EDUCATION OF AI PROFESSIONALS AND THE DIGITAL LITERACY OF THE POPULATION

● No emphasis is placed on the urgent need for multidisciplinary education and on raising the ethical and social awareness of AI professionals and researchers, as well as on the education of society as a whole in aspects of digitization and AI.

As recognized in the report of The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems: «The key is to embed ethics into engineering in a way that does not make ethics a servant, but instead a partner in the process. In addition to an ethics-in-practice approach, providing students and engineers with the tools necessary to build a similar orientation into their inventions further entrenches ethical design practices […] it is recommended that engineers and practitioners come together with social scientists and philosophers to develop case studies, interactive virtual reality gaming, and additional course interventions that are relevant to students". Education for ethically sensitive AI research is essential.

● There is no mention in point 4.C on Skills of the responsibility of educational institutions (in particular Universities) to provide curricula with the necessary multidisciplinarity, including a strong ethics-education component. Future AI professionals will need to have been educated in the ethical, philosophical, social, political, psychological, and risk analysis aspects of the development and implementation of AI technologies.

● There is no mention of the need to promote/subsidize pedagogical research in multidisciplinary curricula in view of all the above. Ethics is already a transversal competence in degrees and a specific competence in subjects such as the Final Project, but this competence is not given due attention or is even ignored. Further research is needed into pedagogical approaches that are particularly suitable for AI ethics education purposes (project-based or experiential, such as service-learning).

● Experts from different areas will have to use and live with AI in their daily work. In a future in which human-AI interaction becomes commonplace professionals from any field will need to deal with AI concepts and understand the causes and consequences of the use of AI systems. This implies a level of technical and also ethical knowledge.

5. ON THE NEED FOR DIALOGUE BETWEEN ALL SOCIAL ACTORS, NOT ONLY INDUSTRY WITH THE PUBLIC SECTOR

● The White Paper does not refer to the dialogue of the public sector with other social actors (third sector, civil society, academia), beyond companies and SMEs. In this sense, it is highly desirable that the discussions on this subject at EU level be transparent, so that the research community, civil society, the social sector and any other actor involved will be able to keep abreast of the state of the discussions concerning the regulation of artificial intelligence. The importance of this lies in the fact that the social value of AI has to be determined collectively and in the fact that enriching the AI ecosystem will need the perspective of many different actors.

● The role of society is not mentioned in the White Paper. Firstly, there is a need to educate society in the use of AI and to raise awareness of the ethical risks and socio-political implications it entails. Secondly, society must take an active role in defining the uses of the data it generates, especially regarding the knowledge extracted from the data and the AI application built using it. The role of AI in society has yet to be defined, and society must be part of the definition process since it will affect all levels of organization, behavior and production.

In this sense, there are some tools that may be useful to support social participation. For example, citizen science and collective intelligence. Citizen science seeks to promote public participation in science, and is beginning to constitute a distinct paradigm of doing science. Collective Intelligence processes are governance mechanisms that could help to ensure that society capitalizes on the potential of AI technologies while at the same time ensuring that they are developed and implemented with due respect for human rights. The wider the participation in these processes, the greater the confidence in, and the social appropriation of, the technology, highlighting the importance of AI literacy.

As we have mentioned, the quantification of benefits and positive impacts is key to motivating citizens and for a scientific activity that allows a responsible evolution of AI. Any analysis of the benefits of AI will need to be continually updated since AI is changing society very rapidly. The same is true for the negative effects and risks, some of which are already being felt by general population.

6. ON THE NEED FOR ETHICAL REFLECTION AND FOR METHODOLOGIES AND TECHNIQUES TO SUPPORT APPLIED ETHICS

Although manuals of ethics and good business practices are necessary, there is also a need in the academic sphere for independent and scientifically rigorous research with an empirical dimension since, to date, this dimension has been mostly lacking. Among the lines of research and development that should be encouraged and subsidized, priority could be given to the following:

● Ethical-philosophical reflection at a theoretical level. There is a need to bring order to the current landscape of overabundance of ethical codes, guidelines and frameworks, many of which suffer from deficiencies such as lack of scientific rigor, subjectiveness, incoherence, superficiality and redundancy, thus generating confusion. Moreover, however many current codes, guidelines and frameworks are produced, the debate is far from closed. In addition, at every step new dilemmas arise from the development and deployment of new applications and from gathering more experience with existing ones, to which is added the permanent evolution of society driven in part by the very use of these technologies.

