المجلس الوطني لحقوق الإنسان

ⵜⴰⵎⵙⵎⵓⵏⵜ ⵜⴰⵏⴰⵎⵓⵔⵜ ⵉ ⵉⵣⵔⴼⴰⵏ ⵏ ⵓⴼⴳⴰⵏ

National Human Rights Council

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

Organization, ethics and guiding principles an international benchmark

Proceedings of the International Seminar
Organized December 3rd, 2021

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

Organization, ethics and guiding principles an international benchmark

Proceedings of the International Seminar Organized December 3rd, 2021

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS
Organization, ethics and guiding principles - an international benchmark

TABLE OF CONTENTS

Conceptual note

The National Human Rights Council (CNDH) considers human rights relevant to the field of Artificial Intelligence within an international context characterized by a holistic reflection on the matter. Numerous initiatives from international, regional, and national bodies are currently developing. Approaching this topic from a systemic perspective, the establishment of a definition of Artificial Intelligence is required. While it may prove challenging to find a comprehensive and conventional definition, given the multiple angles of approach, we have adopted the following definition:

Artificial Intelligence is both a scientific field (integrating multiple scientific ranges: mathematics, informatics, neurology, psychology, engineering, sociology…) that aims to create a technological equivalent to human intelligence, on the one hand; and autonomous intelligent systems with algorithms capable of performing actions that have so far been created exclusively by humans, or that help or make decisions or self-learn through the data at their disposal, on the other.

In today's world where digitization is a lever for societies' growth and evolution, Artificial Intelligence is used in a wide array of fields, such as: in the field of mobility and image processing (facial recognition, automated archiving, localization, cryptography, etc.); in education; in data processing and decision-making assistance; in maintenance; in data transfers and documentation; in banking and accounting; in health and medicine; in planning; in the field of mapping; building simulations; information and communication.

Artificial Intelligence is thus amongst the mechanisms that facilitate the enjoyment of fundamental rights and freedoms by citizens. However, the use of Artificial Intelligence is not devoid of risks to certain rights and freedoms, namely the right to physical integrity and integrity of data, the right to freedom of opinion and expression, the right to access information, the right to privacy, consumer rights, equality and non-discrimination, protection of vulnerable groups (e.g., children, persons with disabilities), the right to physical and psychological integrity, freedom

of elections, the right to employment, freedom of assembly, freedom of peaceful demonstration, ...

The Council shares the conviction of the United Nations High Commissioner for Human Rights that "Artificial Intelligence can be a force for good, helping societies overcome some of the great challenges of our times. But AI technologies can have negative, even catastrophic, effects if used without sufficient regard to how they affect people's human rights... This is why there needs to be systematic assessment and monitoring of the effects of AI systems to identify and mitigate human rights risks."

Considering the enormous opportunities that Artificial Intelligence provides to facilitate access to rights and freedoms, on one side, and the risks that its use poses to certain rights and freedoms, on the other, the Council, through its human rights-based approach, seeks to propose ways to achieve the following objectives:

- Development of Artificial Intelligence in line with a constructive approach to human rights and the values of a democratic society;

- Study and adequately address the effects of artificial intelligence on human rights;

- Artificial Intelligence actors to assume responsibility for its use;

- Citizens to enjoy the benefits of technology associated with artificial intelligence in respect of human rights.

After conducting broad consultations with all national stakeholders, the Council organized an international seminar in Rabat on December 3rd, 2021, to discuss international initiatives in the organization of artificial intelligence with regard to human rights, the various standards, guidelines, and regulations, and governing principles in the field.

- I -
INTRODUCTION TO AI
& HUMAN RIGHTS

The use of AI already affects human rights

Amina Bouayach
Chairperson of the National Human Rights Council – CNDH

In the name of the National Human Rights Council, I would like to express my deepest gratitude to the esteemed participants, national and international field experts, eminent personalities, and friends joining us at the seminar to discuss a topic that haunts the respect of human rights. A topic that is increasingly taking space in our strategic actions as a new and emerging issue related to human rights.

As everyone may know, Artificial Intelligence refers to programmed systems designed to function and act like humans that evolve and develop every day and that we expect will end up being able to imitate and execute the same tasks humans do. The impact of AI is already being felt on human rights issues in the way we receive information nowadays, which influences our choices and how our societies function. The use of AI already affects human rights, and a wide array of sectors and fields raise concerns about privacy, education, employment, healthcare, social care, health, mobility, law enforcement, and even the maintenance of law and order.

In this context, the Council decided to help guide the development of Artificial Intelligence in our country as an institution concerned with the prevention of human rights violations, protection of victims, and promotion of human rights values and principles. Our concern about the negative impacts of AI use on human rights is based on a number of observations and preoccupations. The first concern regards how to monitor the consecration of the exercise of freedom of expression without violating the privacy of individuals in the digital space, more specifically, on social media platforms previously discussed in our 2019 Annual Report [1]. The second concern pertains to the protection of human rights in the field of technology and AI, namely upholding the right to privacy, protection of personal data, and security in the design of applications and algorithms related to AI, also referred to as human rights by design, discussed in our 2020 Annual Report [2]. The Council has organized

consultations at the national level by setting up a digital platform between April and December 2020 called digital expressions "taabirat raqmya" that received contributions of experts and Moroccan citizens concerning the issue of the exercise of freedoms in the digital space. We also organized interactions in December 2019 and January 2020 with journalists, editors, unions, associations, and human rights defenders on the protection of all human rights in the digital space. These first consultations alerted us to the need to set in place consultations of a bigger scale which were organized in April 2021 with technology companies, scholars, professional associations, and think tanks on the protection of human rights in the environment of AI. The CNDH also published a special issue concerning AI in its Arribat journal, with great contributions [3] [4] [5] [6] [7] [8].

Today, we extend these consultations to international and national experts that will shed light, through their evaluations, on the impact of AI on human rights. The threat of the generalized use of AI by states and companies, including profiling, automated decision-making, and machine learning technologies, affects the enjoyment of some fundamental human rights. AI can also cause employment loss and often replaces the role of humans to make processes more efficient by enabling machines to undertake manual tasks once carried out by humans. In 2019, The Economist estimated that 47% of jobs are at high risk of becoming automated and can be carried out by machines in the 2030s. The main risk is how information and data can lead to discriminatory practices that exclude individuals and violate fundamental human rights norms and laws. The challenge that we face is how to keep up-to-date international and national legal instruments to protect and promote AI use that respects and protects the human rights of individuals and society.

Over the past weeks, we have noted with great interest the debate occurring at the level of UN bodies, and their concerns transmitted into recommendations that call on us as an institution of human rights. In September 2021, the United Nations High Commissioner for Human Rights called for imposing a moratorium on certain AI systems, [9] such as facial recognition, and warned against the negative effects that AI technologies can have if their impact on human rights is not taken into consideration. The United Nations Human Rights Council's resolutions 4215 and

4804 urge state members to respect and protect the right to privacy and to take measures to end this right's violations, and create conditions allowing the prevention of these violations, to regularly review their procedures, practices, and legislations in terms of communication monitoring including large scale monitoring and the interception and collection of personal data. This same resolution of the Human Rights Council encourages companies to fulfill their obligation to respect human rights and ensure that respect for relevant human rights is taken into consideration in the design, use, evaluation, and regulation of IT programs.

In November 2021, UNESCO's 193 member states adopted a recommendation on the Ethics of Artificial Intelligence, which can be considered the first international norm on the ethics of AI, outlining common values and principles that will guide the establishment of the necessary legal infrastructure to ensure the healthy development of AI [10]. This recommendation reiterates its call for technology companies and governments to take action to ensure better protection of individuals and thus foresee actions to improve data protection and individuals' knowledge about their own data and their right to control it.

Hence, AI raises urgent concerns for us human rights defenders regarding the possibilities granted to algorithms and AI to make decisions that are decisive for individuals' lives, groups, and even society as a whole. The reflections and observations of each of the seminar speakers and participants will undoubtedly constitute a framework of action for the Council in accompanying the healthy development of AI in Morocco. Within this development of AI, CNDH is particularly concerned about AI impact on the dissemination of content, its customization, moderation, deletion, advertising, targeting, the misrepresentation and undermining of pluralism, risk of opinion homogenization, and the most concerning issue, the incitement of hatred, discrimination, and radicalization. If AI captivates the world, it also calls on us to consider the space set aside for the respect of human rights in this world of algorithms.

These were some of the reflections carried out by the National Human Rights Council (CNDH) that I wanted to share with you that can be considered a

platform to launch a debate, at the national and international levels, on the healthy development of AI in Morocco.

References

[1]   المجلس الوطني لحقوق الانسان، "التقرير السنوي عن حالة حقوق الانسان بالمغرب لسنة 2019 - فعلية حقوق الانسان ضمن نموذج ناشئ للحريات،" الرباط، 2020.

[2]   المجلس الوطني لحقوق الانسان، "كوفيد 19: وضع استثنائي وتمرين حقوقي جديد - التقرير السنوي 2020 حول حالة حقوق الانسان بالمغرب،" الرباط، 2021.

[3]   M. AYAT, "Freedom of opinion and expression in international human rights law: between absolutism and restriction," *Arribat – International Journal of Human Rights Published by the CNDH Morocco,* vol. 1, no. 1, pp. 11-25, 2021.

[4]   A. Bouayach, «Artificial intelligence & digital citizenship. For ethical AI that respects human rights,» *Arribat – International Journal of Human Rights publisehd by CNDH Morocco,* vol. 1, n° %12, pp. 131-140, 2021.

[5]   N. Hilale, "The Evolution of Artificial Intelligence (AI) and its impact on Women: how it nurtures discriminations towards women and strengthens gender inequality," *Arribat - International Journal of Human Rights Published by the CNDH Morocco,* vol. 1, no. 2, pp. 141-150, 2021.

[6]   E. R. GOFFI, L. COLIN and S. BELOUALI, "Ethical Assessment of AI Cannot Ignore Cultural Pluralism: A Call for Broader Perspective on AI Ethics," *Arribat - International Journal of Human Rights Published by CNDH Morocco,* vol. 1, no. 2, pp. 151-175, 2021.

[7]   M. Bensalah, "Toward an ethical code of AI and human rights in Morocco," *Arribat - International Journal of Human Rights Published by CNDH Morocco,* vol. 1, no. 2, pp. 187-203, 2021.

[8]   M. BOSSI, "Processing data to Third Countries or International Organizations," *Arribat - International Journal of Human Rights Published by CNDH Morocco,* vol. 1, no. 2, pp. 176-186, 2021.

[9]   M. Bachelet, «Artificial intelligence risks to privacy demand urgent action,» Geneva, 2021.

[10]   UNESCO, Recommendation on the Ethics of Artificial Intelligence, Paris, 2021.

Artificial Intelligence & Human Rights: a bibliometric study

Jamila El Alami
Director of the National Center for Scientific and Technical Research, Professor, Morocco

My presentation is going to be focusing on scholarly research in the field of human rights and the topic of AI. Moreover, I will be making a brief comparison between the academic research done on the national and international levels about AI and human rights comprising the totality of scholarly research from 2016 to 2020.

Before starting, I would like to explain the choice of the title "bibliometric studies." As previously mentioned, I run the National Center for Scientific and Technical Research, a public institution placed under the authority of the Ministry of Education, Scientific Research, and Innovation. Our primary mission is the promotion of scientific research, its development, and valorization by the scientific community. We are one of the main institutions that implement the government policy in scientific research. Among our missions is to further the dissemination of scientific and technical information for the benefit of the scientific community, which is achieved by making available a wide range of databases such as SCOPUS and Web of Science to academics and researchers as well as conducting technological and scientific monitoring studies.

In this framework, bibliometric studies are conducted at the request of ministerial agencies.

I would like to start by introducing the current scientific research on human rights at the national and international levels before focusing on the intersection between scientific research about AI and human rights. So, let's begin by talking about the conduct of scientific research by means of the SCOPUS database.

SCOPUS and Web of Science are the most relevant databases worldwide, offering a

plethora of accredited scientific journals. Therefore I would like to look at statistics between 2016 and 2020, making this study very recent, regarding some specific keywords meticulously chosen so as not to overlook any publications in these databases, such as human rights, fundamental rights, civil rights, moral principles, moral norms, etc.

SCOPUS 2016-2020

Open access scientific production on the theme of human rights
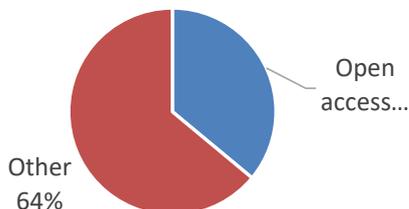
Type of publications on the theme of human rights
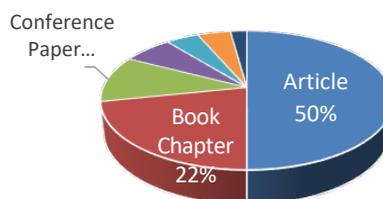
On the SCOPUS database, we have counted 56 676 publications on human rights from 2016 to 2020 at the international level, with only 36% being open access publications while the remaining 64% are paid ones. These publications were mostly journal articles, which make up 50% of the total of the publications, followed by book chapters amounting to 22% and publications on conferences making up 11%.
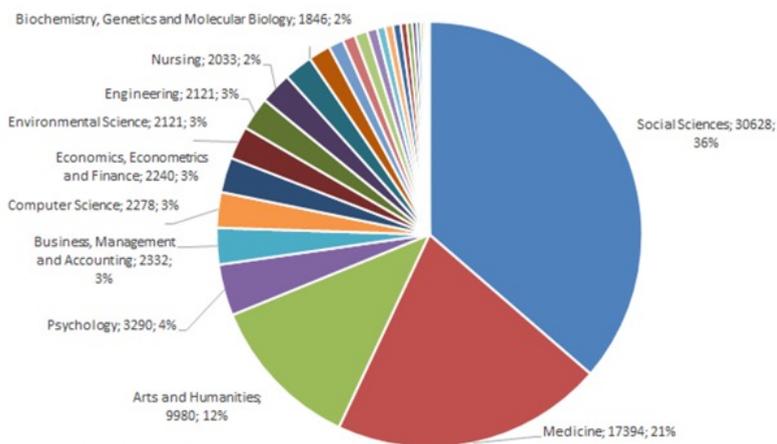
Figure 2: Distribution by disciplinary field of scientific publications on the theme of human rights

When it comes to specific themes, it is quite normal to find that 36% of this global corpus is found in journals of social sciences, followed by 21% of publications in journals of medicine, as human rights are also a topic up for discussion in the field. In comparison, publications in arts and humanities journals make up 12% of the total amount of publications. So, what are the countries that put out most research?
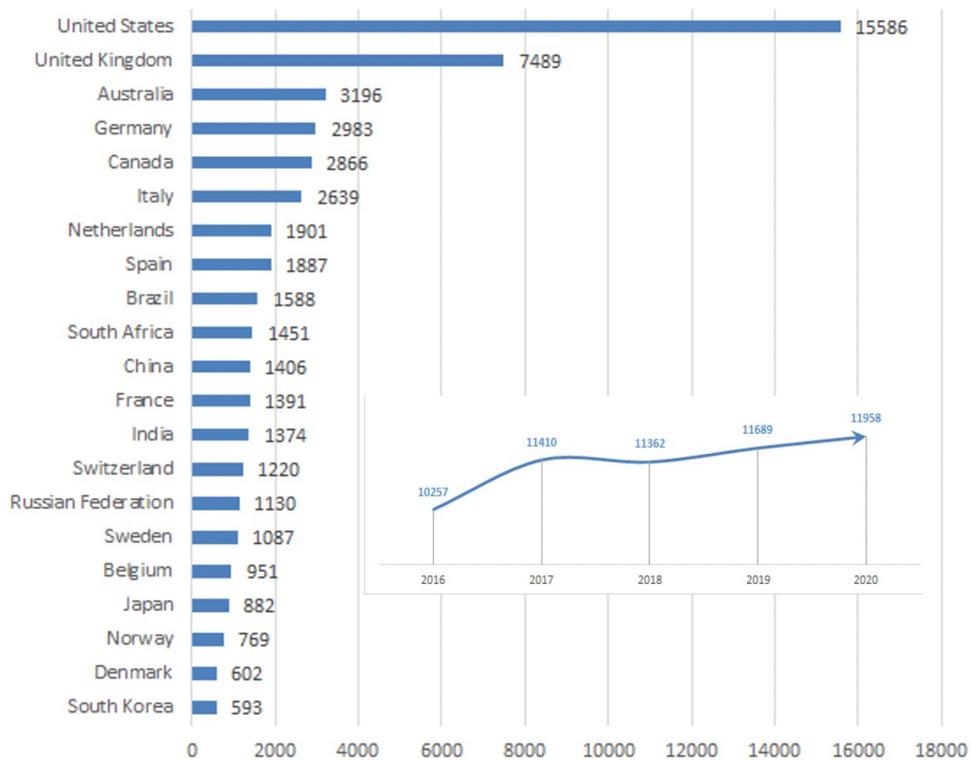
Figure 3: Productive countries on the theme of human rights (Scopus 2016-2020).

Unsurprisingly, the United States is in the lead with 15 586 publications between 2016 and 2020, followed by the United Kingdom and Australia.

The institutions affiliated with these publications are primarily universities such as the University of Toronto and Oxford University, which shows the importance of academic institutions conducting research in the field of human rights.

AFFILIATION	Nombre de publications
University of Toronto	537
University of Oxford	473
University of Melbourne	469
University College London	358
King's College London	352
The University of Sydney	347
Columbia University	340
Harvard Medical School	335
UNSW Sydney	308
Monash University	294
Universitetet i Oslo	291
The University of British Columbia	284
University of Cambridge	266
Yale University	263
Universidade de Sao Paulo - USP	256
The University of Queensland	255
University of California, Los Angeles	249
University of Michigan, Ann Arbor	246
University of California, San Francisco	245
University of Washington	236
Universiteit van Amsterdam	233

Figure 4: University affiliations with regards to publications on the theme of human rights
(Scopus 2016-2020)

The Moroccan corpus only counts 46 publications at the national level, which is a very timid number. The total Moroccan scientific production is that of 38 926 publications, making the percentage of publications on human rights seem very low. The positive sign is that those publications are cited 229 times, the more citatins a publication has, the more interesting it is and the more consulted it is by other researchers, demonstrating that it has an impact at the international level.
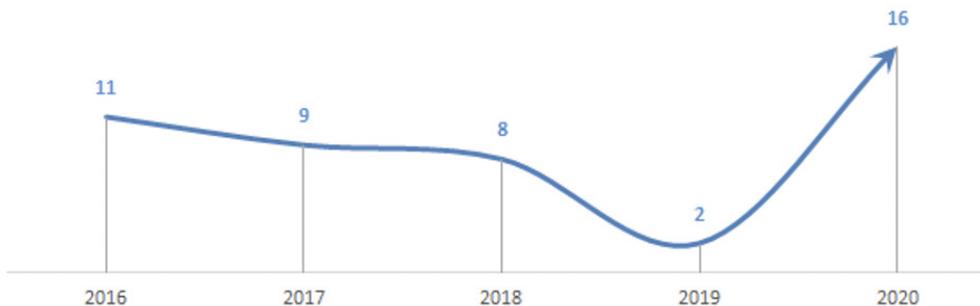
Figure 5: Temporal evolution of Moroccan scientific production on the theme of human rights (Scopus 2016-2020)

Regarding the evolution of scientific production in these years, we can see that there has been a general growth in Morocco, with the exception of the year 2019 where only 2 publications were put out. However, Morocco managed to increase the number with 16 publications in 2020. The most researched disciplines in Morocco follow the global trend with social sciences in the lead, followed by medicine, art and humanities. The Moroccan publications mentioned include both purely Moroccan publications with no foreign collaborators and those that have been collaborated on.

Country	Number of publications (more than 2)	
Morocco	26 local	20 in collaboration
United States		7
Switzerland		6
Egypt		5
France		5
United Kingdom		5
Canada		4
Tunisia		4
Belgium		3
Iran		3
Lebanon		3
South Africa		3
Sweden		3
Algeria		2
Australia		2
Denmark		2
India		2
Libyan Arab Jamahiriya		2
Malaysia		2
Oman		2
Qatar		2
Senegal		2
Sudan		2
Thailand		2
Yemen		2

Figure 6: Moroccan collaborations (Scopus 2016-2020)

In the SCOPUS database, 20 publications were produced in collaboration with other countries. The country of Morocco collaborates with most in the field of human rights is the United States, followed by Switzerland and Egypt.

In all fields combined, France is the first country in collaboration with whom Moroccan research is conducted, followed by the United States and Spain. Concerning collaborations on human rights publications specifically, France is, however, in fifth or sixth position, which is a much lower number than overall collaborations.

Web of Science 2016-2020

The same study was conducted using the Web of Science database for the same period of time.

The findings are as follows: 36 992 publications were found worldwide in the years from 2016 to 2020, 79% of these publications were journal articles, and 61% of these publications had limited access.

Similar to the SCOPUS database, the top contributing countries to research were the United States, the United Kingdom, and Australia, with most research coming from universities.
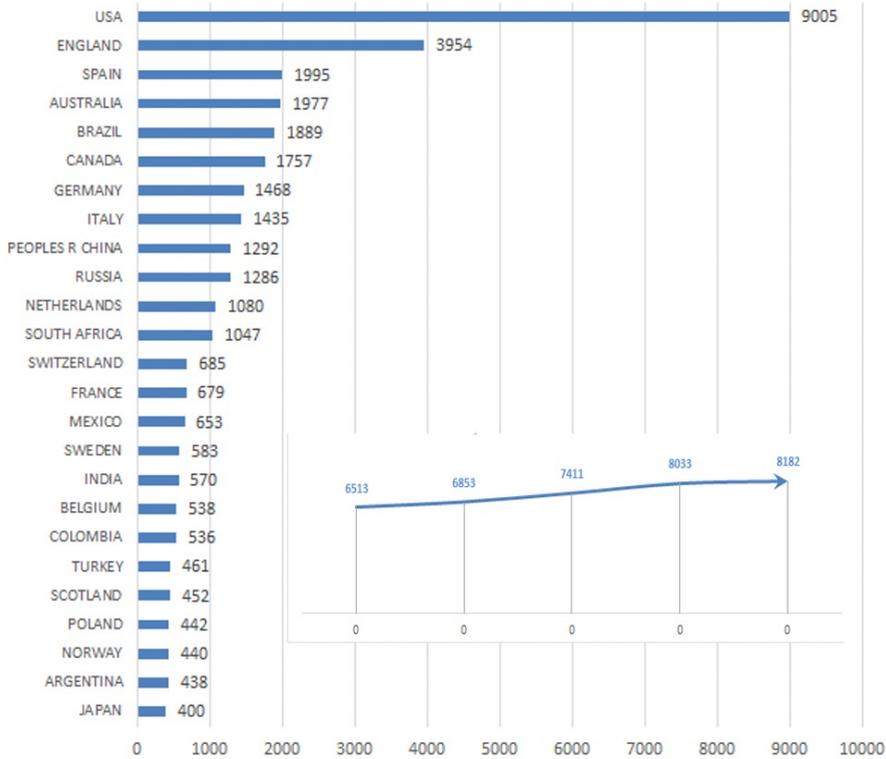
Figure 7: Productive countries on the theme of human rights (WOS 2016-2020)

Why do I insist on mentioning this information? When we conducted the same study on scientific research exclusively, we found research centers, private institutions, and agencies under the authority of the Ministry of Education and other Ministries such as the Ministry of Industry to be the main contributors, but in the case of human rights mostly universities work on the topic.

At the national level, Morocco has put out only 28 publications on human rights, most of which were journal articles.
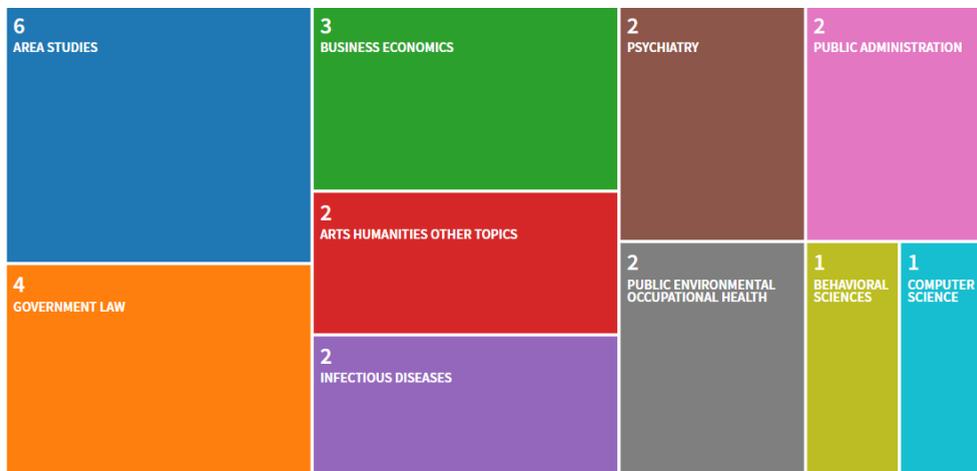
Figure 8: Distribution by disciplinary field of Moroccan scientific publications on the theme of human rights (WOS-2020)

The fields whose work are most cited are governmental laws, regional studies, and business economics. Again, the same pattern can be seen in terms of collaboration on research, with France being the first collaborator, followed by the United Kingdom.

Country	Number of publications (more than 2)	
MOROCCO	19 local	9 in collaboration
FRANCE		4
ENGLAND		3
SWITZERLAND		3
BELGIUM		2
CANADA		2
SENEGAL		2

Figure 9: Moroccan collaborations (WOS 2020)

As for Moroccan affiliations, we have to underline that Al Akhawayn University, the International University of Rabat, Sidi Mohamed Ben Abdellah University, and the National School of Public Health conduct most research in the field and have put out 2 or more publications.

After presenting the findings on human rights in both SCOPUS and Web of Science databases, we conducted the same study but added the search term Artificial Intelligence to the term human rights.
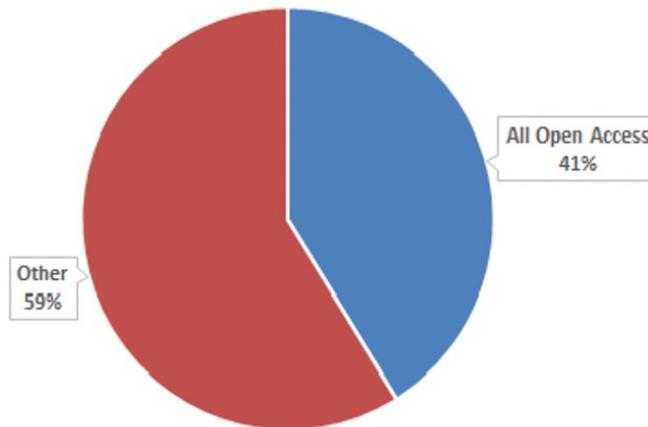
Figure 10: Open access scientific production on AI and human rights
(Scopus 2016-2021)

Using SCOPUS, we found 5724 publications discussing AI and human rights globally, 41% of which are open access and most of which are journal articles. Between 2016 and 2020, we have noticed an increase in publications.

However, it is important to note that the relatively low number of publications in 2020 is not to be taken into consideration since the census was conducted on November 25th, 2021, meaning there is still a month and a half for other publications to come out.
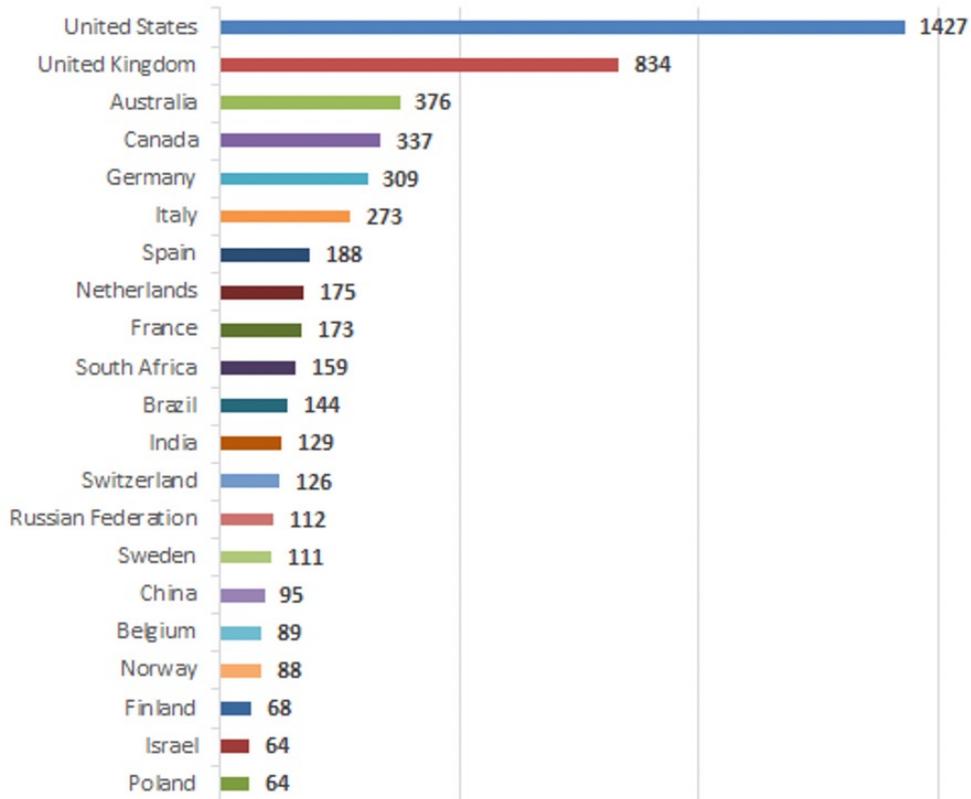
Figure 11: Productive countries on the theme of artificial intelligence & human rights (Scopus 2016-2020)

When it comes to the list of countries that put out the most publications, the same trend as before is observed, with the United States leading the list, followed by the United Kingdom. Similar to the aforementioned studies, it is universities that lead in terms of research in this field.

The Moroccan corpus counts only 6 publications on AI and human rights at the national level, all of which have only been cited 6 times. Morocco actually puts out a lot of research on AI, but when it is linked to human rights, the number of publications from 2016 to 2021 is reduced to 6.

Figure 12: Type of publications on the theme of artificial intelligence and human rights (WOS 2016-2021)

The Web of Science database contains 108 790 AI and human rights publications from 2016 to 2021, mainly comprising journal articles, followed by book chapters and conference papers. The states contributing most to research in this field are the United States and the United Kingdom, and in terms of institutions, universities are in the lead again.

Figure 13: Distribution by disciplinary field of Moroccan scientific publications on the theme of artificial intelligence and human rights (WOS-2016-2021)

Morocco has more publications on the Web of Science database on AI and human rights - even though this database is much more selective than SCOPUS. Indeed, Web of Science counts 73 Moroccan publications, which is reassuring. This amounts to at least 83 publications during the last five years at the national level. Furthermore, these publications were cited 275 times, which is very significant since it indicates the international level's visibility and impact of Moroccan research.

Pays	Nombre de publications	
MOROCCO	30 localement	43 en collaboration
FRANCE		17
USA		9
EGYPT		8
INDIA		8
PEOPLES R CHINA		8
ITALY		7
SPAIN		7
AUSTRALIA		6
BRAZIL		6
ENGLAND		6
CANADA		5
GERMANY		5
IRAN		5
SOUTH AFRICA		5
TUNISIA		5
JAPAN		4
LEBANON		4
NETHERLANDS		4
PORTUGAL		4
SWITZERLAND		4
AUSTRIA		3
BELGIUM		3
CAMEROON		3
DENMARK		3

Figure 14: Moroccan collaborations (WOS 2016-2021)

Morocco collaborates with the same countries on the topic of AI and human rights, with 17 publications in collaboration with France, 9 with the United States, and 8 with Egypt.

Affiliation	Nb de publications
MOHAMMED V UNIVERSITY IN RABAT	16
HASSAN II UNIVERSITY OF CASABLANCA	11
MOHAMMED FIRST UNIVERSITY OF OUJDA	9
UNIVERSITE INTERNATIONALE DE RABAT	8
AL AKHAWAYN UNIVERSITY	6
IBN SINA UNIVERSITY HOSPITAL CENTER OF RABAT	5
CHOUAIB DOUKKALI UNIVERSITY OF EL JADIDA	4
IBN TOFAIL UNIVERSITY OF KENITRA	4
ABDELMALEK ESSAADI UNIVERSITY OF TETOUAN	3
CADI AYYAD UNIVERSITY OF MARRAKECH	3
IBN ROCHD UNIVERSITY HOSPITAL CENTER OF CASABLANCA	3
IBN ZOHR UNIVERSITY OF AGADIR	3

Figure 15: University affiliations affiliations with regards to publications on the theme of artificial intelligence and human rights (WOS 2016-2021)

As for the universities' ranking, Mohamed V University takes the lead with 16 publications, followed by Hassan II University, Oujda University, and the International University of Rabat. There is certainly research being conducted and published at the level of Moroccan universities, but it remains timid.

To sum up, we can say that AI is a topical subject when linked to human rights, with said linkage being indispensable, yet publications are clearly not enough in the field. So, what should we do to resolve this lack of research? Here, I would like to invite the CNDH Chairperson to consider a call for projects on human rights as that implies funding and encourages scholars and researchers to work on the topic. CNRST can assist the Council with this endeavor since we have a high degree of expertise, ranging from drafting terms of references, call for applications, and evaluation, additionally our 240 expert evaluators affiliated with CNRST, can be beneficial.

In conclusion, I am calling for more funding directed toward research on human rights and in order to be able to note that more researchers are working on this crucial topic in the upcoming years.

Universities, Artificial Intelligence & Human Rights: a reflection

Azzeddine El Midaoui
President of Ibn Tofail University, Morocco.

Moroccan universities have always played a central role in terms of the education of the country's leaders and executives. Up until today, 99% of public and private executives are alumni of Moroccan universities. Thus, Moroccan universities have contributed with their alumni to help face the Covid-19 pandemic, judging by the excellent work done by workers and executives in healthcare and other sectors. As university bodies, we are an extension of this country as a whole. Therefore, although we are not perfect, we are always in need of recognition, such as this invitation to partake in this seminar.

Artificial Intelligence (AI) is such an incredibly complex, large, rich, and conflicting topic. It is most interesting to see the National Human Rights Council (CNDH) being proactive by engaging in such discussions, as it is indubitably the subject of hundreds of theses and research projects globally, including in our country. As you may know, AI came out of university labs.

As a chemist, I must admit that I do not work closely with AI, but as a person running a university, I do have some knowledge about it. On the other hand, I have been involved in human rights since my student days.

Today, AI is inevitable. Humans cannot do anything to stop scientific and technological evolution nor resist changes. Printing was invented a few centuries ago and revolutionized the lives of human beings then. Although scientific progress was stalled for a long time, things started evolving very quickly starting in the 19th century. Had the scientific evolution that humanity has known in the last 20-30 years been at the speed of evolution before the 19th century, we would have needed many centuries to get where we are now. Nowadays, technological evolution has caught up and even surpassed humans; just like globalization, we cannot do anything

to stop it. Nevertheless, human intelligence and that of societies and groups are not supposed to stop the hurricane that is scientific evolution, but rather predict what can happen, adapt to it, and be equipped to avoid harm, benefit from it, and generate profit.

We should also be aware that AI is just a transition to other technological evolutions coming our way and that can become operational any day now. So, we should not see AI as an endpoint as it is not; AI is a process, the start of a new chapter. While several studies in labs and research centers are trying to measure the impact of AI, this has proven to be challenging, and only predictions have been made on its impact on jobs or the economy. In contrast, its long-term impact on humans is still unknown. Everything could change dramatically as there is not enough visibility. Still, we need to be prepared as this change bears risks, even more so for consumer societies that do not produce technology and only consume it.

A straightforward example is the telephone, which revolutionized everything. I believe that no one today knows how children that are now 5-6 years old will be in 15 years. The difference between generations evolves rapidly, and no one knows what is coming. Today, we do not educate children the same. They are not being raised like us, where we were constantly face-to-face with our parents. They were aware of the changes we were going through at any time. The telephone educates and transforms parents as well.

Despite the universality of human rights, there are gaps between the latter and technological evolution, as laws and rights evolve much slower than technology. But a time will come when these gaps will be minimized with the necessary readjustments, an opportunity to improve the well-being of human beings.
The only remedy to this is education, training, and raising awareness.

The impact of this technological evolution may be very damaging to marginalized populations, to gender issues, and to rural areas where the literacy rate remains low. So, I believe that the impact of AI on human rights will have consequences on freedoms, gender equality, social and psychological aspects, and the right to

employment….All of this is of interest to experts and human rights defenders. Some predict that by 2050, 60% of jobs carried out today will cease to exist. However, all of this remains unclear; there is nothing exact, just preliminary studies and predictions.

With regards to Morocco and AI, there is work being done. The state is working on the amendment of Law 09-08 creating the National Commission for the Control of the Protection of Personal Data (CNDP), and there is also an active discussion about introducing legislation to protect citizens further.

At the university level, we both have opportunities and challenges. In terms of opportunities, we partake in the creation of change. We are lucky to be witnesses to the change occurring within the youth, we have the means to measure the impact of this change, and we can also play a significant role in raising awareness. Currently, Moroccan universities have 1 100 000 students, which can be a power to reckon with for institutions such as the National Human Rights Council to raise awareness, conduct studies of impact, assess predictions, etc.

To conclude, I would like to reiterate that Moroccan universities, with all of their components, executives, professors, and students, consider themselves strategic tools for the benefit of all state institutions. Therefore, if the Council ever needs to collaborate with universities, and Pr. El Alami has already mentioned this, we are at its disposal to work together to prepare ourselves to face many challenges, including the impact of AI.

Digital Development in Morocco and the Digital Development Agency's contribution to the emergence of an AI ecosystem

Sidi Mohamed Drissi Melyani*, Khouloud Abbeja**
* General Manager of the Digital Development Agency, Morocco.
** Digital transformation director, Digital Development Agency, Morocco.

I will present the Digital Development Agency (ADD) briefly before discussing a partnership with the National Center for Scientific and Technical Research (CNRST) currently being implemented, a call for projects regarding applied scientific research on AI. This experience can be considered a first example of a successful partnership that can be adopted to develop skills, further research, and have national tools adapted to our context.

First, I would like to recall the Royal Directives regarding the development of the digital sphere. We have two excerpts of the Royal speeches of his Majesty the King, may God assist Him. The first derives from a speech of July 29th, 2018, "Government agencies will coordinate and exchange information, using modern computer-based technology to this end."[1] In the second quote from the speech dating back to October 2016, HM the King stated that "E-administration must be generalized following an integrated approach enabling different departments and various services to have a common access to information."[2]

Regarding the digital field in Morocco, various strategies have been adopted for several years now in order to develop the field. The creation of the ADD can be considered a crucial stage within these different strategies since Morocco has a public institution whose mission is to implement strategies in the digital field for the first time ever. At the end of 2017, Law No.61-16 founding the agency was promulgated

---

1 « C'est, en effet, aux services publics qu'il revient d'organiser un échange coordonné des informations, grâce au recours à l'informatique et aux nouvelles technologies. »

2 « L'administration électronique doit être généralisée selon une approche intégrée permettant aux différents départements et aux divers services un accès commun aux informations. »

alongside its implementation decree, and in 2018, recruitment and implementation at the organizational, procedural, and budgetary levels took place. In addition, the National Administration Reform Plan was launched the same year by the Department of Administration Reform. In 2020, an evolution of the regulatory framework compared to what was being done previously took place, i.e., the law regarding the simplification of procedures, the new version of electronic identity cards, and other laws such as a new law related to electronic signature, as well as cybersecurity. The focus in 2020 was evidently on regulatory reforms.

2020 was also the year the pandemic began to spread and boosted the digital sphere in the sense that public institutions and companies acquired tools to work online, which accelerated digitalization. For example, the ADD set up a number of tools and shared applications on behalf of administrations, and we have seen a significant increase in the number of users; 900 administrations adhered to a digital run office (Bureau d'ordre digital) and other shared applications.

Amongst the recent regulatory reforms implemented in 2020 are Law No.72-18 on the targeting schemes regarding beneficiaries of social support programs and the creation of the National Agency of Records (Agence Nationale des registres) - one of the components of this law is the establishment of a citizen identifier which is necessary to digitalize administrative procedures from end to end-, Law No.55-18 regarding the simplification of administrative procedures, Law No.04-20 on the electronic ID card -a new version of the ID card rendered digital, Law No.05-20 concerning cybersecurity, Law No.43-20 on trusted services regarding electronic transactions which facilitates electronic signature procedure so as to generalize its use since the latter is mandatory to carry out electronic transactions.

As previously mentioned, the pandemic has boosted the use of digital technology. These numbers below - a 55% increase in data usage, a 43% e-commerce increase, etc. - illustrate this statement.

Trafic Data	Distance learning	E-commerce	Justice	E-gov
+ 155% increase in data traffic for the mobile segment during 2020	**600 000** daily users of the national platform « TELMIDTICE »	**+ 43%** increase in domestic card-based e-commerce transactions in 2020	**14.161** hearings held remotely in the courts between 27/04/2020 and 29/01/2021	**+900** ministries, local authorities and EEPs to join the digital order desk in 2020

Figure 1: Accelerating the adoption of digital technologies with Covid-19 under 2020.

The ADD is a public strategic institution under the authority of the Ministry of Digital Transition and Administrative Reform. The ministry, having been changed with the appointment of the new government, is in charge of implementing strategies in the field of digital technology.

The agency has four focus areas: E-administration, digital ecosystem and innovation, social inclusion and human development, and environment and digital trust. Since its establishment, the DDA has worked on 15 projects in the different areas previously mentioned. 5 out of these 15 projects are considered a high priority; one of these projects is related to AI.

Other high-priority projects include an interoperability platform that allows users to stop playing "mailman" between administrations; since the latter takes on that role and internally circulates the supporting documents for the procedure that the citizen presents to the central administration. The law on procedure simplification also foresees digitalizing the entire procedure so that the citizen can choose between going to the administration or doing the entire procedure online. The digital factory is a development space within the ADD using Agile methods to benefit the Moroccan administrations. As previously mentioned, an evolution of the regulatory framework is necessary for the regulations regarding the digitalization field. Therefore, it is a major project of high priority for us.

The fifth major project we have undertaken is Digital Generation, which tackles training, specifically on how to accelerate training courses and generalize training in the digital sphere. Training is crucial in AI since it is a new technology. In order for

decision-makers, officials, and other actors to take this technology into consideration, they need to be trained in it to understand it better and give their opinions on AI.

The agency operates in three ways. It is at the service of its stakeholders and cross-sectional since digital technology is featured in many aspects: in the private sector, administrations, higher education, research institutions, civil society… So, we work with a transversal approach in 3 intervention modes. We adopt a 'faire-faire' approach that consists of controlling projects' implementation and evolution in the digital field, an assisting approach in project management to help administrations and economic sectors, and directly implementing projects. Still, it is usually carried out in partnerships since, as mentioned, the digital sphere is cross-sectional.

The last part of my presentation concerns the agency's contribution to the emergence of an AI ecosystem, with the example of the AI Khawarizmi project undertaken in partnership with CNRST.

I would like first to mention some numbers and elements to contextualize trends regarding AI. We are living through the biggest technical revolution since the electric revolution. Some studies predict that 70% of companies worldwide will have adopted AI and that AI will have generated a 15,7 trillion dollars increase in the world's GDP, which translates to a growth of 1,2% per year. Both developments will have occurred by 2030. This can deepen the gap between countries, companies, and workers if public policies implemented are not adapted.
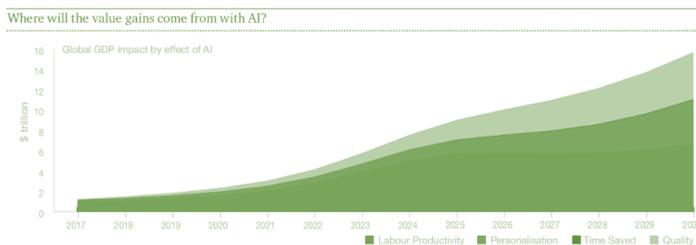
Figure 2: Global GDP impact by AI (Source: PwC Analysis).

In this context, ADD, and CNRST, in partnership with the Ministry of Industry and the Ministry of Higher Education, launched, in May 2019, the AI Khawarizmi program in order to promote scientific research in the field of AI and its applications. The

budget allocated for this project is 50 million MAD, a large sum reserved for research on AI. CNRST launched a call for projects in partnership with ADD, and presently, 45 projects were approved, considering that we had 700 accounts created on the platform and 251 projects submitted. 70% of these projects were from Moroccan public universities, projects were spread out on 16 themes, and 16% of women were project holders.

It is also important to highlight that 40% of Al Khawarizmi projects are related to healthcare; as the call for projects was launched before the pandemic, there is no direct link to Covid-19. Other projects are related to energy, water, environment, industry, agriculture, education sciences and training, logistics and transport, tourism and culture, justice, Big Data, machine vision, and smart cities. We intend to coach this community of researchers in order to facilitate their programs. We will also be organizing a series of thematic workshops with our various partners so as to facilitate institutional sponsorship, access to work resources, and scientific and socio-economic valorization of research projects' results. We want the participants to use applied research that results in concrete projects that can be marketed, and this is why we have matched projects to ministries according to their field. We also intend to include ministries in the digital ecosystem interested in the outcomes of these projects during the workshops we will be organizing. In addition, we will organize trainings and awareness-raising initiatives for the benefit of Al Khawarizmi project holders to assist them throughout their entire research, to anticipate aspects related to the industrial and intellectual property, and the socio-economic valorization of their research results. In the meantime, we plan to organize scientific events during the Al Khawarizmi project on sectorial themes related to AI.

As part of the operationalization of its AI ecosystem project, the ADD will soon be launching a study aiming to develop a national roadmap for Artificial Intelligence (AI) and define an operationalization plan for pilot AI ecosystems. The study will first frame the actors, internal and external contributors of the mission in order to assess the state of the art and the degree of maturation of the sectors suitable for the development of AI ecosystems and presentation of priority pilot ecosystems through a national and international benchmark. This study will be conducted to

contribute to the definition of the national strategic vision in terms of AI. Through it, the ADD will be able to develop its national roadmap for AI by implementing a pilot ecosystems operationalization plan.

The topic of human rights is crucial now more than ever since AI users' data is nowadays more "exposed" because we are increasingly using digital devices every day. Therefore, it is essential to guarantee respect for digital privacy. Moreover, with AI, it is necessary to develop skills to ensure equal access to this technology so as to not create a digital divide. Although many people are already facing this digital divide due to their inability to be connected or lack of tools to access this technology, there is a need to ensure equality, and here the training aspect is really important. Therefore, ADD is open to launching new training programs on AI focusing on inclusivity so that citizens as a whole benefit from it.

- II -
ISSUES & CHALLENGES

Cultural Pluralism as an Essential Part of the Ethical Assessment of AI

Dr. Emmanuel R. Goffi
Co-Director and Co-Founder of the Global AI Ethics Institute, France.

Introduction

There is a lack of outside-the-box standpoints regarding ethics applied to Artificial Intelligence (EAAI). It is thus highly appreciable to see that the National Human Rights Council of the Kingdom of Morocco (CNDH) is opening its mind to non-mainstream perspectives.

Despite all the comments and statements made regarding diversity, the last recommendation adopted by UNESCO included [1], discourses are not always what they seem, and behind each word, there is a reality that invites us to question ourselves, our perceptions. The work conducted by UNESCO is no exception; and even if it is undoubtedly valuable, it is nonetheless built on biases and represents a specific understanding of the world we live in, a cosmogony, a representation of the reality of diversity.

Statements regarding values like democracy, human control, just societies, and social scoring, to mention but a few, are all considered through a universalist perspective that could be disputed for hours. For instance, democracy is always presented as the best and consequently desirable political system when it comes to reaching peace, which has not always been seen this way. A quick look at philosophy through Aristotle's or Plato's works shows that democracy did not benefit from such a high-value consideration. The fact that democracy is the best political system in the world is Western-oriented. In the same way, the fact that we must have control over technology is a Western-oriented perspective. The hierarchical relation to technology in which the idea of control is rooted, does not exist in some societies grounded in an Animist, Buddhist, or Shinto culture.

This idea that we need just societies is questionable for, firstly, "just society" is undefinable absolutely and would mean different things depending on where one lives, the political system one is subject to, the culture, the geopolitics, etc. Just society does not refer to nor summon the same notions in North Korea, for the Taliban regime in Afghanistan, in China, or for the people of the Kingdom of Morocco. By asserting principles and values and presenting them as universal, we deny the right of people with divergent perspectives to be heard. We send them a clear message that they do not deserve our consideration, that their opinions are not relevant, and that their ideas have no place in the debate. Such behavior is clearly the product of a biased stance, mostly the idea that some countries hold an indisputable truth regarding what is acceptable and what is not, what is desirable and what is not, what is right and what is wrong.