● The concept of applied ethics tools. There has been much debate and research into identifying risks but very little on how to mitigate them. Ethical principles codes of conduct and legislation are necessary, but to apply these in a practical way requires tools. The solution to mitigating AI risks can also come from tools that incorporate AI techniques. A coordinated multidisciplinary effort involving researchers, innovators, citizens, legislators, politicians, developers… is needed to create and evaluate these tools. Again, multidisciplinarity is essential in order to give full meaning and from different perspectives to the concepts of explicability, transparency, etc., in order to understand the complexity of human behavior and the impacts that AI technologies can have on it, in order to interpret the algorithmic predictions… In addition, the plurality of values not only of the professionals and producers but also that of the society in general needs to be protected. Many international initiatives conclude that ethics should be embedded in the process of designing, developing, deploying and using intelligent technology, and that ethical principles need to be translated into protocols for design, development, deployment and use. Specific methodological and technical tools to support the development and use of AI applications that meet ethical standards and comply with legislation are required for each development phase. These tools are not meant to replace legislation nor ethics and good practice manuals, but to support their implementation. Academic research, private sector self-regulation and legislation are necessary and complementary actions.

● An approach of growing importance is to consider the ethical behaviour imbued in artificial intelligence systems as a form of control, the idea being to imbue the autonomous entity with values from its conception. As AI achieves more agency, the question of responsibility (a very important legal issue) arises.

● We stress the importance of avoiding a purely economic, technocratic approach. AI impact studies should be wide-ranging, which requires multidisciplinary teams (ethics scholars, technical experts, sociologists, psychologists, philosophers…) and should take into account the dynamicity of AI impacts.

Wide-ranging impact-assessment frameworks necessarily involve gathering, sharing and processing large amounts of multidimensional data and may therefore require large-scale investments in infrastructure. A data-rich ecosystem would allow the SDGs to be used as the impact-assessment framework, by means of indicators that cover multiple perspectives, including social costs and benefits (in terms of schooling, life expectancy, access to basic services and environment, among others), and taking into consideration the cultural values of the communities concerned. Measurement of the corresponding impact indicators should not be too locally-implemented or company-specific firstly, since a significant part of the overall impact may take the form of indirect, and possibly unexpected, effects occurring outside the area of measurement and secondly, to facilitate comparisons. Thus, common, large-scale impact-measurement frameworks would be more appropriate, in which context AI ethical considerations acquire a social and political dimension, i.e. macro-ethics and not just micro-ethics.

7. ON THE NEED TO ENCOURAGE MULTI-SCALE EXPERIMENTATION (WITH THE APPLICATION OF APPLIED ETHICS TOOLS AND EMPHASIS ON IMPACT STUDIES) BEFORE PROMOTING THE MASSIVE USE OF THE PRIVATE SECTOR, INCLUDING SMEs, AND THE PUBLIC SECTOR

● Support tools for the application of ethical principles developed in research laboratories are currently at an early state of research so that their usefulness and impact are yet to be demonstrated. They will have to be scaled up in order for them to be used outside of the research context, pointing to the need to incentivize knowledge transfer in this area. Examples of quantified good practice in the application of applied-ethics tools in real-world projects are also needed. The promotion of R&D projects for the experimentation of AI with an ethical approach involving industry, the public sector and research centres, would help to create such good practices.

● Applied ethics tools also exist in the private sector and are used in internal consultancy. In general these tools are in a very immature state, and are not ready to be adopted by professionals in the short term, although in some very specific fields, effective tools may already have been developed albeit with limited scope. Moreover, for the time being, adopting these tools does not provide any competitive advantage, which is why it is important to encourage them through proactive policies that promote their incorporation into business models. For impartiality, AI ethics assessment should be outsourced in the same way that many assessment, certification and quality control processes are outsourced. For the assessment process to be homogeneous requires standards at national and European level to guide and regulate the assessment tasks.

● Theoretical and empirical research on ethical AI lags far behind purely technological AI research, for which reason the former should be given strategic priority over the latter. It is imperative to accelerate the development of AI ethical assessment frameworks to guide the future of AI in Europe and to guarantee that assessment processes can be carried out in real time and at scale, with a view to mitigating possible negative impacts quickly and to performing learning at scale of the way AI impacts society. A commitment from the private sector is required to prioritize the assessment of algorithms and platforms.

8. IMPORTANT RISKS NOT MENTIONED IN WHITE PAPER

● In all the forums where the ethical risks of AI are discussed, a set of common themes are mentioned: beneficence and non-maleficence, autonomy, justice and explicability. To our understanding the White Paper does not cover all the dimensions of the principles of Autonomy and Justice.