Moving Beyond *A-Priori*

Yet, embracing diversity and respecting human rights to be different request the ability to listen to others beyond disagreements and beyond polarized stances. This does not mean that one should fall into the trap of absolute relativism. This does not mean that people must agree with others. It merely means that we must accept that there is a multiplicity of stances out there, and there are people who do not share our ethical perspective and the values and principles it encompasses. Short of it, we will eventually build norms that will not be applied evenly. Human rights are a perfect example of such norms, based on a narrow perspective and made pointless because of the legislator's inability or unwillingness to listen to divergent positions, its inability or unwillingness to work with all stakeholders, its inability or unwillingness to be open-minded enough and to listen to things that can sound odd, but nonetheless, that must be expressed and listened to.

Another issue that could be discussed for hours are biases. Biases are an element of human behavior that is translated into everyday activities and our interactions, not only with other human beings but also with our environment at large. Biases are unavoidable. The philosophical question here is not *can* they be avoided or mitigated, but *should* they be avoided? Even more, *should* they be removed? And in

some places, biases that are rejected in the Western world, for instance, are part of the way people identify themselves and others. In Mali, people from the South consider themselves as Black, while they label people from the North as White. In many countries, discriminating between genders is totally acceptable, if not desirable. Discrimination stemming from biases can be, in some instances and cultural settings, a necessary element of differentiation and belonging to a specific community.

The ethical issue here is the inability and unwillingness of some people to listen to this reality, and their support at any cost, almost ideologically, to a unique oriented perspective while talking about diversity, respecting and embracing it, requests listening, open-mindedness, and tolerance towards those who do not share our values and conception of the world.

It is essential to embrace diversity in all its dimensions, whether we like or agree with some of them or not. It is essential to listen to divergent perspectives and to respect the rights of each and every human being to have their ethical stance. It is essential not to deny them this right the same way we do not want others to deny us the right to difference.

Defending Cultural Diversity

This right for cultural diversity is so strong that it is even enshrined in international treaties such as the Universal Declaration of Human Rights (art. 22), United Nations Charter (art. 13.2), or the International Covenant on Economic, Social and Cultural Rights. In many reflections and documents, respect for cultural diversity, which encompasses respect for ethical diversity, is a right linked to human security and dignity and seen as part of the demand for freedom from fear.

The Universal Declaration on Cultural Diversity issued in 2001 by UNESCO itself states that "the defence of cultural diversity is an ethical imperative, inseparable from respect for human dignity." Such a strong statement cannot be ignored, even if somehow paradoxical, since it calls for the defence of cultural diversity, building the argument on a narrow Kantian perspective on ethics (the ethical imperative)

and implying that human dignity is an end in itself. Whatever our stance on human dignity, a quick look at the world suffices to see that this notion is not universal and not interpreted the same way, depending not only on cultures but also on circumstances. The United States with Guantanamo, Abu-Ghraib, or extraordinary rendition towards so-called "black sites" are one illustration among others of the variable geometry appreciation of human dignity. To go even further, in some places and some instances, human dignity is not even a concern, let alone an issue.

Looking at freedom or human rights internationally, it is clear that these are concepts unevenly understood and applied. Although the idea of human rights is deemed highly important, it must be kept in mind that at the time when the Universal Declaration of Human Rights was ratified, the United Nations only counted 58 countries (compared to 193 today); among them, many empires (which questions the way the Declaration has been "imposed" to colonized countries). Only 48 out of the UN members supported, signed, and ratified the Declaration (5 abstained and 5 refused to participate), which means that already at that time, many countries, people, cultures, communities were not aligned or did not agree with the concept of human rights and that some communities were left with no other option but to accept it.

There is consequently something paradoxical in saying that cultural diversity is important and at the same time denying this diversity imposing, through hard or soft power, a limited ethical perspective based on a narrow appraisal of what is highly valuable and important.

About Culture

Another element that would deserve further discussion is the notion of culture. Culture is highly difficult to define, and there exist many cultures, none of them being perfect, but I do not think that there is one that can be perfect. One of the most accepted definitions is provided by Kroeber and Kluckhohn:
"Culture consists of patterns, explicit and implicit, of and for behavior acquired and transmitted by symbols, constituting the distinctive achievement of human groups,

including their embodiments in artifacts; the essential core of culture consists of traditional (i.e., historically derived and selected) ideas and especially their attached values; culture systems may, on the one hand, be considered as products of action, on the other as conditioning elements of further action." [2]

This definition underlines that culture encompasses values, an essential element on which ethics is built. Ethics cannot exist without preexisting values. Having a look at values, it appears that, conversely to mainstream assertions, there is no such thing as universal values. Even the work done by Shalom Schwartz [3] fails to technically prove that there would be even one value shared universally. It is important here to stress that words matter and consider that "universal" is mostly understood as referring to something, in our case values, that is shared by everyone, and that is existent or operative everywhere in the world with no exception.

So far, no convincing study has proved that there would be even only one value commonly accepted by the entire world, that is to say, universal in the strictest sense. So, the question: if there is no universal value, how can there be any form of universal ethics? Regarding Artificial Intelligence (AI), short of any universal value, how can we consider elaborating a universal code of ethics applicable to AI? There are different ways to answer this question. The most common one is to consider that "universal" is limited in scope and that exceptions do not detract from the relevance of the supposed universality of some values. Another answer would be to assert the superiority of some values notwithstanding oppositions from some actors. The former answer is misleading in that it leads to believe that wherever one goes, whoever one talks to, all people around the world share some kind of common value. The latter is ethically problematic and dangerous for it both leads to absolutism and to the setting of an artificial hierarchy of values, which would mean a hierarchy of ethics, and eventually a hierarchy of culture if one agrees with Kroeber and Kluckhohn that "the essential core of culture consists of traditional (…) ideas and especially their attached values."

Another option is to consider that establishing a universal code of ethics applicable to AI, or any universal code of ethics at large, is merely impossible, but to impose

one perspective arbitrarily deemed universal. Obviously, such a stance does not go without concerns, notably regarding the risk of sliding towards absolute relativism, which is no more desirable than absolute universalism.

Cultures are contextual, and so are values. What is deemed valuable in France might differ from what is regarded as valuable in Morocco, which itself might differ from what is considered valuable in China. Consequently, talking about cultural diversity requests a deep dive into the values of a given community from which stem a specific ethical setting.

Going further, it is worth stressing that even where some shared value seems to exist, words and their meaning must be questioned. In many instances, "friendship" is presented as a universal value. Building on Ferdinand de Saussure's work, such an assertion must be analyzed through the difference between "signified" and "signifier," which are the two components of signs or words [4]. While the signifier refers to the readable or audible image of the sign, the signified refers to the concept one puts behind the signifier, its meaning. So, translating words from another language, which is already a problematic point, into "friendship" and considering that there is a shared signifier, does not necessarily infer that there is a shared signified. Put simply, a common word does not imply a common meaning. That applies to "human" as well as to "rights," and obviously to "human rights." That even applies to "value" or "ethics," not to mention "Artificial Intelligence." Going beyond the "tyranny of words" is all but easy, yet it is necessary [5] [6].

Dutch sociologist Geert Hofstede [7] remind us that culture is the product of "the collective programming of the mind that distinguishes the members of one group or category of people from another", and this programming is, according to constructivists, made through language, and eventually the words [8].

Whatever the definition, culture is subjective and contextual, and so are values and ethical norms that stem from them. The call to defend cultural diversity, expressed in the UNESCO Declaration, cannot be decorrelated from the defense of ethical diversity and, eventually, the polytheism of values as suggested by Max Weber [9].

Eventually, reflecting on a notion such as human rights requires a deep and thorough consideration of the cultural grounds, including the system of values on which a specific ethical setting has been built and from which stems the notion. It also demands a rigorous assessment of the narrative about the notion. Simply said, it involves as much objectivity and intellectual honesty as possible.

Cultural Pluralism and Ethics Applied to Artificial Intelligence (EAAI)

When it comes to the establishment of an ethical governance of AI, cultural diversity must thus be considered and accepted as an unavoidable factor to be integrated into the debate.

The way words shape how we (mis-)perceive the world, the way we think it, and the way we behave has been very well studied and documented by social constructivist thinkers following Berger and Luckmann's work. Using this framework of analysis, it appears that if culture is the product of our values, it is also the product of the collective programming of our minds, as Hofstede demonstrated it.

Cultural programming is not limited to geographically rooted communities, shared historical journeys, to nations, nor is it to belonging to spiritual or linguistic groups. It also encompasses closer social groups in which one is involved, such as family and friends, profession, or education environment. That explains why the UNESCO, even if they gather representatives of all the countries, has so much difficulty stepping back from their own *a priori* regarding AI when it comes to the establishment of recommendations on the ethics of Artificial Intelligence [1]. The fact that the discussion includes a wide panel of representatives does not exclude the fact that most of them have the same kind of educational backgrounds, training, careers, and interests. Conformism and the confirmation bias at play in such a group cannot be ignored.

Stepping back from the mainstream narrative in this context is difficult, nay impossible. Analyzing things with a surgical eye and putting aside one's own beliefs is not something obvious. Yet, in the field of ethics more than in any other, axiological

neutrality defended by Max Weber in its *Politics as a Vocation* lecture given in Munich in 1919 is vital to make sure that cultural diversity is fully respected.

Unless such an axiological distance is not requested to establish ethical norms, subjective elements such as emotions, beliefs, feelings, and opinions will contribute to the setting of narrow-minded norms that will not be applied and, in some cases, could be violently rejected.

Applied Ethics: Putting Ethics Back into a Context

With the rise of AI and questions regarding associated risks, ethics has gained in importance. At least in speeches. Except for the fact that most commentators dealing with improperly called "AI Ethics" are not experts in ethics, EAAI lacks contextualization. This paradoxical, since applied ethics refers to ethics applied to a specific object, which context cannot be ignored. Yet, in EAAI, things are done as if context does not matter, as if AI was a homogenous object independent of its environment, like if its design, development, deployment, and use were "a-contextual." Added to the belief that there are universal values on which universal ethics can be built, it seems obvious to like-minded non-experts and other stakeholders with vested interests who strongly believe in and promote the idea of a universal code of ethics.

In EAAI, like in medicine, wrong diagnostics lead to wrong cures. Decontextualization and denial of diversity lead to the establishment of an arbitrary one-cure-all ethical solution to frame the development and use of AI. This is not only ineffective, but it can also prove counterproductive and even dangerous. Counterproductive in that instead of regulating AI, the current multiplication of ethical codes pertaining to AI, stemming from the inability to find common grounds on which to build a common ethical frame, tends to deregulate it. Dangerous in that denying the reality of cultural diversity and failing to defend it put us on the slippery slope toward moral absolutism [10]. Along with the growing use of "cosm-ethics," that is a narrative built, intentionally or not, on words of ethics used to hide the impurity of some purposes behind a veil of ethical acceptability, like cosmetics would hide skin impurity, all

elements are present to favor and impose a narrow perspective on EAAI and to legitimate the non-respect of cultural diversity [10].

Therefore, EAAI must be put into perspective. It must be related to a specific ethical setting.
Figure 1 below shows that even if ethics seems to be at the center of the debate over AI, it is necessary to move back and consider the bigger picture made of different interlocked ensembles.

When a specific situation raising questions linked to ethics occurs, the first environment to be considered is the normative one, including both legal and moral aspects. A situation cannot be considered as ethically concerning outside of a moral setting, establishing what is deemed acceptable and unacceptable for the community facing the situation. With the situation appears several ethical perspectives appraising the situation and providing with somehow different insights on the matter. These ethical assessments take into account both moral demands at play and the legal instruments to be summoned. All the more so, as in many instances, morality and laws, even if different, are intertwined and cannot be strictly separated. Ethical thinking about the death penalty, for instance, cannot be withdrawn from its moral and legal context and would thus be different in the United States, Saudi Arabia, France, or Japan, where moral and legal settings differ.

Next, come values on which norms are built. A country praising democracy will not have the same ethical assessment of the impact of AI on privacy as a country adopting a more authoritarian political system. The legal and moral rules stemming from these values will obviously differ depending on the perspective. On the one hand, democracies will tend to defend privacy, while authoritarian systems may lean towards less or no respect for privacy. The same goes with other values, such as the most mentioned ones that are human rights, equality, and freedom. Change one value, and the normative frame will change and lead to a different ethical appraisal of AI-related concerns. If privacy is not a concern, there will be chances that there will not be legal instruments protecting privacy, as well as no strong moral stance in favor of such protection. Consequently, public institutions' ethical appraisals of the

use of data will be way different from those made in a setting where privacy is seen as deserving protection.

Back to democracy, considering it as a universal value leads some actors to move forward and deem it necessary to enforce it by force. The imposition of democracy in Afghanistan has shown the limitations of this poor narrow-minded copy-past-mentality which does not consider the bigger picture, including cultural differences. Democracy is a cultural heritage that has contributed to the cultural programming of some peoples for centuries. Many people, countries, or communities around the world show no interest in Aristotle and Plato's political philosophy. The word democracy can itself be disputed. Democracy in France does not mean democracy in the United States or the Democratic Republic of Congo. Again, behind a common signifier can hide several signified.

All this is included in a bigger ensemble made of cultural features that contribute to the establishment of values based on common preferences and shared perspectives about the world around. The Moral Machine Project is one illustration, among others, showing societies driven by a collectivist perspective have a different appraisal of ethical situations related to autonomous vehicles than individualistic cultures [11]. In collectivist cultures, values will focus on the group; consequently, norms will focus on the community instead of on the individual. In such a context, privacy will not be ethically appraised the same way in both environments. Nor will be human rights or freedoms or equality.

In the same vein, the myth of control at play in the Western world shapes the relation to technology. In non-hierarchical cultures, such as the one adopting Shinto or Ubuntu wisdom, and where human beings are not seen as superior animals but as mere cogs within a bigger ecosystem, the idea of control over nature through technology does not have the same meaning. Neither has control over technological tools that can even be granted with some kind of life. Consequently, the risks and benefits linked to AI will not be assessed the same way in such cultures as they are in individualistic societies where technology is considered as a tool and control (human-centric AI) is a strong demand.

The next ensemble is made of the overall environment, including geopolitics, beliefs (i.e., spiritualities), resources, geography, political systems, and history, to cite but a few. Here again, context is key. For example, Israel considers its army (Tsahal) as the most moral army in the world. Not that they assessed their military through philosophical reasoning, but they consider that since the homeland is under constant threats from their neighbors, their army, which is dedicated to the defense of the country, has a just cause to fight using whatever it takes to ensure the security of the country and its people. In places where war and threats are less present, it is easier to take time debating on right and wrong. Regarding AI, while in Europe, people are concerned about ethical issues such as sustainability endangering, privacy violations, or biased decision-making, in other places, people are concerned about survival and see AI as a potential tool to make benefits and free themselves from misery. Different settings, different perspectives of EAAI.

At the end of the day, changing one feature will lead to the re-assessment of the whole ethical perspective on the igniting situation. That does not mean that everything will inevitably change drastically and shift to a totally different standpoint. It means that more or fewer nuances will be brought to ethical appraisals depending on the variables used in the final equation. Change the geopolitical context or history, and we will end with a different political system, which in turn will influence and modify cultural features, which themselves will impact values, then norms, and eventually ethical appraisals of a specific situation. Conversely, different ethical appraisals of a specific situation can influence norms, then values, cultures, and finally, the overall context. The whole system is in constant motion. Interaction between ensembles leads to constant evolution at each level. Changes can be slow. They can also go fast, especially when a crisis occurs, leading to a paradigmatic shift.
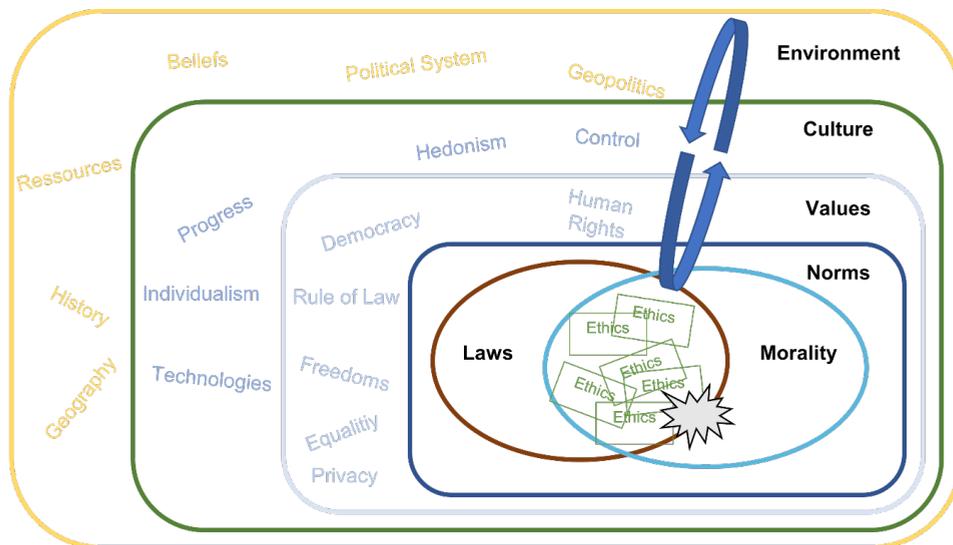
Figure 1: Ethics within the Bigger Picture.

As an illustration, recently, a French company developing and selling computer vision software and video analysis platform promoted solutions of intelligent video protection to ensure security during the Paris 2024 Summer Olympics. Taking the French context into account tends to soothe concerns regarding the use of such technologies. Yet, the company failed to anticipate the forthcoming presidential election. What if the next President sees this technology as a tool to monitor rich people or foreigners? What if France falls into the trap of extremism and populism? In a blink of an eye, the current perspective regarding computer vision and its potential drifts towards facial recognition would be recontextualized and become acceptable. Change the political system, and the ethical perspective will be impacted, potentially in a very wrong manner.

Conclusion

Ethics is not homogenous. It is not unique. It is plural. Failing to accept the reality of this diversity will inevitably lead to failure in terms of AI governance. Worse, it can lead to worsening existing tensions and creating new ones around cultural

imperialism and moral absolutism.

Disagreeing with others does not entail that they must be denied the right to have a divergent perspective, express it, and have it heard and considered.

Ethical appraisal is all about making decisions at a specific time for a specific situation. It needs to be contextualized, and cultural grounds are not only important, but they are also essential.

So far, the Western world has led the debate imposing its ethical standpoints based on its culture, its specific understanding of the world, and its dreams about a better world. Condemning biases, it has never questioned its own ones. It has developed a narrative and worked hard to make it adopted widely. This ethical proselytism is ethically questionable and must be questioned. Words can be weapons, and so can be norms, legal or ethical, based on biased perspectives. Words shape perceptions, contribute to the social construction of the world, and eventually influence behaviors. The ethical appraisal of AI must:

- Pay due respect to cultural diversity.

- Include a wider range of perspectives.

- Consider even those perspectives that the West does not agree with.

- Free itself from the constraints of speeches.

- Avoid the trap of cosm-ethics.

- Reject ideological stances such as universalism and relativism.

- Be contextualized.

Short of all these requirements, ethics applied to AI will be pointless, counterproductive, and even dangerous. It would become cultural tyranny.

References

[1]   UNESCO, "Recommendation on the Ethics of Artificial Intelligence," 2021.

[2]   A. L. Kroeber and C. Kluckhoh, Culture: a critical review of concepts and definitions, Papers ed., Peabody Museum of Archaeology & Ethnology, Harvard University, 1952, p. 181.

[3]   S. S. Schwartz, "Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries.," *Advances in Experimental Social Psychology,* vol. 25, pp. 1-65, 1992.

[4]   F. d. Saussure, Cours de liguistique générale, Paris: Payot, 1916.

[5]   E. R. Goffi, "Ethical Assessment of AI Cannot Ignore Cultural Pluralism: A Call for Broader Perspective on AI Ethics," *Arribat – International Journal of Human Rights,* vol. 1, no. 2, pp. 151-175, 2021.

[6]   E. R. Goffi, "Ethics of AI: Going Beyond Words," 3 March 2022. [Online]. Available: https://www.the-yuan.com/246/Ethics-of-AI-Going-Beyond-Words.html. [Accessed 18 March 2022].

[7]   G. Hofstede, Culture's consequences: Comparing values, behaviors, institutions and organizations across nations, Sage publications, 2001.

[8]   P. L. Berger and T. Luckmann, The Social Construction of Reality: A Treatise on the Sociology of Knowledge,, New York: Anchor Books, 1966.

[9]   M. Weber, *Science as a Vocation,* H. H. Gerth and C. W. Mills, Eds., New York: Oxford University Press, 1946, pp. 129-156.

[10]   E. R. Goffi, "Escaping the Western Cosm-Ethical Hegemony: The Importance of Cultural Diversity in the Ethical Assessment of Artificial Intelligence," *AI Ethics Journal,* vol. 2, no. 2, pp. 1-11, 2021.

[11]   A. Caliskan, A. Narayanan and J. J. Bryson, "Semantics derived automatically from language corpora contain human-like biases," *Science,* vol. 356, no. 6334, pp. 183-186, 2017.

The evolution of AI and its impact on Women

Narjis Hilale
Professor at the International University in Geneva, Author, Member of Morocco's Special Commission on the Development Model, Switzerland.

Basically, we need to take a look at how the picture looks like today and understand the reasons for the evolution of AI and its impact on women now also from a global standpoint on human rights and then see what we can do tomorrow to develop a comprehensive approach for human rights.

Siri, Cortana, Alexa: Most intelligent virtual assistants usually have a female voice. Why? Maybe because a female voice-operated software is easier to develop than a male one? Actually, the answer is no; on the contrary, it is more difficult to develop a feminine voice than a masculine one as they are usually higher-pitched, more complex and more variable, therefore more complex to manage.

So why are we using female voices for these intelligent virtual assistants?

The interesting reason is that we usually use female voices assistants to provide a service or listen, but when we like to provide services related to finances or real estate, we tend to use more masculine voices. There was a study [1] done by a professor at Stanford University, named Clifford Nass, which found that people react differently to a voice depending on gender: masculine voices tend to inspire more respect and expertise while female voices are usually more accepted when it comes to being social and showing empathy.

We can see therefore clearly that there is a gender bias in intelligent virtual assistants. This is just the beginning.

Even though women have pioneered the AI field, as the first coder in the 19th century was a woman named Ada Lovelace, discrimination still remains. What is happening today is that only 22% of women are working in Artificial Intelligence jobs.

According to UNESCO, only one-third of the world's global researchers working in AI are women [2].

In the STEM fields (Science - Technology - Engineering - Math), we see that women are still a minority. There is significantly more discrimination and imbalance in high-profile jobs within those fields.

What is furthermore problematic is that algorithms are biased.

Today, we see that if the data input into the algorithm is biased, it will automatically have a biased data output. Therefore, it is crucial to start thinking about how we could incorporate data that reflects a more equal world. Algorithms are strengthening the biases and reinforcing stereotypes, and in order to counter that, it is crucial to make sure to include women in the discussions so that they can give their point of view and be more inclusive.

The evolution of AI, this fourth industrial revolution, has been impacting women as well as human rights, amongst other things.

Kai-Fu Lee, a futurist and Artificial Intelligence expert, said that "if data is the new oil, then China is the new Saudi Arabia" [3] and I would like to push it further and say that "whoever owns and controls the data controls the world."

Looking at AI evolution cannot be done without looking at market capitalization's growth over the past three decades. Just ten years ago, the top ten most valuable businesses were oil companies, banks and fast-moving consumer goods companies (FMCG), but today we see the hegemony of what is referred to as GAMAM (Google, Apple, Meta, Amazon, and Microsoft). In the beginning of 2021, the combined market capitalization of these five companies, which are the 5 biggest US. Tech companies had reached USD 7.2 trillion[3]. By comparison, it accounts for the third-largest economic power in the world, after the United States and China, and ahead of Japan and Germany, establishing its hegemony.

---

3 Wikipedia contributors. (2021b, December 13). *Big Tech*. Wikipedia. https://en.wikipedia.org/wiki/Big_Tech

Why do we always, as humans, tend to underestimate the impact of technology?

In his Law of Accelerating Returns, Ray Kurzweil explained why we are underestimating the impact of technology. Technology is moving at an exponential rate: The green arrow (in the illustration below) starts very slowly but grows exponentially; this is how technology evolves. However, the human brain is used to a linear evolution (see grey arrow below).

Figure 1: Linear curve and exponential one

Peter Diamandis, Singularity University co-founder together with Ray Kurzweil, has used an excellent metaphor to describe the difference between linear and exponential evolution: If you take 30 steps of one meter each linearly (1-2-3-4-…-30), by the end of the 30 steps, you would have walked 30 meters. However, if you take the same 30 steps exponentially, you would double your steps every time (1-2-4-8-16-32, ...), which is the equivalent of walking 26 times around the earth [4]. Then you can imagine the difference between 30 meters versus 26 times around the earth, that is the discrepancy and gap existing between linear and exponential growth.