● Although the optimistic view of technology also predicts a generally positive contribution of AI to the ability to live a happy and fulfilling life, there are also important risks in this area for individuals, their values and their personal development. It is important to emphasize the warning of possible psychological damage, of how intelligent technology could negatively influence mental health and the development of personal potential, personal fulfillment, human fulfillment, and living a life with meaning and purpose.

No mention is made of important problems associated with the use of this technology such as the alteration of the concept of identity and the nature of human interactions, the difficulty of distinguishing between the real and the virtual, escapism towards virtual worlds, the replacement and deterioration of human bonds, cognitive overload, the loss of meaning and purpose due to being replaced by intelligent machines, etc. Nor is there any discussion of the potential loss of human values: wisdom, creativity, empathy, affection, social skills… It does not warn about the possibilities of control, manipulation, attack on autonomy, etc. that the field of affective computing opens up, given the susceptibility of humans to emotional influence. Nor does it warn of the important effects on mental and physical health that interactive immersive virtual reality applications (with intensive application of Artificial Intelligence techniques) could have. Experimental work with social science experts should be promoted with a view to assessing these risks.

● To ensure that AI really contributes to «human welfare» requires a prudent multidisciplinary and eco-centric approach (in harmony with an authentic anthropocentrism that recognizes the essence of the human being and his or her interests) to AI research, as opposed to an excessively techno-optimistic and technocratic approach. Given its disruptive power, we cannot assume out of hand that AI facilitates more efficient exploitation of resources; to simply assume that the market will regulate good uses of AI is to abdicate responsability.

● While implementing intelligent technologies with a purely economic or technocratic perspective could contribute to economic growth, as a counterpart, it could have an environmental and inequality cost. Some studies estimate that of the sub-targets into which the SDGs are decomposed, AI could contribute positively to 134 (79%) and act as an inhibitor to 59 (21%). The potential of AI to increase productivity could, in turn, increase the overexploitation of resources if economic, social and environmental variables are not integrated, which is not always the case in private sector studies. In addition, while AI can increase efficiency in energy production, advanced AI technology requires massive computing resources only available in large computer centers that require a lot of energy and have a high ecological and carbon footprint (this aspect is addressed in the White Paper).

● The recently published UN report by the Special Rapporteur on extreme poverty and human rights warns of the «risk of tripping over like zombies in a digital welfare dystopia» where «Big Tech has been a driver of growing inequality and has facilitated the creation of a vast digital underclass". The report provides many well-documented examples in different countries of

how dehumanized smart technologies are creating barriers to access to a range of social rights for those without Internet access and digital skills.

● One of the most important ethical requirements of AI is explainability and transparency, since algorithmic decisions can affect the most sensitive areas in people's lives (health, civil and social rights, criminal law, credit). With machine learning, and particularly deep learning applications, however, explaining the decision process is very difficult. As an example of the effect of such lack of transparency, we cite a real life case in which a claimant was informed that he or she did not qualify for a government subsidy. When an explanation was demanded, the only one given was that an algorithm had made the decision. A request to see the algorithm was refused on intellectual property grounds since the decision-making task had been subcontracted to a private company.

● Other relevant risks are commercial and political manipulation, and intensive coercion and surveillance by governments and large corporations, which can damage social cohesion and contravene democratic principles and human rights.

● Privacy is another well-known risk, and though it is treated in the White Paper, in our view, important aspects are not addressed. Privacy regulations must protect people but should also offer solutions to the social and public use of data, since the «non-use» of data in circumstances in which there is a clear public interest in its use is a social disadvantage. So far privacy has mainly served to protect the interests of private companies while what is needed is a move towards models of responsible data sharing. There is a need for a clear public policy on the use of data oriented towards the common good, especially data generated by public institutions (for example open access to results of publicly-funded research).

● The list of high-risk areas mentioned in the White Paper (health, transport, energy and parts of the public sector such as asylum, migration, border control and justice, social security and employment services), is not very complete. Applications that may seem harmless a priori (in marketing or financial or insurance services, or in social assistance provided by NGOs, etc.), may bring about threats to rights if they produce discriminatory, biased results, etc. It seems risky to leave this list open to future revisions and amendments depending on relevant developments in practice, instead of carrying out a deeper analysis now.

● In general AI requires a proactive approach to risk management, involving continual risk identification and handling.

9. CONFLICTS BETWEEN IA AND FUNDAMENTAL RIGHTS

● The White Paper refers to the defense of human rights, but only develops the arguments with regard to civil and political rights (privacy, political rights and freedoms), ignoring social, economic and cultural rights. This is closely related to concerns about the inequality that the massive use of AI could generate.

● As stated in the Zaragoza Declaration (2019): «Sometimes, when it comes to setting responsibilities, it is easy to mistake values for means. That is to say, to establish commitments, but not about what is necessary to protect, nor about what is socially in conflict, rather about the technological principles an artifact is designed with (for example transparency). By focusing on ensuring transparency to guarantee privacy it is understood that it is the user who should exercise the role of supervisor. Is this the best way to legislate such a complex sector? If we look at the food industry, for example, the consumer is not expected to inspect and investigate everything he or she eats. Control is delegated to regulators and inspectors, while the consumer is simply expected to know his or her rights, and how to access appropriate channels in the case of any problems.

The fact of being positioned in one-way or another represents antagonistic conceptions regarding the design of a supervisory system: either user centred or transparency governance. There are values that will never appear in the first approach because they are collective claims. Once we stop viewing the end user as the only person affected by the use of technology, we can begin to consider the effects on society as a whole. This makes it possible for us to consider plurality, cohesion, sustainability and cooperation in technological development. This will only come from public debate and shared reflection.

It is, more than ever, necessary to create environments for public debate, where dialogues between researchers, developers and other members of society can be held. The promotion of spaces for continuous communication about the social and ethical implications of Artificial Intelligence systems is vital».

Assuming the need for transparent governance of AI, we believe that it would be highly advisable to include research on verifying the fairness of algorithms in the areas of research to be prioritized, with a view to integrating such verification mechanisms in a certification process to be carried out by an independent authority.

Quintanilla coined the term «endearing technologies», as opposed to “alienating technologies”, this idea being consistent with the Responsible Research paradigm. We believe that such technologies are more respectful of fundamental rights. Some of their characteristics, as defined by Quintanilla, are: openness (the software is free); versatility (it allows alternative uses); docility (its operation and control depends on a human operator); no planned obsolescence (repair is promoted more than replacement); comprehensibility (basic but comprehensive documentation).

10. PUBLIC DATA INFRASTRUCTURES

The deployment of AI requires the computation of large volumes of data that is currently done in data processing centers. Administrations need to have a public custody system for data processing.

Institutions and social organizations should also have access to data for public use. Open data implies interoperability and interconnection of diverse data sources. It also requires designing services with adequate security and privacy and that optimize the usage and accessibility of data.

In addition, the question of data ownership should be addressed. It should not be assumed that citizen-generated data belongs to the companies that capture it because this fact has not been actively accepted by society. This question acquires particular relevance when the data is to be used for a public purpose. There needs to be a move towards data sharing models that facilitate the use of data for public purposes, especially in the case of humanitarian emergencies.

Conclusions

The guidelines of the European Commission's White Paper advocate individual commitment and self-regulation considering that regulatory intervention considering that regulatory intervention could limit potential innovative capacity; in addition, they advocate the assumption of responsibilities based on the design of the algorithms but not on the social effects and impact of their use; finally, they are in favour of accountability with respect to a series of commitments on technological principles (e.g., transparency) but not on the preservation of values (e.g., equity). All this assumes an AI conceptual framework that is reductionist. The danger is that, by using notions of regulation, self-regulation and ethics in an imprecise and interchangeable way, we are closing the way to the possibility of true spaces in which to define a common standpoint. AI is a technology with a strong transformative potential; it is an opportunity for greater European cohesion if it is view from the perspective of promoting the common good of the peoples of Europe. Moreover, the potential of AI is greater the larger the scale of application.

The White Paper reduces AI to an instrument of a commercial nature. Responsibility is restricted to the protection of the individual rights of the client or consumer. It has been demonstrated that the value of data and of AI transcends the concept of technological service and therefore requires mechanisms to ensure proper management of this value.

To use an analogy given in the Zaragoza declaration, AI has the dimension of a social technology, so the control that is needed over its development and deployment is not simply the quality control of individual products like, say, the quality control of a car. Since its impact is not limited to individual users, the required control is that exercised over infrastructure developments like, say, the planning of the construction of a new bridge. "The design criteria (territorial cohesion, care for environment) are not centred on guaranteeing a specific right to a few individuals but to the mobility within a community." The impact of AI will change society, for which reason AI policy and regulation demands collective debate. Moreover, this debate must be multidisciplinary since, on their own, technologists cannot find the solution to many of the problems that arise with the use of AI. We will need to integrate critical and constructive action to design a society empowered by AI and that satisfies our idea of a society in which we would wish to live.

List of signatories of this proposal is available here: http://blogs.uned.es/workshopadvancingtowards/list-of-signatories/.