"We always overestimate the change that will occur in the next 2 years and underestimate the change that will occur in the next 10. Do not let yourself [be] lulled into inaction." [5].

Why do we tend to overestimate the change that is going to happen in the first years?

If we go back to the grey arrow, we can see that the linear growth is higher at the beginning in comparison to the green arrow. This makes us overestimate change in the first years and underestimate it in the next ten years, and we should not let ourselves, as Bill Gates said, " [be] lulled into inaction."

The situation we have today is that we are lagging several steps behind, specifically when it comes to regulations. We are constantly playing catch-up, in a reactive mode: just responding to the technological advancements and developing initiatives that are already obsolete by the time they are launched.

The thing is that this race will never be fair because the more you run, the more the first contestant (AI technology) is running much faster while we (humans, regulations) are running into hurdles and trying to overcome them, to fix them along the way. This is the current state of what we are facing with AI and regulations.

What can we do tomorrow in order to develop a comprehensive approach to human rights?

Something we can do is to think "pluridimensionally" as this has several dimensions, and there is no one solution that fixes all but different angles to be taken into consideration.

The first thing is definitely a shift of the mindset; we need to think exponentially - since AI is evolving exponentially, then we need to do the same - and even think ahead and take bold steps forward with our initiatives. The initiatives that we have today are not proactive, we should not act in reaction to what technology does, but we need to be thinking 2, 5, 10, 15 years in advance in order to 'play fairly'.
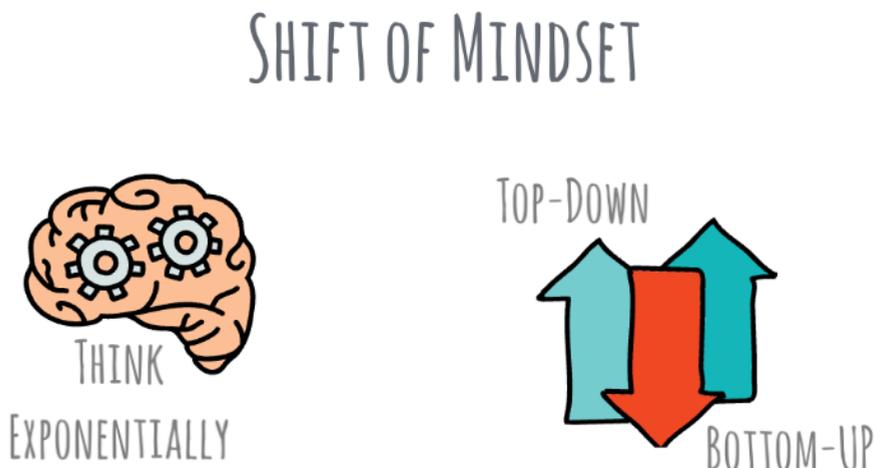
Shift of Mindset

Figure 2: Shift of mindset.

The other element that is also very important is that we have used a lot of the top-down approach with big institutions, the United Nations, the private sector, governments, ... coming up with initiatives and rolling them out. It is time to consider the power of the bottom-up approach by ensuring that citizens and civil society are involved in working towards solutions.

How can we achieve that?

When talking about Artificial Intelligence today to end-users (e.g., citizens, civil society), we are mainly evaluating the pros: how it is positively changing our lives, from the internet of things to connected homes, to machine learning, to algorithms (that learn and influence our preferences) to virtual reality (e.g., Meta launching its virtual reality environment). This makes us realize that AI is basically becoming an extension of ourselves. Therefore, it is very important, however, that we, as citizens, we start understanding as end-users the cons as well, understanding that we are giving our data that are feeding AI for free, freely given by the end-users without all of them being fully aware of the high price they pay in the counterpart of using these services.

This is really about opening the debate so that all the stakeholders and end-users are fully aware of all the facets of AI technology.

Human-Centric

Figure 3: Technology development must be human-centric

Another element is, of course, the human rights component and making AI human-centric. Not only do we need to raise awareness in this regard, but also prepare humans for this digital revolution, for them to be able to ride this fourth industrial wave. We also need to enhance our soft skills in order to empower human beings to face this digital revolution. Nowadays, things are starting to change, with institutions and academics talking more and more about soft skills, although hard skills remain at the core of any job, and 90% of work resumes focus on hard skills. Soft skills are still under-evaluated, while they really are what differentiates us, humans, from the technologies and the machines, so we must praise them and look at the way institutions, academics, the private sector, governments, work environments … are approaching these soft skills so that they are praised and better considered more to enable and empower humans to be fully prepared for the digital revolution.

Women being more affected by multiple forms of discrimination and violence, and

this is another element concerns demonstrated by the numbers on the rise since the beginning of the pandemic. Although women represent 50% of the population, they are still considered a minority. They are disproportionately affected by the impacts of AI.

What is interesting to note is that moving forward, we might not just talk about women and men, but we are going to be talking about humans versus technology so all of us will all be in the same situation, facing the same discriminations. Therefore, it is extremely important that we think human-centric and put the human at the center of all our initiatives.
Every transformation requires a change of mindset: we have to start thinking like AI. We cannot protect human rights [6] if we are always lagging behind and not thinking ahead [7].

References

[1]   C. Nass and L. Gong, "Speech interfaces from an evolutionary perspective," *Communications of the ACM 43,* vol. 9, pp. 36-43, 2000.

[2]   J. Lewis, S. Schneegans and T. Straz, "UNESCO Science Report: The race against time for smarter development. Vol. 2021," UNESCO Publishing, 2021.

[3]   K.-F. Lee, Interviewee, [Interview]. 2018.

[4]   P. H. Diamandis, What Does Exponential Growth Feel Like?, 2017.

[5]   B. Gates, N. Myhrvold, P. Rinearson and D. Domonkos, The road ahead, 1995.

[6]   M. Bensalah, "Toward an ethical code of AI and human rights in Morocco," *Arribat - International Journal of Human Rights Published by CNDH Morocco,* vol. 1, no. 2, pp. 187-203, 2021.

[7]   N. Hilale, "The Evolution of Artificial Intelligence (AI) and its impact on Women: how it nurtures discriminations towards women and strengthens gender inequality," *Arribat - International Journal of Human Rights Published by CNDH Morocco,* vol. 1, no. 2, pp. 141-150, 2021.

Technology, the Environment, and Ethics

Dr. Workineh Kelbessa
Professor of Philosophy at Addis Ababa University, Ethiopia.

I will start by summarizing my presentation that addresses the issues of guiding ethical principles as well as the positive and negative effects of technology through examples related to biotechnology and intellectual property rights, technology and social life, information and communication technologies and E-waste. The main objective of this presentation was to examine the impact of technology on society and the environment and the ethical problems and issues associated with it.

Concerning the positive effects of technology, I point out that human beings have used different forms of technology to modify, subdue, control, and engineer nature. Technology has enabled humanity to improve production and human welfare, control different diseases, develop better communications, generate electricity and address some of the most pressing environmental challenges.

To illustrate this, we can see examples such as the successes of high-tech agriculture. According to Lester Brown [1], "technological advances have tripled the productivity of world cropland during this century [twentieth century]. They have helped expand the world grain harvest from less than 400 million tons in 1900 to nearly 1.9 billion tons in 1998. Again, agricultural productivity has also enabled different groups of people in different parts of the world to reduce global poverty and hunger."

We can also mention the example of the advances in genetic knowledge and technology. Thus, according to W. Brock [2], "we face the prospect of being able to take control over and to design human nature and the nature of our progeny. What was once in the hands of God or the natural lottery will come to be increasingly within deliberative human choice and control".

Regarding the role of reproductive technology, Rolston [3], an American philosopher, claims that "Humans have proved capable of advanced skills never dreamed of in our ancient past: flying jet planes, building the internet, decoding their own genome, and designating world biosphere reserves."

It is also important to mention the role of the internet in modern life that has changed the way we live before addressing the question of the role of AI in shaping the future. According to Risse [4],

""Intelligence" is the ability to make predictions about the future and solve complex tasks. AI is such an ability demonstrated by machines, in smartphones, tablets, laptops, drones, self-operating vehicles, or robots. Such devices might take on tasks ranging from household support and companionship (including sexual companionship), to policing and warfare."

AI can contribute to meeting some elements of sustainable development goals; it has contributed to economic growth, poverty reduction and health promotion, and the realization of human rights. We can also mention the application of AI in the fields of agriculture, energy, health care, public services, financial services, wild animal management, and smart cities. In particular, some African countries and other developed countries have benefited much from AI in different fields.

AI can also detect health problems and avoid health and safety risks. To give an example, robots such as South Korea's DRC-Hubo helped mine workers to avoid health and safety risks, which is very significant. Furthermore, AI can detect financial crime and money laundering. According to Ridley (2001), technological innovation rather than treaties, global energy policies, or consumer restraint can fix various global problems, including climate change.

However, technological advancement and application have incorporated both creative and destructive forces. Amongst the negative effects of technology, I would like to mention that modern technology has contributed to pollution, resource depletion, and biodiversity loss because of specialization and a host of other problems. Modern technology has also led to different types of environmental destruction, one can mention the example of the "Three Mile Island" accident in 1979 at the "Three Mile Island" nuclear power station in the U.S.A, the Bhopal chemical disaster in 1984, in India, the explosion of the U.S.A space shuttle orbiter Challenger in 1986, the nuclear accident at the Chernobyl nuclear power station in the former Soviet Union in 1986, the Exxon Valdez oil spill in 1989, the breakup of the U.S. space shuttle orbiter Columbia in 2003, and the Fukushima Daiichi nuclear disaster at the

Fukushima Daiichi ("Number One") plant in northern Japan in 2011.

Despite its advantages, the Internet has created dramatic vulnerabilities that have been affecting peoples' lives: crime, cyberbullying, pornography, sextortion, internet predation, the proliferation of fake news, lack of privacy, decreased safety and security, harm to social relations and communities, and so forth.

Another important finding is that technology reinforces inequality. In particular, the production of AI has the tendency to maintain inequality in different parts of the world. AI is actually strengthening power differences in the world. It has widened the technological gap between the rich and the poor.

Moreover, technology can also be misused by bad actors, companies, rogue states, using AI surveillance technology to dominate their own citizens, and at the same time, terrorist organizations can also misuse the AI to violate human rights. Kriebitz and Lütge considered this as "Malicious AI." [5] AI can also lead to unemployment, job losses, and unjustifiable and morally repugnant military attacks; some countries or groups can use the drone to violate privacy and at the same time security.

In addition, incorrect programming or training of AI can lead to unintended violations of human rights. When we look at the autonomous and semi-autonomous weapon systems, we find that they won't be in compliance with international humanitarian law, as they cannot make a distinction between combatants and civilians, which is another problem related to the use of autonomous and semi-autonomous weapon systems.

AI can also lead to environmental degradation, for example, overuse of RMs, use to produce AI, as well as high-energy consumption, can lead to an increased carbon footprint that contributes to climate change.

Even technologies that were created to promote wellbeing can destroy sustainable indigenous life. If we look at the impact of the Green Revolution, in the 1940s, 50s, and 60s, we can see that it enabled some Latin-American countries, India, including some African countries to get out of poverty, however it has also led to biodegradation, biodiversity loss, and other related problems. The indiscriminate application of Western technologies, particularly in developing countries, has led

to the destruction of cultural heritage, because modern values, particularly in developing countries, tend to destroy their own values and embrace western values. So, this has had a negative impact on the development of different countries.

With regard to the negative impacts of biotechnology on the environment, biotechnology can permanently destroy some parts of the natural environment, amongst others: genetically engineered organisms have led to 'biological pollution'. Biological pollution has a tremendous impact on the environment, for example, it can overrun indigenous crops, through cross-breeding. The modern technology system simultaneously creates and destroys values. So, what is the way out of this problem? In particular, in developing countries, we can talk about intermediate-scale technologies or appropriate technology. Developing countries are required to choose what we call the appropriate technology. Since there is no one unique appropriate technology, different governments ought to test different forms of technology and adopt technologies that are useful to their purposes.

I also would like to mention the role of biotechnology and intellectual property rights. Intellectual property rights are intended to protect from theft of various intangible products of human intellect. Intellectual property is intangible personal or corporate property and has economic benefits for the holders of property rights, which can include copyrights, patents, and trademarks.

Originally, the major purpose of the patent system was to meet industrial needs. In the beginning, the patent system did not involve agriculture and living things, but later, developed countries introduced intellectual property rights to protect agriculture and at the same time living things.

A patent is granted for an invention that is novel (recent and original), involves an inventive step, and is capable of industrial application. Because of these criteria, those who invented the concept believe that Indigenous environment knowledge cannot be patented because it doesn't have itself an independent inventor, because it is invented by a collective corporation of different groups, therefore, Indigenous environmental knowledge was excluded from the sphere of patent law.

More recently, intellectual property rights (IPRs) have allowed patents on living

organisms, including the knowledge fostered by indigenous people in different parts of the world. Therefore, biotechnology brings the corporate patenting of new life forms and the voracious global search for indigenous seeds and plants to patent and market, with devastating effects on developing countries' agriculture, ecology, and human rights. For example, farmers are not allowed to use seeds for subsequent years unless they pay royalty rights. Plant breeders use new genetic technology to prevent farmers from saving seeds to replant them in the future. It has had devastating impacts on food security, and agricultural biodiversity. "Because it is a potentially 'lethal' technology, Rural Advancement Foundation International (Canada) (RAFI) has dubbed it the 'Terminator Technology' [6]. "

I will also talk about technology and social life. Technology can undermine the ability of technology-dependent individuals to think and act outside of technology. This happens because the individuals who completely rely on technology are detached from the natural world, which puts our nations in front of a "Nature deficit disorder". This is also related to what Tavani [7]   called "ambient intelligence (or AmI) -a technology that enables people to live and work in environments that respond to them in 'intelligent ways'."

I want to address the issue of information and communication technologies and electronic waste (e-waste). Despite the fact that technology is very useful, many technological processes use up natural resources and produce unwanted by-products that pollute the environment. For instance, particularly in developing countries, it is very expensive to clean electronic waste (e-waste). Some countries and corporations decided to export their e-waste to Asia, Africa, and South America in the name of "aid" and second-hand electronics, however, the deal was not genuine and ethically defensible. In developing countries, it will be very difficult to get rid of this waste.

Then I move on to the issue of the guiding ethical principles. Technology alone cannot solve all problems. Geoengineering's promise of an ultimate technofix is far from addressing the problems humanity faces. According to Rolston, "human beings should bring themselves under control and reduce their ambition for endless growth and wealth, and learn to manage themselves as much as the planet." Therefore, we

need to examine the foreseeable consequences of various technologies for human beings, animals, and the natural world.

A number of writers have tried to formulate different normative and ethical principles to manage AI and other technologies. Mark Ryan and Bernd Carsten Stahl [8] identified 11 normative principles for developers and users of Artificial Intelligence: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity, and solidarity. A range of stakeholders including policymakers, users, and developers but also educators, civil society organizations, industry associations, professional bodies, can use AI ethics guidelines.

The General Conference of UNESCO adopted the recommendation on the ethics of AI at its 41st session on November 24th, 2021. It has identified the following ethical principles that can guide the development and application of AI technologies: proportionality and do no harm, safety and security, fairness and non-discrimination, sustainability, right to privacy and data protection, human oversight and determination, transparency and explainability, responsibility and accountability and awareness and literacy.

I also emphasize with further guiding ethical principles, amongst others, the precautionary principle that states that although there is no scientific consensus when there is reason to expect possible disasters, preventive measures should be taken to avoid them. So "this principle should be applied to anti-technology campaigns as well as to technological innovations since the former as well as the latter could put people at risk of unnecessary harm." [9] According to Attfield, engineering students are required to acquaint themselves with "a clear view of the ethical aims, principles, and constraints relevant to their work as professionals and as citizens", and all engineers should study the ethics of climate change to address ethical issues related to technological risks and uncertainties.

Others reject the role of the precautionary principle for opposing innovation: it is based on fear, and unpleasant remarks "rather than sound science." Although I do not agree with this claim, even if sometimes the precautionary principle may have negative or positive impacts, what matters is that it depends on how much suffering

would result from applying it.

While quoting UNESCO that emphasizes the need for social justice, I insist that decent social justice implies environmental justice, which is very important. Therefore, environmental justice requires other individuals to distribute environmental benefits and risks… in a fair manner, that opposes technological apartheid. Respect for people is equally important; autonomous people should make their own decision about the use of technology and its impact on their health. Informed consent and confidentiality are based on these principles.

To conclude, Technology cannot only benefit but harm us as well. Some countries and transnational corporations have continued to develop new technologies without taking their long-term implications into account. Primarily, they have been motivated by unlimited profit. According to Rolston, "Technology, coupled with capitalism, drives people, rich and poor, ever to want more, more, more, with increasing power to get it. Human nature continuing into the Anthropocene Epoch, Pleistocene appetites or not, can at once offer promise of success and simultaneously escalate the threat of our undoing. For the first time in history, the future of earth is at stake". All concerned stakeholders should oppose this move and try to make technology more humane and environmentally sustainable. His paper suggests that we should try to reduce the impact of technology on the environment. Some ethical principles, including justice, should be in place in the distribution of the fruits of technology.

- Ethics can help humanity to reduce or avoid the negative impacts of technology on humans and the environment;

- The introduction of new technologies requires strict codes of ethics and responsibilities towards human beings, non-human beings, and Mother Earth;

- The new form of corporate social responsibility and accountability should be in place to minimize the negative effects of technology;

- Human beings should not be removed from AI and AI systems should not be given full responsibility;

- Unlike human beings, machines cannot control their actions. Only humans can make adequate judgments and avoid the risk of accidents.

I finally conclude by saying that African worldviews do not support the replacement of human beings with lifeless machines. Therefore, I suggest that AI should not be detached from human beings. We should retain human judgment in charge of AI systems. The right balance between humanness and technology should be in place.

References

[1]   L. R. Brown, "Feeding Nine Billion," in *State of the World*, C. F. H. F. F. L. S. Lester R. Brown, Ed., W. W. Norton, 1999, p. 119.

[2]   D. W. Brock, "Genetic Engineering," in *A Companion to Applied Ethics*, R. G. F. &. C. H. Wellman, Ed., Blackwell's, 2003, p. 357.

[3]   H. Rolston, "Technology and/or Nature: Denatured/Renatured/Engineered/Artifacted life?," *Ethics and the Environment,* vol. 22, no. 1, p. 52, 2017.

[4]   M. Risse, "Human Rights and Artificial Intelligence: an urgently needed agenda," *Human Rights Quarterly,* vol. 41, no. 1, p. 2, 2019.

[5]   C. L. A. Kriebitz, "Artificial Intelligence and Human Rights: a Business Ethical Assessment," *Business and Human Rights Journal,* vol. 5, no. 1, pp. 84-104, 2020.

[6]   Rural Advancement Foundation International (RAFI), *RAFI Communiqué: The Terminator Technology — New Genetic Technology Aims to Prevent Farmers from Saving Seeds,* March/April 1997.

[7]   H. T. Tavani, Ethics and Technology: Controversies, Questions and Strategies for Ethical Computing, 4th ed., John Wiley & Sons, Inc., 2013, p. 370.

[8]   B. C. S. Mark Ryan, "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications," *Journal of Information, Communication and Ethics in Society,* 2020.

[9]   R. Attfield, "Engineering Ethics, Global Climate Change and The Precautionary Principle," in *Contemporary Ethical Issues in Engineering*, S. S. Sundar, Ed., IGI Global, 2015, pp. 38-47.

Disinformation 2.0: protecting the integrity of the information eco-system in a technologized world

Dr. Alina Bârgăoanu
Professor, Dean of the College of Communication and Public Relations at the National University of Political Studies and Public Administration

The topic of disinformation 2.0 is both related to artificial intelligence and ethics, even if collateral confirms that technology is a two-edged sword. At the core of this new phenomenon of disinformation 2.0 is technology.

Disinformation 2.0 is a new phenomenon of interest when discussing AI and machine learning concerns. We are talking about a new phenomenon because it occurs under conditions of information over-abundance, not information scarcity as used to be the case of traditional propaganda during World War I, World War II, or the Cold War, for example. The next argument to support the idea that this is a new phenomenon is that it is technology-driven; it is co-substantial with the Internet, the explosion of social media, and increasingly with the possibilities to fake the amplification by using machine learning and Artificial Intelligence. Technology, big data, and AI create possibilities to go straight to the people so as to spread it directly; hence the human mind becomes the battlefield. Different terms underline this kind of reverse propaganda that goes straight to the people with no intermediaries and no human gatekeeping; sometimes, it is called peer-to-peer propaganda, bottom-up propaganda, ampliganda, or crowdsourced propaganda.

The theoretical and philosophical question for me is, and I am still working on the answer, whether technology is a driver or an amplifier?

Coming back to some conceptual notes:
Disinformation 2.0 differs from other forms of traditional propaganda or disinformation in that it is a 2-fold phenomenon as it involves information manipulation and blurs the distinction between what is factual and not factual but at the same time – and this is definitely fed by technology, provides the opportunity

for manipulating the engagement meaning the manipulation of the social component of social media.

This is very different than what we have witnessed before the digital revolution as contemporary propaganda "ampliganda" or disinformation 2.0 manipulates information, the social content, the context, and the engagement.

With disinformation 2.0, the content is somehow secondary and what matters most in my understanding is how this kind of content circulates, the way it is massively deployed, technologically driven by platforms of big data and Artificial Intelligence. In my public statements, I try to make very vigorously the point that to understand what is going on in liberal democracies, in the western world, in the European Union, and outside these borders, we need to understand that it is platform- and big data- driven and increasingly driven by machine learning and Artificial Intelligence.

Some examples of propaganda 2.0, disinformation 2.0 do not dwell on the binary classification of "true or false" – this would have made our lives very easy, but can encompass all mentioned below:

- "True, false, and anything in between";

- Non-factual content that is plausible;

- Non-factual content that is realistic;

Here, we should underline the possibilities to come up with deep fake, fake writing and the possibility of fake engagement;

- It can cover non-factual but political/cultural/ identity-based content;

- It can rely on factual but hyper-partisan content, sensationalized, clickbait-y;

- Factual but misleading (framing, out-of-context, cherry-picked);

- Factual content but algorithmically amplified by means of AI, big data, and machine learning.

Hence another theoretical and conceptual question: is computational/ algorithmic truth still truth?

- It can also dwell on things that are beyond factual such as satire, visual discourse, memes, etc. This kind of content that is already very complex can be amplified by artificial means by using big data, machine learning, and AI.

This new phenomenon of disinformation [1] relies on different tactics and objectives compared to traditional propaganda. Contemporary propaganda/disinformation 2.0 seeks to weaponize every major controversy in a society, what is called "hot button issues," and to amplify positions on both sides of the debate – it is a popular myth and misconception that propaganda only amplifies one side according to my understanding and to the technological root of contemporary propaganda, it seeks to amplify both sides of a debate.

Hence, the objective might not necessarily be to hijack a candidate or a person or a political party but rather to hijack the public conversation altogether and to alter the terms of the public conversation so as to polarize, disrupt and strategically shape public discourse to kill the very possibility of civil discourse. These statements are inspired by my research on the western culture and socio-political space, but I would say that this can also be used to understand what is happening at a global level.

In terms of effects, contemporary propaganda, compared to the traditional one, is different from anything that we have seen so far because it creates within a society a specific type of polarization which is cognitive and emotional at the same time and creates a mindset of all or nothing, us vs. them, irrespective of specific topics.

It cannot be emphasized enough that this specific type of polarization has a technological root and is enabled by technology by the use of microtargeting, targeting users with hyper-personalized content utilizing big data in order to understand their preferences and digital fingerprint to feed them only content that they already appreciate. Disinformation 2.0 creates, as a result, a pan-ideology of exclusion, a feeling that it is us vs. them, all or nothing, and that we cannot live together in the same society.

"We lose our ability to distinguish the real from the unreal, the actual from the imagined, or the threat from the conspiracy. The powers working to disrupt democracies through memetic [cognitive] warfare understand this well. Contrary to popular accounts, they invest in propaganda from all sides of the political spectrum. The particular narratives they propagate through social media are less important than the immune reactions they hope to provoke." (Rushkoff, Team Human, 2019)

So, this is the biggest effect that can be felt around the western world and liberal democracies. Cognitive warfare dwells in provoking reactions and the viralisation of reactions, irrespective of the nature of these reactions (rejection, approval, outrage). Contemporary propaganda with a strong technological root provides a megaphone to both sides of the debate and can lead to the implosion of the middle ground.

Here are some examples from the Romanian public space in the context of the vaccination campaign, you can see below some very violent covid-19 nazi analogies:

Figure 1 : violent covid-19 nazi analogies.