References

Alston, P. Report of the Special rapporteur on extreme poverty and human rights. UN. https://digitallibrary.un.org/record/1629536?ln=en . (2019). Accessed: May 15, 2020.

Australian Human Rights Commission document, Human RIghts and Technology. Discussion paper, 2019.

BBVA (2020) El trabajo en la era de los datos

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

Bostrom, N. Strategic implications of openness in AI development. Global Policy, 8(2), 135-148. (2017).

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., … & Anderson, H. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.

Data Revolution Report. 'A WORLD THAT COUNTS: Mobilising The Data Revolution for Sustainable Development. ' Presented to Secretary-General. https://www.undatarevolution.org/report/

EU-Citizen.Science. The platform for sharing knowledge, tools, training, and resources for Citizen Science. https://eu-citizen.science/. Accessed: May 15, 2020.

European Commision, Ethics guidelines for trustworthy AI. (2019). https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419. Accessed: May 15, 2020.

Floridi, L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philosophy & Technology, 1-9. (2019).

IEEE, Ethically aligned design. Tech. rep., IEEE. https:// ethicsinaction.ieee.org/ (2019). Accessed: May 15, 2020.

International Telecommunications Union (2020) AI for good global summit 2020. https://aiforgood.itu.int/. Accessed: May 15, 2020.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.

Luengo-Oroz, M. (2019). Solidarity should be a core ethical principle of AI. Nature Machine Intelligence, 1(11), 494-494.

Morley, J., Floridi, L., Kinsey, L. & Elhalal, A. From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics, 1-28. (2019).

Mueller, M.P., Tippins, D. and Bryan, L.A. The Future of Citizen Science. Democracy and Education, 20 (1), Article 2. (2012).

Ochigame, R. The invention of "Ethical AI", The Intercept. December 20 2019. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ Accessed: May 15, 2020.

Organisation for Economic Co-operation and Development (OECD), May 2019. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Accessed: May 15, 2020.

Quintanilla Fisac, M.Á.: Engaging technologies: criteria for an alternative model of technological development. In: Laspra, B., López Cerezo, J.A. (eds.) Spanish Philosophy of Technology: Contemporary Work from the Spanish Speaking Community, pp. 103–123. Springer, Cham (2018)

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., … & Jennings, N. R. (2019). Machine behaviour. Nature, 568(7753), 477-486.

Ryan, M. , Antoniou, J., Jiya, T., Macnish, K., Brooks, L., Stahl, B.C. 'Technofixing the Future: Ethical Side Effects of Using AI and Big Data to meet the SDGs'. IEEE SmartWorld. (2019).

Stilgoe, J., Owen, R., Macnaghten, P. 2013. "Developing a framework for responsible innovation". Research Policy 42(9), 1568-1580.

UN Global Pulse (2012) Big data for development: opportunities and challenges – White paper. https://www.unglobalpulse.org/document/big-data-for-development-opportunities-and-challenges-white-paper/ Accessed: May 15, 2020.

UN, AI for Good Global Summit. Switzerland: UN. (2020). https://aiforgood.itu.int/. Accessed: May 15, 2020.

UNICEF Global Innovation Centre, Generation AI. (2018). https://www.unicef.org/innovation/stories/generation-ai Accessed: May 15, 2020.

Université de Montréal, Montréal Declaration for Responsible Development of Artificial Intelligence. Montréal: Université de Montréal et Fonds de recherche du Québec. (2018). https://www.montrealdeclaration-responsibleai.com/ Accessed: May 15, 2020.

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch,S.. & Nerini, F.F. The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1), 1-10. (2020).

Von Schomberg, R. 2013. "A Vision of Responsible Research and Innovation?, in R. Owen, J. Bessant and M. Heintz, Eds., Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society, Chichester, UK: Wiley, pp. 51-74.

World Economic Forum, Harnessing artificial intelli- gence for the earth. (2018) http://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf. Accessed: May 15, 2020.

World Wide Web Foundation (2017) Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries. Tech. rep., World Wide Web Foundation. https://webfoundation.org/docs/2017/07/AI_Report_WF.pdf Accessed: May 15, 2020.

Zaragoza Declaration: For a Deontology in the Design and Interaction with Intelligent Systems (2019). https://www.fundacionzcc.org/es/noticias/publicamos-declaracion-zaragoza-deontologia-diseno-interaccion-sistemas-inteligentes-2496.html

List of signatories of this proposal is available here: http://blogs.uned.es/workshopadvancingtowards/list-of-signatories/.