With strong visuals comparing the green pass to a new Auschwitz (Holocaust), the vaccine compared to a totalitarian dictatorship and terror, hospitals compared to concentration camps: the big narrative is that we are dealing with medical nazism. This kind of content cannot be judged in terms of factual or non-factual because it has a lot of visual components, a lot of framing, and is very brief. I have chosen examples that are heavily amplified by bots, trolls, click factories, and fake followers, which seek to create a reaction, and amplification is crucial.

Here is another example of strong visuals using nazi symbols to discredit the pro-vaccination messages:
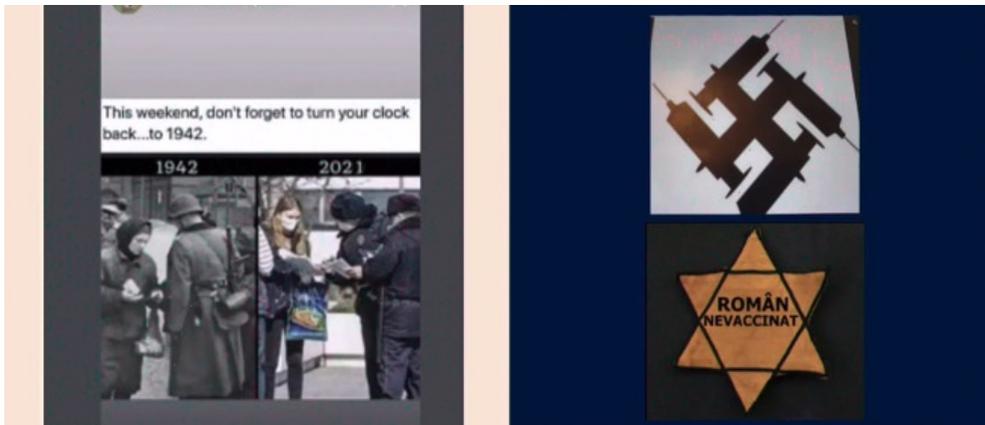
Figure 2 : using nazi symbols to discredit the pro-vaccination messages.

In Romania this kind of framing is transitioning slowly but surely towards geopolitical aims and what was inconceivable before the pandemic – to talk about Romania exiting the European Union, is now brought to the mainstream by means of strong visuals, sketchy content, and by amplification mechanisms with trolls, click factories, clickbait content, automated advertising, etc.

Figure 3 : Desinformation and propaganda about 'RoExit'.

Stanford professor Renee DiResta who coined the term "ampliganda" said that in order to tackle disinformation, "We have to move away from treating this as a problem of giving people better facts, or stopping some Russian bots, and move towards thinking about it as an ongoing battle for the integrity of our information infrastructure – easily as critical as the integrity of our financial markets." (DiResta, The digital Maginot line, 2018)

I think that there are short-term solutions that can be implemented. The first is to invest a lot in literacy and education but not in terms of wide literacy rather, invest in very specific literacy addressed to high-level decision-makers such as digital literacy, technological literacy, algorithmic literacy, data literacy, and AI literacy.
Long-term solutions to avoid political fragmentation, intolerance, radicalization, and extremism are to address the internal vulnerabilities of societies or what can be called the "kernel of truth."

In order to solve this kind of "information disorder," we should pay attention to tackling inequality and sources of popularist thinking and resentment, investing in good media and information systems, including public broadcasters, investing in inclusive leadership, and pay attention to good governance and the capacity of contemporary government systems to provide public goods.
My concluding point is that we should pay attention to the issue of ethics, including how Artificial Intelligence, big data, and machine learning as technological possibilities are used to amplify this kind of political fragmentation.

References

[1]   A. B. R. B. a. O. Ş. Nicoleta Corbu, "Does fake news lead to more engaging effects on social media? Evidence from Romania," *Communications ,* vol. 45, no. s1, pp. 694-717, 2020.

[2]   D. Rushkoff, Team Human, W. W. Norton & Company, 2019.

[3]   R. DiResta, «The digital Maginot line,» 2018.

- III -
EMERGING STANDARDS

COMEST/UNESCO: a comprehensive global standard-setting instrument to provide AI with a strong ethical basis.

Dr. Peter-Paul Verbeek
Chairperson of the UNESCO World Commission for the Ethics of Science and Technology (COMEST), Professor at Twente University, the Netherlands.

There are two lines of work done at the Commission on the Ethics of Scientific Knowledge and Technology (COMEST) that I would like to discuss. The first one is the COMEST concept note on the ethics of science and technology [4] that we have been writing as a basis for asking the question if it would indeed be able to proceed as a global normative instrument for Artificial Intelligence. The second one is the recommendation[5] on the ethics of science and technology that was adopted at the UNESCO General Conference on November 24th, 2021.

It is essential to explain why it is such an important theme for UNESCO: AI is about thinking; it is about the mind and how we understand the world. Artificial Intelligence is about technology that helps us make sense of the world around us, one could say, which makes it of interest to UNESCO as it concerns our thinking, science, education, culture, and communication.

---

[4] UNESCO, Concept note of COMEST on the ethics of science in society: lessons from the COVID-19 pandemic?, 2021 https://unesdoc.unesco.org/ark:/48223/pf0000379990
[5] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021 https://unesdoc.unesco.org/ark:/48223/pf0000380455
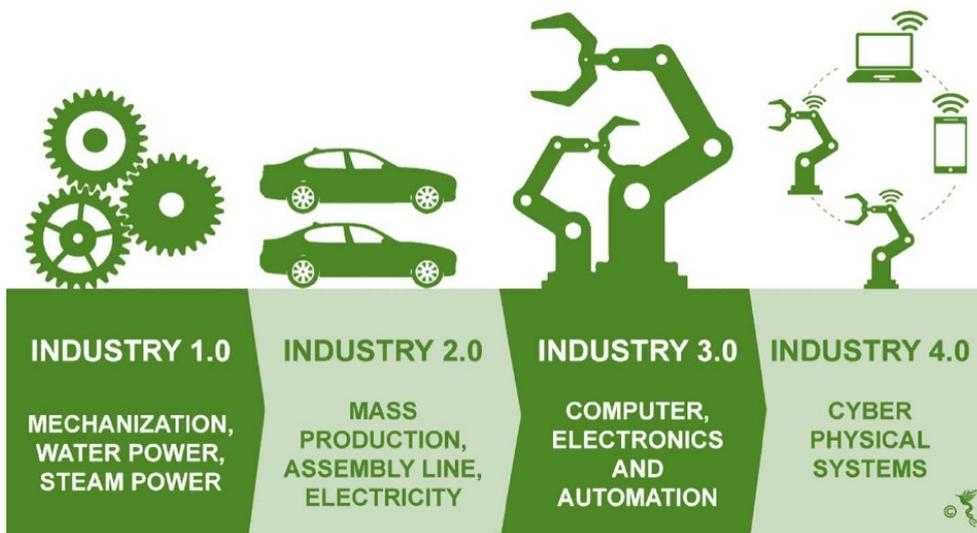
Figure 1: Industrial revolutions in human history

It is also a technology that we wanted to study as COMEST as it is part of the digital revolution, also called the fourth industrial revolution, where after the mechanization of labor came the second industrial revolution with mass production. The third one is the information revolution, and the fourth one brought the cyber-physical systems, where the internet becomes an internet of things, AI is embodied in robots, and where digital technologies have become a part of our society. Some even say that we have entered what one could call society 5.0, where this digital revolution takes us into a completely new world. This requires thinking about it in a profound way from an intercultural point of view, which is very relevant, and that is why UNESCO embarked on this journey to keep thinking about Artificial Intelligence and see it as a key element of UNESCO's work.
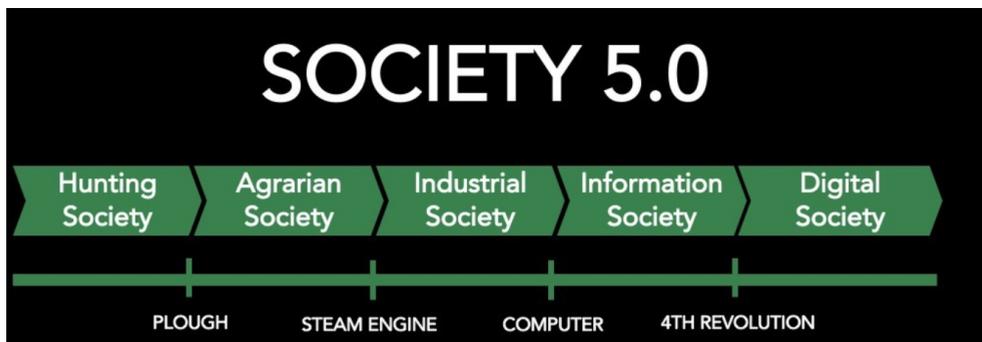
Figure 2: Society 5.0, or digital society.

The first step of this work consisted of the COMEST Extended Working Group on Ethics of Artificial Intelligence writing a preliminary study on the ethical questions regarding AI.

Why UNESCO? As previously mentioned, it is because education, science, culture, and communication are the key elements of UNESCO. The way in which we have come to do ethics of science and technology over the past years is an engaged form of ethics. We want to be critical from within and engage with the technology, profoundly understanding what is at stake not only on the external assessment, as it were, but also by giving ethical feedback and complementing the design, use, and implementation of the technologies. The intercultural perspectives are also critical; we should make sure that nobody is left behind. Leave no one behind has always been the ambition of UNESCO, and that is also what we have tried to do in our work.

Thus, we first tried to analyze the impact of AI on the key domains of UNESCO, for instance, the realm of education. What is the societal role of education? Is the societal role of education changing now that we are actually facing a world of Artificial Intelligence? What will happen to the labor market? How can we educate people to make sure that they are prepared for a new labor market? Also, how can we make sure that in education, we foster other things like literacy, digital literacy,

and critical thinking that people are not equipped with a critical approach to what AI brings to them so that they do not just take for granted what they hear about AI systems, but develop a critical perspective.

It is also vital that we educate engineers about ethics. We need to educate engineers that ethics is not just an idea in human minds but also in the machines and that a lot of AI can be designed from this ethical point of view and that we need to equip engineers as well with this capacity to do that.

The second realm of UNESCO is science, and questions pop up like what does AI do to scientific explanation? What does this new wave dealing with large data sets mean for how we make combinations between correlations and causation? What does it do? Also, particularly through the social sciences, where data sets have got to play an enormous role, how can we think more deeply about decision-making based on AI? How does AI help medical doctors to arrive at a diagnosis or judges reach a verdict, or politicians make decisions? How can we then make sure that this decision-making process is supported by scientific facts in such a way that AI will not disrupt the solid scientific bases that we often need to make decisions?

Then, there is culture, the realm of culture, where AI plays an enormous role. For instance, in the realm of language, what will automated phone translation do to the more minor languages? Can one translate anything? Will some subtle details get lost if we do that? If so, how can we make sure that will not happen? Artistic practice, what will the forms of creativity seen in AI systems do to our ideas of creativity? What about intellectual property rights? Who owns the property, who owns the artworks? Furthermore, on cultural diversity in AI, we must design in such a manner that fosters cultural diversity rather than be an obstacle. To make this very mundane, could we design the AI systems in platforms like Spotify or Netflix that people use in specific regions of the world to not lock people up in their own taste, in their own cultural preferences, but also open them up for other genres of culture, to educate them interculturally.

There is then the realm of communication, where momentous elements like disinformation and journalism play a role. How can we deal with the growing importance of disinformation online? How can we equip people and AI systems to detect it? Also, how can we ensure the safety of journalists and the existence of journalism in a world that needs independent journalism that algorithms cannot replace?

We also discuss some global issues in our work, for instance, peace. How can AI be given a responsible role in conflict? What will it do to the character of a given conflict? What will cyberwarfare mean, and how can we ensure it is kept out of the systems? How can we prevent dual-use, for instance? Moreover, how can we make sure that AI systems contribute to responsible decision-making by world leaders in situations of conflict?

The specific role of Africa has always been present in the work of UNESCO. How can Africa be profoundly involved in the agenda-setting and development of AI systems? How can we make sure that attention is paid to the role of African women specifically? How can AI play a role in educational settings, especially in African countries?

Gender matters ended up being quite a substantial theme in the recommendation adopted at the 41st session of UNESCO's General Conference, which is now this worldwide normative instrument to deal with member states, a guideline for all the member states of UNESCO (UNESCO, 2021). Gender equality is all about challenging the bias that can be implicit in systems, and equality is also needed in AI engineering, so how can we make sure that data sets with which AI systems work do not have a gender bias? Furthermore, how can we make sure that we have a fair representation of women in the development of AI systems?

We also have the environment and the environmental sciences that can benefit from AI systems; how can we use the latter to get a better idea of, for example, how climate systems and rare-earth elements develop and the accuracy needed for it. Finally, how can we deal with AI systems' energy and climate impact?

Although the disaster and risk management field seems entirely distant from AI, it is also concerned with it. How can we anticipate risks in nature in a good way with AI systems? How can we design our AI systems to deal with that in a responsible manner?

All of this resulted in a first list of ethical principles, and we believe that a number of key principles should play a role in the design, implementation, and use of AI systems.

- Inclusiveness, leaving no one behind.

- Flourishing; AI should not only be designed from something negative one wants to keep out but also something positive that one wants to contribute to – AI for good.

- Awareness, literacy, and explainability are essential. People should be ready to understand what AI is giving to them in order to maintain their freedom and responsibility. That means that AI should be explainable; AI systems should be able to explain how they arrived at their conclusions. The data sets with which they were trained need to be transparent so that we can check how the AI system learned as it were to do what it is doing, which enhances the responsibility of people who need to make choices based on what AI is doing but it also enhances the responsibility of developing AI systems.

- Accountability is one layer further than responsibility. Therefore, we should arrange a way in which people can be held accountable, to be clear on who is to be held accountable for the decisions made by AI systems.

- Good governance. Governments should provide regular reports about how they use AI in policing, intelligence and security. All of these things end up in the recommendation that was adopted.

- Democracy is the last thing to be mentioned, as it would not work without the other principles

- Autonomy. Of course, AI should respect human autonomy. I also mention this last because we want to avoid it as the only thing to be considered important. In many cultural frameworks worldwide, autonomy is not the only principle from which to worry.

On the basis of the work we did on AI ethics a few years ago, there was an Ad Hoc expert group formed, with experts from all over the world, that has been writing a draft recommendation on the ethics of Artificial Intelligence, which took about a year, back in 2020. After that, the text went to all member states and was negotiated further, and ultimately it was adopted in late November 2021.

Let us walk through the key elements of the recommendation, not to explain them, but a summary of what is listed on the UNESCO website as the core of the recommendation will be mentioned. There are values set apart from basic guidelines and principles in the recommendation.

Four key values should guide our thinking about AI the recommendation sets: respect, protection, and promotion of human rights, fundamental freedoms, and human dignity -crucial for the seminar- then the environment and the flourishing of ecosystems which is an element that one does not very often see in a lot of existing frameworks. Ensuring diversity and inclusiveness is very prominent for UNESCO; we want to have a global inclusive framework and live in peaceful, just, and interconnected societies.

So, I would say this is a quite specific UNESCO point of view. There is a number of principles in the recommendation that are extensively worked on:

- Proportionality and Do No Harm;

- Safety and security ;

- Fairness ;

- Sustainability ;

- Privacy and data protection ;

- Human oversight ;

- Transparency and explainability ;

- Responsibility and accountability ;

- Awareness and literacy ;

- Multi-stakeholder approach.

It is the principles that one often sees in AI frameworks, but the characteristic of what UNESCO has been doing is to give them a global intercultural character with a focus on sustainability and gender [2].

Perhaps the most crucial thing in our recommendation is that it ends with quite an extensive section on policy action. I think the most consequential observation of the recommendation is that we have a lot of ethical frameworks for AI and that we should actually move from talking to doing. It is not only about understanding what ethical basics are but also about doing ethics and implementing ethics in practice. That is why the recommendation has quite an extensive list of areas where policy is needed; we need an ethical impact assessment. We need to be able to assess what AI is doing in society. We need governance, stewardship, and policies to deal with data in an ethical way. Also, we need to keep a sharp eye on development, international cooperation, the environment, and ecosystems are substantial areas to work on, and all kinds of ideas are discussed in the recommendation regarding the areas mentioned above. In addition to gender, culture, education, communication, economy, the labor market, health, and social well-being, all the elements at UNESCO.

The key message can be summarized in four elements that are also listed on UNESCO's website.

- The first concerns data, making sure that we do things beyond what the big tech firms and governments are currently doing to guarantee the freedom of individuals. Therefore, we need to make sure that we deal carefully and responsibly with data.

- The second one is about banning social scoring and mass surveillance. This is really about how member states should deal with Artificial Intelligence, and we should ensure that we do not use AI to randomly look in data sets and do social scoring or mass surveillance.

- The third one is helping to monitor and evaluate, offering tools that will assist in the implementation of the recommendation, such as tools for Ethical Impact Assessment and methodology to assess the readiness, the societal readiness for AI systems.

- The fourth, and most importantly, is to protect the environment and show that AI is becoming a more prominent tool to fight against climate change and work positively on environmental issues.

AI changes how we think and how we understand the world; this is why it is vital that we take responsibility for it, and that is why we hope to contribute to this at UNESCO.

References

[1]   UNESCO, "Recommendation on the Ethics of Artificial Intelligence," Paris, 2021.
[2]   COMEST/UNESCO, "Preliminary study on the Ethics of Artificial Intelligence," Paris, 2019.

Work of the Ad Hoc Committee on Artificial Intelligence (CAHAI)

Gregor Strojin,
fmr. Chair of CAHAI at the Council of Europe

The role of international human rights law in tech regulation

In the aftermath of the Second World War, the international community attempted to preventtotalitarian approaches, human rights violations, and atrocities by redesigning the rule-based world order and putting a significant emphasis on human rights legal instruments and mechanisms. As a crucial part of this, the United Nations General Assembly adopted the Universal Declaration of Human Rights in 1948.

Drawing inspiration from the Universal Declaration, the European Convention on Human Rights (ECHR) was adopted by 12 European countries in Rome in 1950. Notably, it also included mechanisms for supervision and enforcement, thus becoming binding and highly influential in the European and global space. The Council of Europe (CoE) was formed based on this in 1949 and is now an intergovernmental organisation composed of 47 countries with more than 830 million inhabitants.

Regulation of new technologies to balance their risks and opportunities with the expected standards of civilisation is by no means novel. We can find it in many other areas of CoE's competence. Through its work, standards for the protection and promotion of human rights, democracy, and the rule of law have developed in many areas of human endeavour. They have since addressed numerous technological developments and their impact on society. Many of them also serve as examples and a basis for future regulation of other new technologies, including artificial intelligence (AI).

Among these, we can emphasise the Convention on the Elaboration of a European Pharmacopoeia (1964), the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981, also called Convention 108), the

Convention for the protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine (1997, also called Oviedo Convention), and the Convention on Cybercrime (2001, also called Budapest Convention).

One of the main reasons for the increasing destructiveness of human conflict during the wars of the past century was the high-impact use of technology. Technological developments served as catalysts for the progress of human civilisation, and various types of materials used for human warfare serve as eponyms for historical eras. However, it was only in combination with the Industrial Revolution that the technology reached the level capable of destroying much more than the immediate enemy.

While military technology is explicitly excluded from the mandate of CoE and limited to the United Nations alone, it does provide an important message: The more effective the technology is, the more impact it has, and the more damage it can do.

This is an important message that relates to the power of AI, as its digital, highly integrated, and often obscure, multiple purpose nature allows it to have an immediate impact on individuals and society on a mass scale. This characteristic can be relevant in situations where it functions correctly or according to its expected specifications, thus significantly augmenting the capabilities of its human users and problems when it makes unintended mistakes, which can then be replicated in all instances.

Techno-solutionist, mechanistic approach was also a significant contributor to various theories developed and utilised in the past two centuries to justify different racist and colonialist views and can now provide valuable insight into the pitfalls of reliance on pseudoscientific methods.

Regrettably, we can still find their modern analogues in specific uses of AI technology. Correlation is often confused with causation. Attributes of an individual are inferred based on automated statistical methods, not reality. Modeling uses unrepresentative datasets or classifies by using arbitrary parameters, etc.

Various misplaced approaches to using AI have steadily trickled into awareness over the past few years. While the future might put them into the same category as, for example, Cesare Lombroso, it requires significant efforts by affected people, activists and lawyers toidentify, analyse and challenge them. As our awareness of them is not comprehensive, their paramount importance is that they have provided evidence to address the matter in a systemic, regulated manner.

Feasibility study

The Ad Hoc Committee on Artificial Intelligence (CAHAI) was established in 2019 and tasked with a two-year mandate to prepare a feasibility study and elaborate the elementsof a potential legal framework for the design, development and application of AI in line with the CoE standards on human rights, democracy and the rule of law [1].

Numerous previous activities and instruments by the CoE and other international governmental organisations have paved the way to its formation[6]. CAHAI's goal was to amend the previous vertical, sectorial instruments relating to AI with a horizontal, transversal approach and prepare a foundation to transition from non-binding instruments, such as recommendations and guidelines, to binding ones. As part of its process, CAHAI has relied on a thorough preparatory work [2].

At the end of 2020, CAHAI unanimously adopted a focused and condensed feasibility study [3], which explored the reasons as to if and why we need a legal framework for AI. CAHAI identified several substantive and procedural gaps in the process.

AI technologies have great potential for our individual lives and societies. However, they can also harm a wide range of civil, political and social rights. The study also provided an essential complementary perspective concerning AI technologies' far-reaching and disruptive effects on the rule of law and democracy.

---

6 Various instruments and other work on AI regulation by CoE bodies can be found on:
https://www.coe.int/en/web/artificial-intelligence/work-in-progress

From a substantive point of view, the rights and obligations formulated in existing legal instruments tend to be articulated too broadly or too generally and very often do not meet the specific challenges raised by AI systems.

Moreover, many essential principles relevant to protecting human rights, democracy, and the rule of law in the context of AI are currently not explicitly assured. They include, for example, the necessity to ensure human control and oversight over AI applications, to ensure their technical robustness, and to secure their adequate transparency and explainability.

In addition to substantive gaps, the study has indicated that current legal safeguards are insufficient and gaps in the current level of protection provided by international binding andnon-binding instruments.

Indeed, we are not in a legal vacuum. Many international legal instruments, including CoE treaties such as the ECHR Convention 108, or the Budapest Convention, are already applicable to AI systems. However, the number and diversity of instruments render them difficult to interpret and apply to the AI context consistently and comprehensively, leading to varying protection levels.

These gaps show a need for a more comprehensive governance framework and effective international legal response to address the challenges and opportunities raised by AI systems.

The critical issue is not the technology itself but the use of this technology without clear rules being provided to AI system designers to ensure that they develop it in line with standards on human rights, democracy, and the rule of law. For example: without the transparency or explainability of an impactful AI-enabled decision, we cannot assesswhether a human right – such as the right to non-discrimination – is actually ensured.

Legal gaps also lead to uncertainty for stakeholders, who lack a predictable and sound legal framework to design and implement AI systems. Uncertainty risks hampering

beneficial AI innovation. Gaps and a fragmented approach to applying these instruments to the context of a globalised community raise uncertainty regarding tackling the transboundary nature of impact generated by the development and use. A lack of standard norms might also hamper cross-border trade of AI products and services.

Ethical guidelines and soft-law instruments have guided CAHAI in this respect. They have helped structure the debate about the fundamental principles that should guide the development of AI applications. Their non-binding character and the lack of mechanisms of oversight of their actual implementation are, however, a significant limitation and clearly show that an appropriate international legal response cannot limit itself to these instruments alone but must consist of a set of complementary and mutually strengthening instruments[7].

The feasibility study has indicated that regulation and innovation are not contradictory but mutually reinforcing. Certainty and coherence are essential for AI operators. Furthermore, when assurance is provided to the citizens that innovation is compliant with legal standards that protect their rights, they will feel more reassured and confident, which will foster the uptake of AI technologies.

CAHAI has reached a consensus that a risk-based approach targeting high-risk applications for human rights, democracy, and the rule of law is needed. A legal framework should provide an enabling regulatory setting in which beneficial AI innovation can flourish while addressing the identified risks and substantive and procedural legal gaps to ensure both its relevance and effectiveness amidst the existing instruments.

Moreover, the feasibility study identified some key principles that a future legal instrument must secure in the context of AI. These include:

• Human dignity,

---

7 Expansion of various initiatives over the past decade can be observed on CAHAI's website:
https://www.coe.int/en/web/artificial-intelligence/national-initiatives

- Prevention of harm to human rights, democracy, and the rule of law,

- Human freedom and human autonomy,

- Non-discrimination,

- Gender equality,

- Fairness,

- Diversity,

- Transparency and explainability,

- Data protection and the right to privacy,

- Accountability and responsibility.

On this basis, CAHAI identified concrete rights that individuals can invoke and examined the requirements that developers and deployers of AI systems should meet. These included existing rights, further clarifications of existing rights, and newly tailored rights to meet the challenges and opportunities raised by AI.

Also, CAHAI looked at possible practical and follow-up mechanisms to ensure compliance with and the effectiveness of a legal framework.

Finally, the CAHAI concluded that a comprehensive legal framework combining binding and non-binding legal instruments that complement each other is the way forward. A binding instrument of horizontal character, a convention, or a framework convention could consolidate general common principles. In addition to this, additional binding or non-binding sectoral instruments could address challenges brought by AI systems in specific sectors. Such a combination would also enhance AI stakeholders' legal certainty and provide the required legal guidance to private actors who wish to undertake self-regulatory initiatives.

Potential elements of a legal framework

During the second phase of its work, CAHAI focused on preparing potential elements of a legal framework on AI based on the CoE's standards on human rights, democracy, and the rule of law.

Among other general remarks, the final document recommended drafting a legally binding transversal instrument to facilitate accession by States outside of the CoE region. Such an approach would increase the impact and efficiency of the proposed instrument and provide a much-needed level playing field for relevant actors, including industry and AI researchers, which often operate across national borders and regions of the world. The standards of the CoE on human rights, democracy, and the rule of law are sufficiently universal to make this a realistic option.

CAHAI also recommended that existing and upcoming legal and regulatory frameworks of other international and regional fora, which are involved in developing various forms of standards related to AI systems, need to be taken into account to ensure global and regional legal consistency.

The purpose of an international legal framework should not be to lay down any detailed technical parameters but to establish certain basic principles and norms governing the development, design, and application of AI systems and regulate, in a consistent and deliberate manner, if and on what conditions AI systems potentially posing risks may be developed, designed and applied by all types of organisations. For this, it should contain certain fundamental principles of protection of human dignity and the respect of human rights, democracy, and the rule of law, irrespective of whether the actors are public or private.

CAHAI favoured a combination of both the establishment of certain direct, concrete, and positive rights of individuals concerning the development, design, and application of AI systems and the establishment of certain obligations upon Parties. This approach would ensure the introduction in their domestic law and practice of measures aimed at protecting the rights of individuals concerning AI systems

and a more uniform application of the legally binding transversal instrument among Parties.

The instrument should establish a methodology for the risk classification of AI systems. The criteria used for assessing the impact of the application of AI systems in this regard should be concrete, clear, and objective. The assessment itself should be done in a balanced manner, thus providing legal certainty and nuance. Risk classification should include several categories (e.g., 'low risk', 'high risk', 'unacceptable risk') based on a riskassessment concerning the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law.

Specific AI systems, which might present an unacceptable risk, should be considered for afull or partial ban or moratorium. Such prohibited practices can include some AI systems using biometrics to identify, categorise or infer characteristics or emotions of individuals, particularly if they lead to mass surveillance, and AI systems used for social scoring to determine access to essential services, as applications that may require particular attention. However, a moratorium or ban should only be considered where an unacceptable risk has been identified on an objective basis and if, after careful examination, there are no other feasible and equally efficient measures available formitigating that risk and given the specific sphere of application.

CAHAI also recommended that Parties establish "regulatory sandboxes" to stimulate responsible innovation in AI systems by allowing for the testing of AI systems under the competent national regulator's supervision. To promote a multi-stakeholder approach and raise awareness in society about the impact of AI, parties should promote evidence-based public deliberations and inclusive engagement.

Recommendations also included specific provisions on:
- Preventing unlawful harm,
- Equal treatment and non-discrimination,
- Gender equality,
- Vulnerable groups,

• Data governance,

• Robustness,

• Safety and cybersecurity,

• Transparency,

• Explainability,

• Auditability and accountability,

• Sustainability throughout lifecycles.

A legally binding transversal instrument must ensure a necessary level of human oversight over AI systems and their effects throughout their lifecycles.

Elements regarding the public sector should include:
   • Additional provisions of access to an effective remedy.

   • Mandatory right to human review of decisions.

   • Adequate human review for processes.

   • Adequate and effective guarantees against arbitrary and abusive practices within the public sector.

Recommendations also highlighted the risk of unlawful or undue interference in democratic processes. They underlined the need to respect the right to freedom of expression, including the freedom to form and hold opinions, receive and impart political information and ideas, and the right to freedom of assembly and association.

A series of provisions on legal safeguards should be applied to all applications of AI systems used to inform decisions impacting the legal rights and other significant interests of individuals and legal persons. These should include:
   • The right to an effective remedy before a national authority (including judicial authorities) against such decisions,
   • The right to be informed about the application of AI in the decision-making

process,

• The right to choose interaction with a human in addition to or instead of an AI system,

• The right to know that one is interacting with an AI system rather than with a human,

• Protection of whistle-blowers.

In order to ensure the effectiveness of a legally binding transversal instrument, it should also include provisions:

• Obliging Parties to take all necessary and appropriate measures to ensure and establish effective compliance mechanisms and standards,

• On the establishment and position of national supervisory authorities,

• On defining cooperation between Parties and mutual legal and other assistance,

• On the establishment of a committee of the Parties to support the implementation of the instrument.

In addition to this, CAHAI adopted proposals for two possible additional legal instruments considered necessary for the future legal framework: Human Rights, Democracy, and the Rule of Law Impact Assessment (HUDERIA), and complementary elements relating to AI in the public sector. CAHAI considered such impact assessment an element of the overall legal framework on AI systems. Still, it does not necessarily form a constituent part of a possible legally binding instrument.

Perspective

By elaborating the elements, CAHAI completed its mandate and sent its output to the policy-setting body of the CoE, the Committee of Ministers, which will frame the next stage of the process. The work of CAHAI will thus form a basis for the negotiations of a future legally binding transversal instrument or an AI treaty, which are to start by May 2022.

The developments at the CoE are concurrent with several other international initiatives. They are all advancing at different speeds, from different starting positions, and with different perspectives and aspirations. They have different membership structures and statutory and organisational goals, influencing their ability to adopt documents at varying breadth, depth, and effectiveness levels.

The CoE deals with human rights, democracy, and the rule of law, the European Union (EU) with the common market and fundamental rights, UNESCO with science, culture, and education, and OECD with economic cooperation and development.

However, only the CoE and the EU are capable and mandated to adopt legally binding instruments. In their timelines, we can observe some common characteristics.

2020 was the year when the need for regulation of AI was established.
2021 was the year when the key elements of regulation were elaborated.
2022 will be the year when verbal commitments will be put to the first real test, as the interests of key stakeholders will become involved at various levels of negotiations.
And hopefully, 2023 will bring about the adoption of effective binding instruments.

References

[1]   C. Muller, "The impact of Artificial Intelligence on Human Rights, Democracy and the Rule of Law," Council of Europe, Strasbourg, 2020.

[2]   I. Ben-Israel, J. Cerdio, A. Ema, L. Friedman, M. Ienca, A. Mantelero, E. Matania, C. Muller, H. Shiroyama and E. Vayena, Towards regulation of AI systems: Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law, Strasbourg: Council of Europe Study DGI, 2020.

[3]   D. Leslie, C. Burr, M. Aitken, J. Cowls, M. Katell and M. Briggs, Artificial Intelligence, Human Rights, Democracy, and the Rule of Law: a primer, Strasbourg: The Alan Turin Institute; Council of Europe, 2021.

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

John C. Havens
Executive Director of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Member of the World Economic Forum Global Future Council on Human Rights and Technology, USA

I would like to acknowledge Elizabeth D. Gibbons and all the experts in human rights; whether it is in a legal background or any other human rights background, I have deep respect for that realm.

My realm of expertise is in technology, in the applied ethics aspect of AI and technologies. When you hear me talk about human rights, my goal is to support the formal work of people like Gibbons and people that worked in the human rights arena for years. This is all intended to be a yes-and, and a complement to the work all of you are giving into human rights. Thank you for the work that you do.

I respectfully submit the idea that responsible innovation and ethically aligned design will hopefully complement human rights work. Elizabeth D. Gibbons was the chair of a committee in our document Ethically Aligned Design, which I will talk about later, and we were genuinely grateful to have her involved in that work.

If you do not know about IEEE, it is the world's largest technology association; it is genuinely the global heart of the engineering community. It was founded about 100 years ago, has members in over 160 countries and 15 out of the 20 top academic journals. I work in the IEEE Standards Association, and the reason we are all talking today is that we are on Wi-Fi, and IEEE created the standards for Wi-Fi. When I say IEEE created the standards, it means that they convened people globally to come to a consensus around the technological interoperability regarding Wi-Fi. It is essentially a communications tool as well as a policy, not a formal legislative policy instrument or what people call soft governance; but a standard is a very powerful

tool because it means people agree on a certain subject, and they can actually create technology based on it.

The tagline, however, for IEEE is advancing technology for humanity, but the word "for" is critical.


Unless what you mean by advancing and for who and how this technology advances is defined, areas such as people's human rights can be violated or not honored because things are not well defined or well designed. I should state that even though I work for IEEE, these views are my own; they do not necessarily represent all of IEEE because IEEE is a very large organization. This is work I helped conduct with people like Gibbons and others who have volunteered their time.

The focus of non human-centric innovation

In general, the focus of non-human-centric innovation or when human rights and human-centric design are not prioritized at the outset of design -this example used here is not to be negative to Facebook or Silicon Valley or any business- but when you ask the question: what are we optimizing for? This is a design question; it is not a moral question with moral ramifications. When you build artificial intelligence (AI) or any other technology, what are you trying to do? What are you optimizing for?

The example used is from Jonathan Stray, a former research fellow at The Partnership on AI, with his idea "what are we optimizing for?".
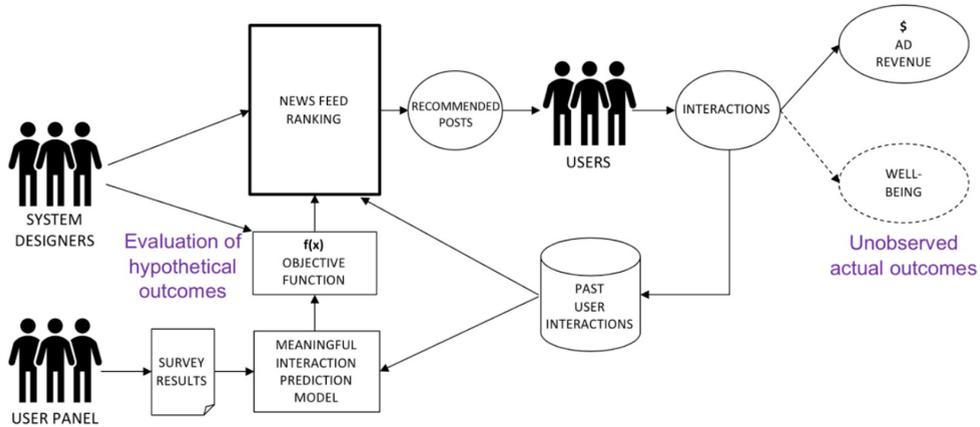
A reconstruction of Facebook's 2018 "meaningful social interaction" changes

He gives an example from 2018 where Facebook tried this experiment called "meaningful social interactions." It was quite an admirable attempt to have people connect to each other well online based on their sentiments and emotions. However, the logic of the underlined goal that is optimization was that if people connected would drive ad revenue, that was the ultimate business goal. It is important to mention here that if this is the key performance indicator at the outset, then that means that other things are deprioritized or not even necessarily focused on, and that can include human rights.

The basis of "What are we optimizing for" logic is the idea of gross domestic product (GDP), and there is nothing morally wrong or evil as it were about GDP. However, exponential growth, which means not just profit, as profit is how we all pay our bills, but means that every decision for a business, government, or person is made based on the question: How can I maximize this growth?

If you only have fiscal measurements, that one set of metrics focused on economics and financial issues, then you naturally will not know how to think about these other areas. It does not mean that people are not thinking about the environment or matters like mental health; it is that they do not have formal indicators or metrics at

the design level when they first start to say, "when I build this, what happens to the environment and the people?".

It is completely different if you think about these things at the beginning of design than after building it, putting it out on the market as fast as possible, and only at that point do your best to keep that system from harming. It is, again, a critical distinction: if you build something, even if you are trying to avoid risking harm, but then put it on the market – especially with things like algorithms that are invisible or that can do a lot of things we are not yet aware of or, in the context of this seminar, if you do not include human rights experts in the design team and the marketing team then both teams, the engineers and data scientists will not even know what risks they do not know about.

During these past 5 years, most of our work has been about how to complement and honor engineers and data scientists; so they do not have to be put in situations that do not make sense for them. Having cross-pollinating or having cross-disciplinary work, which is why it was so helpful to have Gibbons drive one of our committees, allows for conversations such as when a lawyer says the word values, an ethicist or a philosopher, and then a human-computer interaction or robotized set says the same word, there are actually 3 very different definitions according to those 3 areas.

That is just one word in English, coming from a Western perspective; it is a wonderful opportunity to recognize that ethics that are only Western in nature tend to be dualistic but also focused on rationality, based on Aristotle and Greek philosophy. However, we have also to recognize that for a global perspective, to use Eastern traditions say Confucian or Ubuntu ethics, global southern traditions, and certainly, traditions from Africa and around the world. So, one metric to rule all these things I just talked about simply does not make sense.

Thus, it must be recognized from a design sense that if you are not prioritizing the planet and people and prioritizing exponential growth and exponential profits, then that is what is going to continue to be prioritized, and human rights on the planet will continue to suffer.

The focus for human-centric/responsible innovation

What has been used in society for about 70 years is myopic or single-focused, single bottom-line measures for everything we do. So, even if you build AI for good, that for "good" is built on a single bottom line versus a society of people, planet, and profit.

Currently, there is a lot of leading legislation in the EU since they have been thinking about how to create AI for a while now.

"We believe that AI has the potential to significantly transform society. AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation. In particular, AI systems can help to facilitate the achievement of the UN's Sustainable Development Goals…and supporting how we monitor progress against sustainability and social cohesion indicators." [1]

This direct quote is from the Ethics Guidelines for Trustworthy Artificial Intelligence, which have been informing our work, and ours has helped inform theirs, as we have worked with OECD, UNICEF, and UNESCO. The EU experts here said that AI is not an end in itself: any technology is not good or evil, but it does have intentions in one sense as it does not emerge; it is released and affects people. It is a promising means to increase well-being, and it can monitor progress against sustainability. In monitoring progress, you need to have indicators and metrics like the human UN SDGs or the social progress index, or the World Health Organization metrics about mental health, or Gallup Inc. People tend to think that there are not enough metrics or data about matters like mental health, which is factually untrue. There are so many that can be used and need to be used, especially in the wake of Covid and mental health deteriorating globally.

At IEEE, we have been working on this paper that bears the same title as this talk: Ethically Aligned Design Responsible Innovation. Ethically Aligned Design is a paper we started working on at the end of 2015. In 2016, we had a 100-page paper with a hundred experts working on it. We released it as a request for input, and we got 500 hundred pages or more of feedback, most of which said, "this is very western," and

that we must include the perspectives of ethics and philosophical and faith-based traditions not only from the West.

So, in version 2, we added a wonderful chapter called Classical Ethics in AI, and this is part of being responsible if you want to build something globally. It is not responsible to say that one person's region is the right answer for everyone. It does not mean that there is something wrong with your region or anyone else's region; it just means it is poor uninformed design. We are trying to stress this idea of design versus necessarily going to ethics unless the word "applied ethics" is involved, especially in the business world. As someone who used to work at a top 10 PR firm, I get it; my clients or customers would think that we were doing something wrong ethically. Of course, those questions have to be asked in terms of human rights; you have to honor those things that are law. Morals or not is not the question to me; it is rather, are you honoring the law?

The point here is to help the engineers, the data scientists, the ethicists, the social scientists, and the lawyers. Part of the biggest revelation we had is to bring all of these people together in the same room. Even in English, people from the same regions are saying: lawyers, do you understand what marketers are saying? Social scientists, do you understand what the philosophers are saying? Then, after two or three hours of conversation, all these wonderful epiphanies where people realize, "it is this conversation we just had, a three-hour conversation that saved us dozens of hours of work." So, as I mentioned, our document Ethically Aligned Design was used by OECD, UNICEF, and IBM; it has been cited hundreds of times. It is now a 200-page document, the version that came out in 2019 is creative comments designed to be used by anyone; it is pretty evergreen, still all very relevant, but we are hoping to update it.

All the committees that I drive or lead are free and open to join, we will welcome, and we need memberships from Morocco, Egypt, and Africa in general. We really want more people from those regions to get the perspectives of where you are from but also the expertise of what you do.

We are very proud of the standards coming from Ethically Aligned Design. It is the first time in the history of the Standards Association, which is about 50 years old, while the larger organization is about 100 years old.

The standard you are looking at below is the first in the history of the IEEE Standards Association that formally and fully considers the idea of technological interoperability with socio-technological issues.

IEEE 7000™- 2021
Addressing Ethical Concerns During System Design
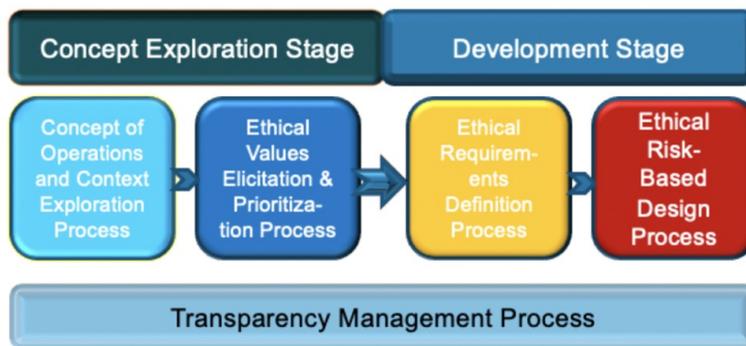
It was a very big project that took five years. Basically, what it does is that it helps anyone who picks up the document have a methodology for what is called value-based engineering. Instead of just wondering what end-users would think about privacy, it is a formal methodology to weigh in on such questions.

The logic here is that if people do not understand applied ethics, which would make sense if one is not an ethicist, they can ask what the artificial intelligence we are building or designing is and what are the various values of the end-users. Once you identify them and you break them down, this standard helps you translate as it were those issues into the process at the beginning of design that you can think about.

It came in about three months ago, there is a free, read-only version of it and our

Standards are about $100. If you end up wanting to try it, then please let me know. We are actually asking governments from different regions of the world, so if you use the standard, let us know what you think about it so that it can be updated.

IEEE Std 7010-2020
IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being
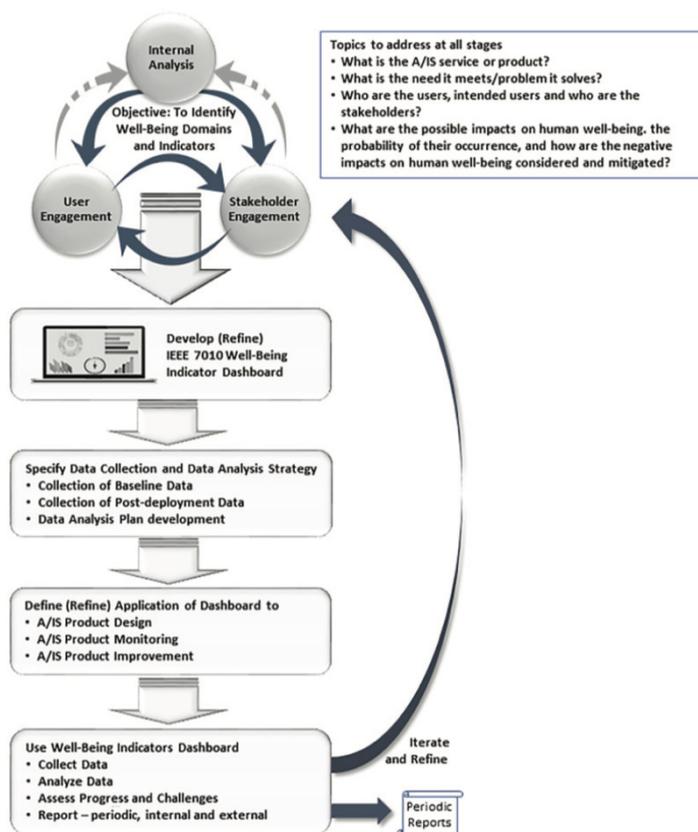
IEEE 7010 standard is another standard that I really think illustrates an idea of what we are doing in IEEE with regards to the conversation on human rights.

This is not a human rights-oriented standard, but it introduces people to metrics

like UN SDGs and others of its kind. As you are about to build your tool along with other end-user values like with the 7000 standard, what are the key performance indicators that you may use at the design level to ensure that people, planet, and profit will be honored in what you build. Not use only metrics for profit, or metrics safe for your brand, or to increase the number of media hits you get -which remain important. How do you know you are going to increase environmental flourishing?

Increasing environmental flourishing is different from just not harming. A big thing that we are focused on is with regards to risk: for engineers and data scientists, you have to avoid risk and harm; that is the job. That is the core job as well as building the tools. However, a world where there is just no risk does not mean the world continues safely. We have to build flourishing for both the planet and people, not just avoiding risk, we have to increase the positive. On that note, we are doing some fantastic work with children as we have a new AI certification 2089-2021 mark.

In conclusion, I would like to leave you with these three things about responsible innovation in terms of human rights:
– "Responsible" Innovation must prioritize human well-being, which I also would say is human rights. That is the first principle in Ethically Aligned Design, thanks to people like Elizabeth D.Gibbons, who brought that to our attention.

– "Responsible" Innovation must prioritize environmental sustainability

– "Responsible" Innovation must honor and utilize methodologies and metrics that honor human rights and values.

References

[1]   High-Level Expert Group on Artificial Intelligence, "The Ethics Guidelines for Trustworthy Artificial Intelligence (AI)," European Commission, 2019.

The Australian Human Rights Commission's experience with AI

Edward Santow,
Industry – Responsible Technology Professor at the University of Technology Sydney,
Former Australia's Human Rights Commissioner

If I were to summarize what we learned in a single sentence from that process: Artificial Intelligence has enormous potential to help us with economic development and greater efficiency in running organizations from governments to the private sector and to make our world more inclusive, but there are also very significant risks and threats of harm, and as a world community we have focused enough on those risks and threats of harm to our basic human rights.

Part of our work at the Commission was to address that imbalance to provide a clear picture of how AI can cause harm to us as humans and present a clear way forward of how we can make sure that we can get the benefits of AI while addressing those risks.

To give a quick picture of a 3-and-a-half-year project culminated with a final report [1] tabled in the Australian parliament in May 2021. It is important to acknowledge that report itself is a government report of 250 pages; despite efforts to make it as simple as possible, the other material on the website is much easier to go through.

Before moving to the key things from the report and explaining what we see as the way forward when it comes to protecting human rights in AI, I consider that we need to start by asking a fundamental question: what is artificial intelligence?

The term AI that is widely used is not a scientific term; it is, if anything, a marketing term. At the heart of artificial intelligence are four key technologies and techniques coming together: machine learning, big data, algorithms, and massive increases in computing power. The combination of these four things has seen the rise of AI over the last ten to twenty years, and we are seeing a shift of AI previously being in the

laboratory and now very much being central to decision-making in the real world.

Perhaps the most important of all technologies and techniques that make up AI is machine learning. Machine learning is essentially a technique that starts with the idea that one needs to teach a machine how to make a particular type of decision, and we do that by categorizing particular training data into, ideally, two categories.

Starting with the example of e-mail, almost everyone has used either Hotmail, Gmail, or Microsoft Outlook; one of the major e-mail applications and central to how that works is machine learning. The developers of such applications would have gone through literally millions or billions of e-mails and labelled those e-mails into two categories, first would have been genuine e-mails, and second, would have been spam.
Over time, a machine learning system would be able to learn the characteristics associated with genuine e-mails – ones that we actually want to receive; and the characteristics associated with spam – in other words, advertising or scam e-mails we do not want to receive. Over time, the e-mail application would automatically be able to categorize e-mails into either the genuine category or the spam category, which is really at the heart of the rise of AI.

It is important to acknowledge that this technique of machine learning can improve decision-making; it can make decisions more rational and less subject to human bias or prejudice. To use a real-world example that is seen in Australia but is also true worldwide: let us imagine a bank trying to decide whether to accept someone's application for a home loan. Up until the last twenty years or so, that decision-making process was pretty chaotic, as the bank would have had a lot of data and information in the form of a big pile of papers to sift through. Ultimately, the decision made by the bank would be a combination of some data – but very hard to access the right data, but also some past practices and, frankly, a significant level of prejudice sprinkled over the top.

Machine learning and AI have enabled us to make sense of that information. To use an analogy, it was like moving all of that massive chaotic information into some order, bundled into some categories. This has substantially improved the way in which decisions are made.

To put this another way, the cartoon drawing below explains well the process in which machine learning can change the way in which we make decisions:
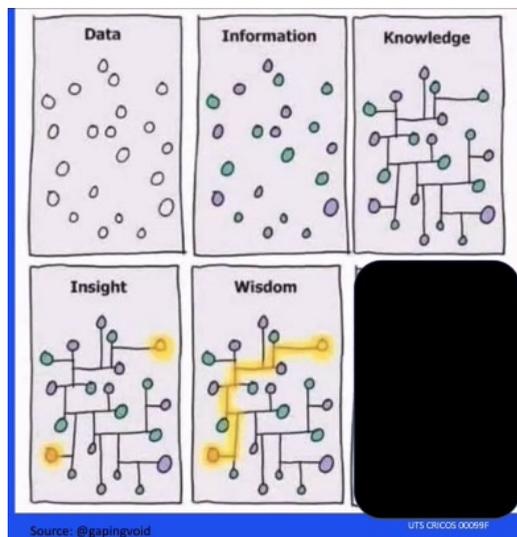
Machine learning and AI allow us to enter data and information to start making sense of what we are seeing before us: from information we can get knowledge, and critically from knowledge we can get insight.

Back to the example of home loan application at the bank, machine learning systems might identify characteristics that seem to be associated with people who were recipients of home loans – people who made their home loans payments back on time, the machine learning system would correlate these characteristics commonly associated with good home loan customers.

Over time, one can build wisdom because we can start to understand why particular characteristics are associated with people who pay back their home loans on time. The problem is that we do not always end up with wisdom as correlations do not always mean something, so just as easily as we can end up with wisdom, we can end up with an incorrect view of the world, as seen on the cartoon drawing below:
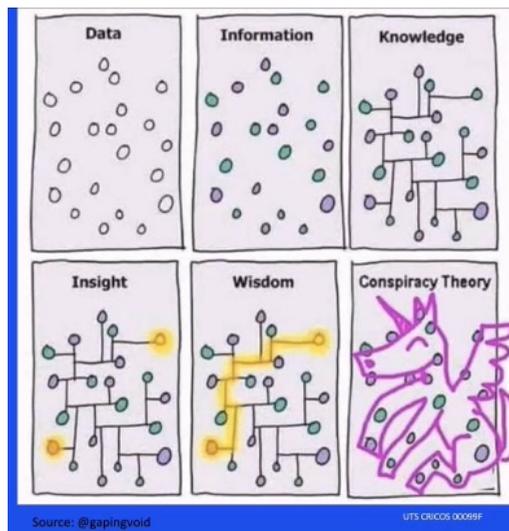
From the Human Rights Commission of Australia's perspective, the concern is that machine learning systems too often end up not with wisdom but with inaccuracy and, indeed, with bias.

Going back to the example of home loans, if we train our machine learning or AI system using the heads of previous home loans decisions, what has been seen in Australia but also around the world is that certain categories of people may be less likely to be granted home loans and that is unfair. We have seen that, around the world, certain racial groups have been categorized as less likely to be able to pay back home loans, and that is just not true; of course, one's race should have no impact on one's capacity to pay back a home loan. Similarly, we have seen that a person's sex or gender can be a factor that machine learning systems can incorrectly learn as being relevant to determine whether to grant someone a home loan.

When we bring all of this together, this really shows that there are some significant human rights risks associated with artificial intelligence, and the Commission's work showed, in particular, three main risks for human rights that we should be worried about:

The first is equality or fairness. Too often, what we have seen with AI systems is that they learn incorrectly that certain categories of people are likely to be, like in the example given previously, bad customers or somehow should be categorized as worse than other people. In human rights terms, we have seen way too often that

this categorization happens by reference to protected attributes, in other words, characteristics about people that we cannot control, such as race, age, gender, disability, etc., and this constitutes a fundamental challenge to our commitment to equality and non-discrimination.

The second big challenge to human rights is accountability. One of the common pictures of decision-making that uses AI is what is known as black box decision-making. What that means is that often when we have a decision made by AI, all we get is the outcome, but we will not get the reasons for the decision, and that is a huge problematic because unless we know how a decision was made, we cannot know whether that decision was fair or even if that decision was lawful.

Returning to the example of home loans: if all the decision one gets is "no, you are not going to be given a home loan," then that person may be left with an unsettling feeling that perhaps that decision was made because of their gender or their race or their disability, but they will not be able to know for sure whether that was the basis for the decision. If all the bank can do is to say that "look, you were denied a home loan because that is what the computer said," then there is never a way of truly getting to the bottom of the decision-making process.

In human rights terms, this represents an existential threat to the right to remedy, which is fundamental to the entire human rights system; if someone's human rights have been violated that they are able to access remedy or redress. Black box decision-making, which is too often a part of AI, is a challenge to that. The solution to this problem is to insist that decision-making systems that use AI provide those reasons that we all rely on.

Perhaps most obviously, the way in which people's personal information is used is a challenge to our right to privacy. What the Commission had heard over and over when it was doing its public consultation was that people in Australia and other countries were saying, "I have just realized that with AI, my personal information can be used against me," and that is in a sense the most fundamental articulation of why we have the right to privacy. It is not just so that we can live a life secluded from the rest of the community – although that is part of the right to privacy, but the most

fundamental aspect in my view is that your personal information not be used against you and not to be used unfairly to harm you. This question was right at the heart of the report.

To bring this into the real world with the previously given example of home loans and banking, we have seen in Australia and throughout the world that these problems have started to become more and more common, which shifted from being theoretical problems to problems that we see in the real world and that started to be reported on [2].
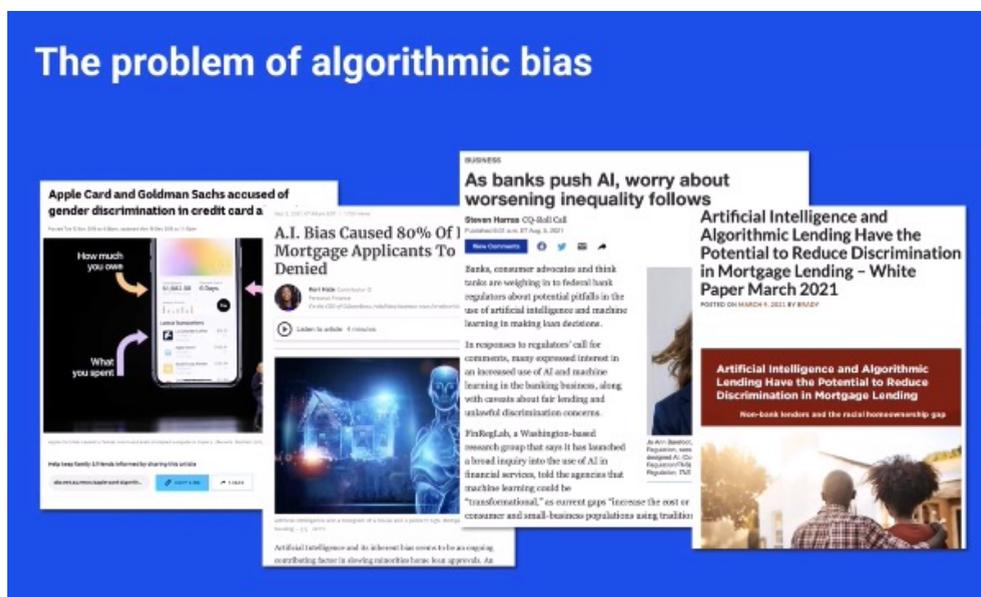
As a final observation, I ask: what is the way forward?

When the Commission asked the community what they wanted when it came to the use of artificial intelligence, it really boiled down to four key things:

– Decision-making systems must be fair. The right to equality is so central to the way in which personal information is used in AI that it must be upheld, and more needs to be done in that area. The technical paper [3] done by the Commission is a deep

exploration of that issue.

– Decision-making systems have to be accurate. As a human rights lawyer, my principal concern is that if a woman or a person of color is being denied a home loan, such as in the example used previously, there is a human rights violation, but thankfully it can also mean that it is just an error. We know that if these problems are not addressed, then the decision-making process will not just be unfair; it will simply be wrong. The focus on making sure that AI works accurately was another of the four key things that the Commission has heard in its public consultation.

– AI must be fit for purpose. AI can be used in some areas very effectively, very accurately, and very safely, but that same technology is used in other areas in a way that is not fair, accurate, or safe. A good example is with respect to facial recognition; we are increasingly becoming accustomed to using facial recognition essentially like a password to unlock smartphones or computers, which can be reasonably safe and accurate. However, we know that a very similar technology of facial recognition used in the policing context to identify criminal suspects can be far less accurate. There are some critical technical reasons why those two things are very different, although they seem to be relying on the same technology. Therefore, making sure that we use those technologies in an appropriate way that is safe and accurate.

– Accountability is at the heart of making sure that AI is reliable and safe. Black box decision-making should be something that is pushed back, and its antidote is to make sure that decision-making systems are designed in an accountable way. On an individual level, people affected by decisions should always have a right to a remedy if the decision is unlawful and/or unfair. At the systemic level as well, decision-makers must be able to look at the output of their systems to make sure that the system-wide level is accurate and fair and is not causing problems with privacy.

There is an enormous positive potential [4] associated with the rise of AI, which is exciting, but that positive potential will never be truly realized unless we get to grips with the risks and threats of harm that the Australian Human Rights Commission puts the focus on in its project.

References

[1]   Australian Human Rights Commission, «Human Rights and Technology Final Report,» 2021.

[2]   E. Santow, "Can artificial intelligence be trusted with our human rights," *Australian Quarterly,* vol. 91, no. 4, pp. 10-17, 2020.

[3]   Australian Human Rights Commission, «Using artificial intelligence to made decisions: Addressing the problem of algorithmic bias,» 2020.

[4]   E. Santow, "Emerging from AI utopia," *Science,* vol. 368, no. 6486, p. 9, 2020.

In search of accountability: A human rights critique of Guidelines for Ethical AI

Elizabeth D. Gibbons,
Instructor, FXB Center for Health & Human Rights, Harvard University, USA.

The experience of working on the IEEE Ethically Aligned Design was one of the best in my professional career because of all the interdisciplinary work we did and the different, very necessary, perspectives we catalyzed together as the Sustainable Development Committee to produce the outcomes for IEEE.

1) John C. Havens has referred to the great promise of AI also the many risks and potential harms. The focus of my remarks today is on how we are going to ensure accountability for these harms, if and when they occur. How will there be remedy to affected populations? What I am presenting is based on a paper [1] that I co-authored with Sakiko Fukuda-Parr, who is a professor of International Affairs at the New School. We researched specifically how accountability and other key human rights principles are reflected in a sample of the many guidelines professing to provide a framework for the ethical application of Artificial Intelligence.
How is AI defined, and, as Havens said, do different constituencies use the words in the same way? I accompanied my remarks with a PowerPoint presentation, which can be found in the annex to these proceedings. For ease of reference, I have numbered the sections of my remarks according to the relevant PowerPoint slide.

2) There is a proliferation of stakeholder guidelines from many different parts of society: corporations, governments, multilateral organizations, and civil society. Algorithm Watch [2] started an inventory of guidelines in 2019, and by 2020, the inventory had doubled to 160 guidelines on the ethical development of AI. While guidelines proliferate, there is an absence of common national/international standards, regulations, and legal frameworks to address the challenge of AI, although, as John C. Havens presented, there are some emerging and exciting efforts to develop a common mechanism.

At the UN, a lot of human rights issues related to AI-DDD (design, development, and deployment) are those related to the right to privacy, the surveillance state, the right to assembly, racial discrimination, poverty, digital welfare states, and of course the big issue of corporate accountability. There is also a very significant discussion between the human rights community and the ethics community over which is really the best framework for ensuring that AI advances society and serves humanity. Philip Alston, the former UN Special Rapporteur on Extreme Poverty, stated, "As long as you are focused on ethics, it is mine against yours. I will define fairness, what is transparency, what is accountability. There are no universal standards." [3] That is the key difference with the human rights framework, which has established universal standards; while these human rights standards are, of course, adapted to different cultures and different legal frameworks, they *are* universal. That is why the international human rights community is promoting a human rights framework, with its universal standards and principles, as more effective in ensuring that AI assists human society to realize its potential.

3) With this proliferation of guidelines, there have been, of course, several efforts to analyze them across various domains to see whether common themes are emerging. John C. Havens had mentioned the Harvard Berkman Klein Review, which was one effort [4]; Algorithm Watch and Professor Jobin [5] both reviewed extensive numbers of guidelines, while Professor Asaro [6] analyzed 28 specific to the private sector. One potential outcome of these reviews is that the identification of common themes across a great diversity of guidelines allows these themes to serve as the basis of global norms and legal standards for the ethical design of AI.

Human rights and accountability are both themes that emerge clearly across all these different reviews. The question again is, are these guidelines that use the same words actually talking about the same thing? So, we conducted a detailed analysis of 15 guidelines, which we selected from those that the Harvard Berkman Klein mapping review (referred to in John's presentation) considered to be strong in human rights content.

4) We sought to answer what the guidelines really meant when they state "respect

for human rights." We did this by developing and applying an analytical framework anchored in (1) International Human Rights Law and (2) Human Rights core principles of equality and non-discrimination, participation, and accountability. These principles are the process by which every human right should be realized. Given that the right to privacy is ubiquitous across the guidelines, we gave specific attention to that unique human right: the right to privacy.

5) There are issues particular to AI when applying these core HR principles. With respect to the principle of equality and non-discrimination, we know that there is often an inherent bias within algorithms, reflecting the imperfect judgment of humans, but also the availability of data and how much of the data is embedding historical inequalities. Low- and medium-income countries (LMIC) or marginalized communities within countries may not produce sufficient data to train an AI application, or the data produced may be improperly captured or irrelevant to the problem or even to the population at risk. Infrastructure constraints in countries/communities exclude them both from contributing to training data and accessing the potential benefits of AI. There are still countries in the world where over 90% of the population has no access at all to the internet.

Finally, applications designed in high-income countries (HIC) may not function properly in LMIC, both from a social and infrastructure perspective. Therefore, there is a high risk that AI reinforces or even accentuates existing inequalities, not only inequalities between and within societies but also between countries.

6) With respect to participation in AI-DDD, some leading impediments are the knowledge and power asymmetry between companies and people, as few in a population are going to have sufficient knowledge to understand how to hold companies accountable. That is partly due to the companies' proprietary information data protection mechanisms, whereby they prefer not to share or will not share the source of an application's training data, the algorithms used, etc. This knowledge asymmetry poses a risk to the right to information, without which there can be no meaningful participation; this, in turn, limits the possibility for the public to hold the tech industry accountable.

In terms of accountability, the OHCHR defines accountability as having three interdependent elements: responsibility (who or what entity has the responsibility to protect rights), answerability/transparency (when the public demands answers on how a decision was reached, there is a transparent process and product for providing those answers), and finally enforceability and remedy (a mechanism for enforcing accountability for the impact of an AI application, with access to remedy for those it may have harmed). The current weakness of the AI regulatory regime impedes enforceability and remedy, thereby posing an ongoing risk to the public of potential short and long-term harm.

7) Regarding the right to privacy, as we know, AI depends on access to and monetization of personal data, over which individuals are unable to assert their ownership or privacy. That is an ongoing concern. One of the professors quoted in the Rathenau Institute report was damning in her assessment "We the citizens have been reduced to raw material-sourced bartered and mined in the 'privatized commons' of data surveillance" [7]. With the right to privacy obliterated, there are irreparable consequences to the realization of other rights: assembly, equal protection, justice, health, education, etc.

8) Bearing in mind the risks to human rights principles particular to AI, we analyzed a small subset of guidelines to see if and how these risks were addressed and the solutions proposed. We selected only 15 out of the universe of 160 guidelines on ethical AI, and these were chosen due to their attempt to address human rights concerns of AI-DDD. What we found was that less than half (7) of these guidelines incorporated the key elements of a human rights framework, and 8 adopted an ethics framework, meaning only 7 include all elements of the human rights framework grounded in international human rights law, and address the principles of equality and non-discrimination, participation and accountability including remedy. Although most of the guidelines recognized AI's potential for human rights violations or harms and the importance of preventing harm, only 7 committed to the human rights' legal framework as a means for doing so. These same 7 also incorporated the core human rights principles in their guidance.

With respect to the principles of equality and non-discrimination, the risk of bias

was addressed by all guidelines, but few proposed proactive measures to reduce bias and its impact on socio-economic inequalities within and between societies.

9) Regarding participation, most guidelines mentioned stakeholder consultation, and some referred to power asymmetries and the need to empower users, but only 6 recognized participation as a right; only 2 had participation as a cross-cutting principle. Some ethics guidelines even considered participation equivalent to market research (which, of course, is simply another way to find out how to monetize their product).

Accountability is most problematic. Although it appears as a universal principle across all guidelines, those employing ethics frameworks did not reflect the full human rights definition of accountability, limiting it to responsibility and answerability/ transparency of, for example, the decision-chain in AI development.

However, enforceability, restitution, or remedy for AI-DDD's negative impact on society is absent from guidelines adopting ethics framework. This is what, I think, is the crux of the dispute between the ethics community and the human rights community, with the goal of ensuring maximum benefit and minimal risk of AI to society. Without the possibility of enforcing the protection of human rights, the protection of society (especially the most marginalized) from potential harms, or remedying actual harms of an AI application, we cannot have an ethical or functioning regulatory system.

10) On privacy, all 15 guidelines addressed privacy and data protection. Only 6 considered privacy a human rights issue, and neither ethics nor human rights guidelines identified how privacy and data protection can be fully operationalized. All of us feel some hope seeing how the EU is making significant progress in handling this very difficult issue. However, Access Now had a rather blunt statement "With people creating a trail of data for every aspect of their lives…it is questionable whether data protection is even possible" [8].

11) Through a table in the article, we captured how each of the 15 guidelines

measures up against the different elements of the analytical human rights framework we developed. Of course, if interested, you can certainly read the article itself.

12) In the end, we concluded that corporations are claiming the "ethics" space, defining norms and redefining "human rights" to suit their business model, rather than meeting the standards demanded by these international instruments. Guidelines for "ethical AI" that claim to respect human rights must be scrutinized to ensure they have provisions that actually: protect against bias and discrimination, allow for meaningful participation, and understand accountability beyond responsibility and transparency, to include enforceability and remedy.

Of course, human rights and ethics are not mutually exclusive but complementary; and indeed, sometimes the human rights infrastructure takes time to evolve, and ethics guidelines can be a necessary way to bridge the gap while the legal system catches up.

13) Both human rights and ethics frameworks, we found, fall short in operationalizing their principles. However, AI-DDD guided by self-defined ethics is unenforceable, thus allowing corporations to escape true accountability to the public. Voluntary corporate guidelines on ethical AI cannot check the power of Big Tech, while their proliferation risks forestalling the development of necessary legal norms/ enforcement mechanisms. However, in the meantime, while awaiting these norms and mechanisms, we have a framework to hold AI developers accountable: the existing human rights legal framework is universal - every country in the world has certainly ratified at least one human rights treaty, if not several.

While human rights instruments are evolving, the existing framework has operational standards and principles for holding AI-DDD producers accountable for any potential harm to society. Of course, none of this will be easy, given the incredible complexity of AI; however, international human rights offer an existing, operational, legal framework. We believe we should use it.

14) If you would like to know more, either by reading this article or exploring

related topics, the Global Policy Journal set out a special issue on Digital Technology and the Political Determinants of Health Inequities, which includes our article.

References

[1]   S. Fukuda☐Parr and E. Gibbons, "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines," *Global Policy,* vol. 12, pp. 32-44, 2021.

[2]   Algorithm Watch, "AI Ethics Guidelines Global Inventory," 2019. [Online]. Available: https://algorithmwatch.org/en/ai-ethics-guidelines-global-inventory/.

[3]   Human Rights Council, "Report of the Special Rapporteur on Extreme Poverty "Digital technology, social protection and human rights"," Geneva, 2019.

[4]   J. Fjeld, N. Achten, H. Hilligoss, A. Nagy and M. Srikumar, "Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI," Berkman Klein Center Research Publication, 2020.

[5]   A. Jobin, M. Ienca and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence,* vol. 1, no. 9, pp. 389-399, 2019.

[6]   P. M. Asaro, "AI ethics in predictive policing: From models of threat to an ethics of care," *IEEE Technology and Society Magazine,* vol. 38, no. 2, pp. 40-53, 2019.

[7]   J. E. Cohen, "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power," *Surveillance & Society,* vol. 17, no. 1/2, pp. 240-245, 2019.

[8]   Access Now, "Human Rights in the age of Artificial Intelligence," 2018. [Online]. Available: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf.

[9]   Rathenau Institute, "Human Rights in the Robot Age: Challenges arising from the use of robotics, artificial," 2017.

- IV -
NATIONAL INITIATIVES

AI Movement: The Moroccan International Center for Artificial Intelligence at UM6P University

Amal Fellah Seghrouchni,
Head of Ai Movement, Mohammed VI Polytechnic University

Ai movement, the Moroccan International Center for Artificial Intelligence is a center of excellence in Artificial Intelligence that aims to foster the emergence of Moroccan expertise in Artificial Intelligence and Data Sciences. It is both:

• An articulating and consolidating tool of various actions related to the field of AI, with the ambition of making Morocco a regional AI hub impacting its ecosystem, on strategic, educational and industrial levels.
• A lever to anticipate and accompany the evolutions and transformations related to Artificial Intelligence and Data Sciences, the aim of which is to provide innovative, operational, resilient and ethical solutions to the problems of society, environment, market, economy and technology.
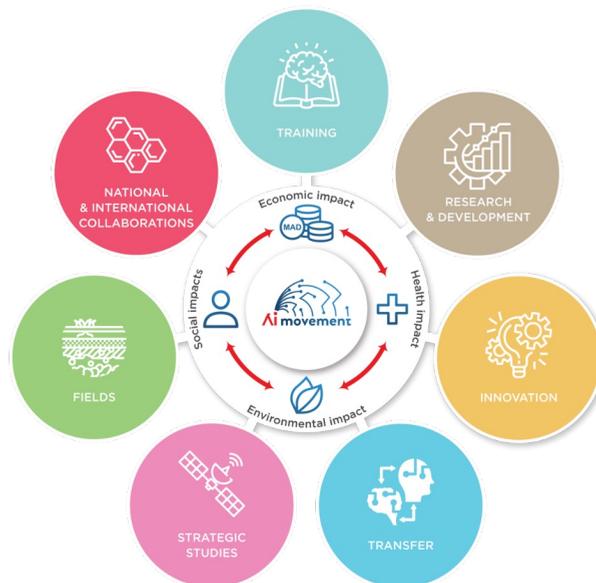
Mission

As a hub for Artificial Intelligence transformation, Ai movement is based on 7 pillars that structure and define its missions:

- Foster the emergence of an attractive ecosystem
- Attract international expertise in AI and Data Sciences to work in collaboration with national researchers.
- Create international synergies and bring out national talent.
- Create and strengthen international quality partnerships.
- Develop a field-oriented approach
- Rise society's awareness and understanding to accept transformations related to the dematerialization of services and AI.
- Leading change in a way that is adapted to different strata of society.
- Study the various fields and identify the needs and expectations of different societal strata.
- Promote inclusive training

• Design and implement appropriate training to support the appropriation of AI and its evolution: initial training, training for young people from middle school, lifelong training for adults, vocational training, etc.

▪ Combine theoretical and applied research, technology development, innovation, and industrial transfer

• Develop research structures and connect them to the business world to create synergy between research, industrial and societal needs.

• Develop research that fosters innovation, creativity, and development of disruptive and innovative tools of high added-value technology products (including tools to deploy for educational innovation).

• Put a particular emphasis on start-ups and youth entrepreneurship.

• Transfer the results of the research conducted in partnership with the economic world (R&D) and innovations to companies on the one hand and to society on the other.

• Promote the creation of an R&D ecosystem (fostering multidisciplinarity and collaboration with existing research centers, national and international research laboratories), entrepreneurship, and innovation (FabLab, Start-ups, SMEs, VSEs, etc.) with short cycles and synergies for creativity.

▪ Strategic studies

• Ensure and maintain a strategic and geopolitical vision concerning transformations induced by AI.

• Develop tools such as "Think Tanks" or "convergence" institutes to shed light on the societal, economic, ethical, and (geo)political impact of AI developments.

In addition to scientific research work, the training includes a large part on the governance of AI. For example, out of 360 hours of training for the Master AI Governance & Practice, 90 hours are devoted to data management, data regulations and data ethics. Another 50 hours are dedicated to the study of AI applications in management, human resources, energies, education, agriculture, ...

AI GOVERNANCE & PRACTICE

EXECUTIVE MASTER

A PIONNER AI EXECUTIVE PROGRAM

LOW-TECH/ DEEP-TECH

The first Executive program to propose bridge courses to build the skills to aboard AI from a technological standpoint.

GENERIC/CROSS INDUSTRY

Target senior executives from a broad set of sectors, and keep everyone interested!

Ai movement
EXECUTIVE MASTER

AI GOVERNANCE

Learn about AI project governance and good practices.

VERSATILE

Adapt to the group needs and propose UpToDate course material.

Ai movement
Program Units
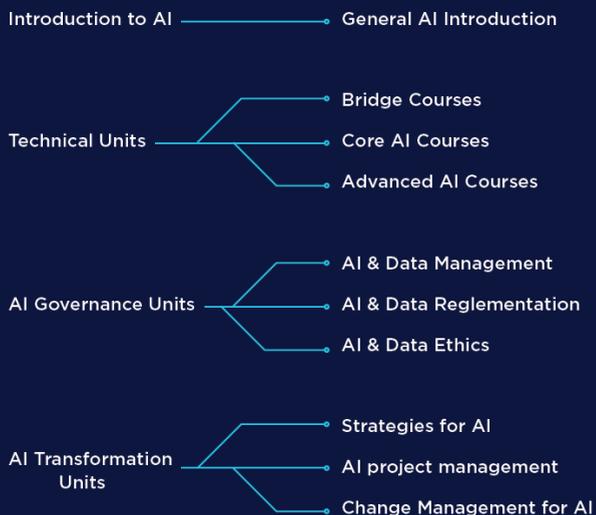
Introduction to AI ———— General AI Introduction

Technical Units
- Bridge Courses
- Core AI Courses
- Advanced AI Courses

AI Governance Units
- AI & Data Management
- AI & Data Reglementation
- AI & Data Ethics

AI Transformation Units
- Strategies for AI
- AI project management
- Change Management for AI

PROGRAM SYLLABUS:
THE CERTIFICATES

C1
AI Foundations Certificate — 90H

Learn about AI foundations and explore the potential of machine learning in different fields

Introduction to AI — 4H
- Gain a basic understanding of AI fields and learn common keywords used in the tech industry today.

AI Foundations — 30H
- Learn about AI foundations algorithms for Planning, Problem Solving
- Decision Making and Optimization.

Machine Learning — 30H
- Explore the power of Neural Networks & Ensemble techniques for an informed application of artificial intelligence in business management

AI Business case — 26H
- Learn the techniques and methodologies involved in building AI business case.

Bridge course (optional): Math & CS tools for AI — 10H
- Math and computer science tools for AI foundations

C2
AI Application Certificate — 90H

Explore the full potential of specific AI architectures for multiple industry applications to form a strategic understanding of AI (*)

Advanced ML and Computer vision — 20H
- Acquire a bird-eye view of advanced machine learning architectures, with a focus on Computer vision applications.

Natural language processing / AI for robotics — 20H
- Learn about two main case application of Machine learning in business: Natural Language Processing and Robotics.

AI Application 1 — 25H
- Management/RH/ Renewable Energies/ Marketing/Digital Transformation/ Education/ Cybersecurity/ Agriculture

AI Application 2 — 25H

C3
AI Governance Certificate — 90H

Learn about the latest reglementation in AI & Data management, the best practices for a responsable and Ethical AI

Introduction to AI Governance — 4H
- Introduce AI and the importance of appropriate Governance skills for innovative leaders covering trustworthy AI, data management, reglementation and ethics.

AI & Data Management — 30H
- Gain insights with regards to the best practices in data management for a seamless AI system deployment (Cloud, Data Center, embedded AI...).

AI & Data Reglementation — 26H
- Learn about reglementation that corporations and governments should follow in order to responsibly design and deploy data-sourced AI systems.

AI & Data Ethics — 30H
- Develop the tools and skills to deploy ethical AI and manage the diverse impacts of this technology on your organization and the society.

C4
AI Transformation Certificate — 90H

Learn how to manage a complex AI project and successfully lead the AI transformation in your organization

Strategies for AI: — 30H
- AI genesis and history
- Value Chain for AI
- Links to organizational models, shared services and processes
- Core and supporting AI strategies

AI project management : — 30H
- Analytics project lifecycle
- Specificities of an AI project and related risks
- Upstream and downstream integration

Change Management for AI: — 30H
- Develop skills to implement organizational change with key stakeholders fostering an organizational learning culture that embraces AI transformation.
- Apply conceptual frameworks, models and analytics that support integrating AI transformation in business strategy.

* Management / RH / Renewable Energies / Marketing / Digital Transformation / Education / Cybersecurity/ Agriculture

Ethics in action of the Ethics Observatory of the House of Artificial Intelligence

Saïda Belouali
Professor of Applied Ethics, Co-responsible for the House of Artificial Intelligence, Mohammed I University (UMP)

The project of a new partnership

The project of an Artificial Intelligence House at Mohammed I University is inspired by La Maison de l'Intelligence Artificielle (MIA) of the Alpes-Maritimes. The idea is sponsored by EuropIA Institute. MIA was conceived as "a space entirely dedicated to AI and its applications to enable everyone to grasp a technology that reshapes the contours of the future." Marco Landi emphasizes that this is the first house of its kind in Europe. He describes it as a place of 'SMART Education' with the aim of preparing and training the population to understand the challenges ahead and grasp the new possibilities offered by AI.

AI offers new avenues for the dissemination of expert and ethical knowledge to the general public while providing opportunities to develop innovative and collaborative projects by bringing together different actors in the AI ecosystem. Data, artificial intelligence, and the Internet of Things are present in our lives and have disrupted our relationship to knowledge and the economy, which could reshape our future and transform our societies. Similar to the MIA affiliated with the University Sophia Antipolis, Mohammed I University wants to imagine a place where reflections and experiments related to AI are focused on uniting all actors and stakeholders around a future shaped by AI.

Why the Eastern region?

AI is already producing an unprecedented rupture in society. We do not want this

rupture to be a technological barrier or ethical divide for our country. Quite the opposite, we want to be amongst the key stakeholders responding to and steering these changes. Each territory is able to position itself in terms of AI capacity, such as human capital and high levels of IT expertise. Mohammed Premier University is increasingly committed to these technological changes and has, amongst other things, in its arsenal:

- A critical mass of teachers and researchers engaged in these issues
- A strong dimension of training (engineering and master's degrees) in IT, AI, and Data Science
- An anchor in a socio-economic environment and a vehicle for federating partnerships and projects
- A computing center (in progress)
- Prototyping and 3D printing center

The University has already established a space that supports researchers in initiating IT research and development (R&D) via an incubation ecosystem. The Knowledge Campus (Campus du Savoir**)**, with an area of 30,000 m², which consists of several spaces for R&D. It is fundamentally dedicated to these different initiatives. These different spaces reflect the choice of Mohammed I University to put in place all necessary and strategic devices to help produce knowledge that can become a real economic value for the region.

Components

At the Campus du Savoir and on a surface of 1300 m², the MIA-UMPO-Morocco will be designed to better understand AI and its transformations. It is a place where we will discover, understand and experiment with this technology.

- Ethics monitoring observatory
- Exhibition hall
- CreaLAB
- Coworking space

The aim is to prepare the eastern region to understand the changes linked to AI and adopt a human-centric approach to AI. This space aims to boost institutional,

academic, and industrial collaboration around new technologies and their challenges. The space will be animated regularly by events on innovation and AI. The various events will allow the public to learn, exchange and develop knowledge.

Missions

- **AI in practice**

  – Offer a demonstration showroom

  – Develop use cases/experiments

- **Ethics observatory**

  – Measuring the societal and ethical impacts of AI

- **Animate and federate the actors**

  – Building a collective dynamic: events, fairs, etc.

  – Promote and raise awareness

  – Promote new uses of AI by raising awareness among different audiences

  – Promoting scientific excellence awards, publications, and R&D projects.

Projects

The MIA wants to create a real dynamic of acculturation to AI through public experiences but also promote applied research with a substantial societal and economic   impact.

Applications have been carried out and are presented at the Maison de l'Intelligence Artificielle

 – **InkadSearch**[8]: Intelligent search engine for research articles in medicine and biology

 – **Aladilemma**: The Platform of exercise in Ethical Decision-Making

---

8   Project funded by the Centre National pour la Recherche Scientifique et Technique, Maroc.

– **Confilearn**[9]: Smart Offline Learning Platform

– **FrBMedQA**: First Dataset Questions/Answers in French in the biomedical field

– **InkadBot**[10]: Covid-19 Analysis Help Chatbot

InkadSearch, Confilearn, and InkadBot are all projects funded by the Moroccan National Center for Scientific and Technical Research (CNRST).

Current projects at the MIA include

– Smart Water Campus: a life-size project which aims to produce intelligent monitoring of all water activity on the Knowledge Campus (rain, drinking water, wastewater, etc.).

– Robotic caregiver: an intelligent robotic caregiver likely to take over in pandemic situations.

---

9 *Idem.*
10 *Idem.*

The role of National Human Rights Institutions (NHRI) in promoting human rights-based ethics of AI: Case of CNDH Morocco

Mounir Bensalah,
Secretary-General of the National Human Rights Council – CNDH

From a human rights perspective, regulations include "rules and other processes aimed at moderating individual and organizational behavior in order to achieve established goals" [1] and are divided at the national level into three categories: – Self-regulation: when actors in a particular field organize themselves to develop codes, standards, or guidelines that must be voluntarily complied with to address problems identified.

– Co-regulation: when the actors and the government work together to develop strategies and codes to be followed in this field.

– Legal regulation: when legal or regulatory texts require compliance with the rules and/or facilities of monitoring and/or control institutions.

In numerous fields, the organizational structure varies according to the adopted approach and the maturity conditions in the concerned field. For example, the authorities in Morocco adopted the regulation of communication through legal regulation in accordance with Law 24.96 relating to postal and telecommunications services, as amended and supplemented by Law 55.01, which provided for the creation of the National Telecommunications Regulatory Agency (ANRT). In contrast, the adoption of self-regulation in the field of press and publishing was based on the requirements of Law 90.13 establisVhing the National Press Council, its bylaws, and the National Charter for Journalism Ethics.

With regard to digital audiovisual communication, the High Authority for Audiovisual Communication (HACA) in Morocco organized[11] beginning of 2020 an international forum on "Regulating the media in a digital, mobile and social environment: Adaptation necessities, challenges of reorganization," in which national and international institutions, technology companies and specialists participated. Recommendations in the direction of self-regulation and co-regulation were issued.

International conventions, recommendations from UN organizations, treaty committees, and mandate holders do not refer to any best regulatory formula in the field of Artificial Intelligence regulation to protect human rights.

The United Nations considers[12], in the words of its High Commissioner for Human Rights Michelle Bachelet from September 2021, that artificial intelligence technologies can have adverse or even catastrophic effects if used without adequate consideration for their impact on human rights. The greater the risks to human rights, the more stringent legal requirements for the use of artificial intelligence technologies should be. The High Commissioner for Human Rights added that urgent action is needed to assess the risks for the various systems that rely on artificial intelligence; since the risks' assessment and consideration can take time, countries should impose a moratorium on the use of potentially high-risk technologies.

There have been many international initiatives in recent years that aim to think of ways to protect human rights in the field of Artificial Intelligence. A reference book [2] has numbered more than 126 initiatives to establish guidelines for the ethics of Artificial Intelligence across the globe. Experts [3] record that 24% of them are located in the United States and 16.7% in the United Kingdom. They also note some initiatives' "stalemate," including those related to the issue of cultural diversity, and regret the attempt to impose a Western vision on the rest of the world [4]. In this sense, experts note that African countries and Latin America are not present in terms of initiatives outside the scope of international organizations [5].

---

11 https://maroc-diplomatique.net/haca-conference-a-rabat-sur-la-regulation-des-medias/
  12https://news.un.org/en/story/2021/09/1099972

New research notes that "'ethics' guidelines, disproportionately from corporations and other interest groups, are also weak on addressing inequalities and discrimination" and that "this exposes an urgent need for action by governments and civil society to develop more rigorous standards and regulatory measures, grounded in international human rights frameworks, capable of holding Big Tech and other powerful actors to account" [6].
The United Nations Secretary-General launched the "Big Data and Artificial Intelligence for Development, Humanitarian Action and Peace" initiative under the name UN Global Pulse[13]. In 2019, UNESCO also launched a "preliminary study on the ethics of Artificial Intelligence," which worked to propose a set of relevant guidelines towards the adoption of a normative agreement [7]. Audrey Azoulay, UNESCO's Director-General, stated that artificial intelligence is the new frontier of humanity[14]. Once this frontier is crossed, a new form of human civilization will emerge. The guiding principle of AI is not to be independent or replace human intelligence. Still, we must ensure that it is developed according to a humanistic approach based on values and human rights. We are facing a crucial question: What kind of society do we want for tomorrow.

The AI revolution is opening up exciting new horizons, but the anthropological and social upheavals it generates are worth comprehensive reflection and consideration. The International Telecommunication Union (ITU) held the fourth session of its international conference Artificial Intelligence for Good, during which its Secretary-General Houlin Zhao stated: "The AI for Good Summit is the main platform for the United Nations for a comprehensive dialogue on AI. The Summit identifies practical applications of AI to accelerate progress towards achieving sustainable development goals and promotes collaboration to help these applications achieve global impact" [8]. In the Global Partnership on Artificial Intelligence Joint Declaration, the founders affirmed their commitment to support the development and the use of responsible and human-centered AI while respecting human rights, fundamental freedoms and their shared democratic values, as elaborated in the OECD Recommendation on Artificial Intelligence [9].

In addition to governments, United Nations, and international agencies,

---

13 https://www.unglobalpulse.org/
14 https://en.unesco.org/news/unesco-holds-first-global-conference-promote-humanist-artificial-intelligence

professionals and experts have also addressed this issue. Indeed, the IEEE Global Initiative on the Ethics of Autonomous and Intelligent Systems has articulated the ambition of this project and its "ethical biased design" standard in a vision to prioritize human well-being with autonomous and intelligent systems. Its General Principles Committee "seeks to articulate high-level ethical concerns that apply to all types of autonomous and intelligent systems (A/IS*), regardless of whether they are physical robots (such as care robots or driverless cars) or software systems (such as medical diagnosis systems, intelligent personal assistants, or algorithmic chat bots)" [10]. Amongst the giants of the IT and AI industry[15], a hundred partners (academics, industry experts and NGOs) came together for the Partnership on Artificial Intelligence in order to think about questions that will help us most in exploring ethical issues and the unintended consequences of creating and deploying emotional intelligence that keep influencing us as a society. They believe that we need to ask ourselves together, and if we want to develop and use artificial intelligence that we can feel connected to, can recognize, can influence it, and can imitate.

In their reports of recent years, the UN human rights system mandate holders unanimously address the problems of respecting all human rights in artificial intelligence systems [11]16] [15] [14] [13] [12]]. The most important recommendations in this regard can be summarized as follows:

Recommendations for States:
    – Enhancing and facilitating access to digital technology, bridging the digital divide, and not imposing restrictions on its use to exercise various rights and freedoms.

    – Preparing and publishing reports on transparency that summarize all of its interactions with technology companies related to human rights.

    – The optimal implementation of its duties to protect from the violation of rights and freedoms by business enterprises by taking appropriate steps to prevent, investigate, punish and redress such violations through effective policies, legislation, regulations, and judicial rulings; These laws should only be adopted after an inclusive, participatory process of consultations with all stakeholders.

---

15 https://www.partnershiponai.org/

Recommendations for digital technology companies:

– The need for companies to exercise due diligence in observing human rights in order to identify, mitigate and confront violations of rights and freedoms, including by:

– Conducting impact assessments on human rights when developing or modifying its products and services. The impact assessment process should always include consultations with civil society actors and other experts and be endorsed by an accredited external third party with human rights expertise;

– Incorporate the results of impact assessments by taking the necessary steps to:

Increase knowledge and awareness of rights and freedoms by providing training and issuing guidelines for management, employees, and other business-related actors... and supporting research and development of appropriate technological solutions to harassment, disinformation, and online propaganda.

Taking everything mentioned above into account, the National Human Rights Council (CNDH) adopted an approach based on the analysis of 84 documents containing guidelines (issued by various bodies) and categorized the principles contained in coherent groups that allow enumerating the percentage of their presence in these documents. These guidelines concern topics that should be considered in the design and use of systems based on artificial intelligence, based on a human rights-based approach, and the table below summarizes the results of this analysis [17] [18]:

Ethical principles	Occurrence	Themes
Transparency	86,90%	Transparency, explainability, understandability, interpretability, communication, disclosure, demonstration
Justice and equity	80,95%	Justice, equity, coherence, inclusion, equality, equity, (non-) prejudice, (non-) discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-Maleficence	71,43%	Non-maleficence, security, safety, damage, protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	71,43%	Responsibility, accountability, compliance, act with integrity
Privacy	55,95%	Confidentiality, personal or private data
Beneficence	48,81%	Benefits, beneficence, well-being, peace, social good, common good
Freedom and autonomy	40,48%	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment
Trust	33,33%	Trust
Sustainability	16,67%	Sustainability, environment (nature), energy, resources (energy)
Dignity	15,48%	Dignity
Solidarity	7,14%	Solidarity, social security, cohesion

References

[1] Australian Human Rights Commission, "Human Rights and Technology - Discussion Paper," 2019.

[2] Y. Meneceur, L'intelligence artificielle en procès: plaidoyer pour une réglementation internationale et européene, Paris: Bruylant, 2020.

[3] E. R. Goffi, "The importance of cultural diversity in AI ethics," Institut Sapiens, Paris, 2020.

[4] E. R. Goffi, "The importance of cultural diversity in AI ethics," Institut Sapiens, Paris, 2020.

[5] M. I. a. E. V. Anna Jobin, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence,* vol. 1, p. 389–399, September 2019.

[6] S. Fukuda-Parr and E. Gibbons, "Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines," *Global Policy,* no. 32-44, p. 12, 2021.

[7] COMEST / UNESCO, « Étude préliminaire sur l'éthique de l'intelligence artificielle ,» UNESCO, Paris, 2019.

[8] S. U. Houlin Zhao, "AI for Good Global Summit: Now stronger than ever," in *2020 AI for Good Global Summit to scale AI-powered problem solving for global impact*, 2020.

[9] Partenariat mondial sur l'IA, «Lancement du Partenariat Mondial sur l'Intelligence Artificielle par quinze membres fondateurs,» ٢٠٢٠. [Online]. Available: https://www.diplomatie.gouv.fr/fr/politique-etrangere-de-la-france/diplomatie-numerique/actualites-et-evenements/article/lancement-du-partenariat-mondial-sur-l-intelligence-artificielle-par-quinze.

[10] IEEE, "Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems," IEEE, Toronto, 2017.

[11] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بالحق في حرية التجمع السلمي وتكوين الجمعيات - 2019،" جنيف، 2019.

[12] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بتعزيز وحماية الحق في حرية الرأي والتعبير،" جنيف، 2018.

[13] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بالحق في التعليم،" جنيف، 2016.

[14] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بالفقر المدقع وحقوق الانسان،" جنيف، 2019.

[15] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بالحق في الخصوصية،" جنيف، 2018.

[16] مجلس حقوق الانسان، "تقرير المقرر الخاص المعني بالحق في الخصوصية،" جنيف، 2020.

[17] M. Bensalah, "Toward an ethical code of AI and human rights in Morocco," *Arribat – International Journal of Human Rights,* vol. 1, no. 2, pp. 187-203, 2021.

[18] M. Bensalah, Artificial Intelligence and Human Rights: Action plan & recommendations for human rights- sensitive and ethical artificial intelligence, Doctoral Thesis ed., Hawaii: Atlantic International University, 2022.

- V -
FINAL DECLARATION

Declaration of Rabat

Artificial Intelligence and Digital Citizenship:
For an Artificial Intelligence respecting Human Rights

We, national and international experts, participants in the International Seminar on Artificial Intelligence and Human Rights, organized in Rabat, on December 3rd, 2021, by the National Council for Human Rights CNDH, while welcoming the initiative, express our adherence to:

- The recommendations of the annual reports of the National Human Rights Council:
- « Opening a public debate on freedom of opinion, expression, and the press, with the participation of all stakeholders, taking into account the changes in this field, **notably in the digital space and in particular social networking platforms**, for the consecration of the practice of this freedom without violating the privacy of individuals » (2019 annual report);
- « Opening a public debate on **the protection of human rights in the field of technology and artificial intelligence** ... [and] the necessity of taking into account the protection of human rights, including the rights to privacy, personal data protection and security in the design of artificial intelligence applications and algorithms – **Human Rights by design** » (2020 annual report).

- Taking note of the broad and inclusive consultations undertaken by the National Human Rights Council with stakeholders in Morocco for the purpose of:
- Establishing a digital platform (April to December 2020) « taabirat raqmya » (Digital Expressions), which received contributions from experts and Moroccan citizens on the issue of the practice of freedoms in the digital space;
- Organizing workshops (December 2019, January 2020) to exchange with journalists, publishers, unions, associations, and human rights defenders on the protection of all human rights in the digital space;
- Editing a special issue of the National Human Rights Council scientific journal (Arribat) on Human Rights and Artificial Intelligence;
- Holding a national consultation seminar (April 2021) with technology companies,

academics, professional associations, and research centers on human rights protection within the artificial intelligence environment.

- Taking into consideration the international and national initiatives on regulatory frameworks for artificial intelligence based on the protection of human rights presented at the international seminar, and taking note of:
- The recommendations on the Ethics of Artificial Intelligence of the UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology COMEST[16];
- The Global Initiative on the Ethics of Autonomous and Intelligent Systems, IEEE[17] (Institute of Electrical and Electronics Engineers) ;
- The Council of Europe's recommendations Towards regulation of AI Systems (the Council of Europe Ad hoc Committee on Artificial Intelligence CAHAI[18]) ;
- Human Rights and Technology, final report[19], Australian Human Rights Commission;
- University initiatives and academic research (Mohammed VI Polytechnic University, Benguerir, Morocco; Ibn Tofail University, Kenitra, Morocco; National Center for Scientific and Technical Research, Rabat, Morocco; George Washingon University, Washington, USA; International University in Geneva, Geneva, Switzerland; Harvard University, Cambridge, USA; Addis Ababa University, Addis Ababa, Ethiopia; University of Technology Sydney, Sydney, Australia).

We recommend to the National Human Rights Council to:

- Continue its work in protecting human rights in the artificial intelligence environment;
- Expand the broad national consultations with stakeholders, in order to propose guiding principles regulating the development of artificial intelligence for the respect of human rights;
- Share its findings and recommendations with its partners and public opinion at the

---

16   https://en.unesco.org/themes/ethics-science-and-technology/comest

17   https://standards.ieee.org/industry-connections/ec/autonomous-systems.html

18   https://www.coe.int/en/web/artificial-intelligence/cahai

19   https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021

regional and international levels.

Welcoming the National Human Rights Council's commitment to organize an international colloquium, in Morocco, in July 2022, we hereby declare the constitution of a scientific committee to advise in the implementation of the above recommendations.

Composition of the scientific committee:

Chairperson:

● Amina Bouayach, Chairperson of the National Human Rights Council;

Members:

● **Peter-Paul Verbeek,** Chairperson of the UNESCO World Commission for the Ethics of Science and Technology (COMEST), Professor at Twente University, The Netherlands;
● **Gregor Strojin,** Chair of CAHAI at the Council of Europe, Magistrate at the Supreme Court, Slovenia;
● **John C. Havens,** Executive Director of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, Member of the World Economic Forum Global Future Council on Human Rights and Technology, USA;
● **Elizabeth D. Gibbons,** Chairperson of the Sustainable Development Committee, IEEE Ethically Aligned Design, Senior Fellow and Director of the Child Protection Certificate Program at Harvard University, USA;
● **Amal El Fallah Seghrouchni,** former member of COMEST, Head of AI UMP6P/AI Movement, professor UM6P, Morocco, Paric La Sorbonne, France;
● **Edward Santow,** Industry and Responsible Technology Professor at the University of Technology Sydney, Former Australia's Human Rights Commissioner, Australia ;
● **Alina Bârgăoanu,** Professor, Dean of the College of Communication and Public Relations, National University of Political Studies and Public Administration, Romania ;
● **Workineh Kelbessa,** former member of COMEST/UNESCO (2012-2019), Professor of Philosophy, Addis Ababa University, Ethiopia;
● **Sidi Mohammed Drissi Melyani,** Director General of the Digital Development Agency (ADD), Morocco;
● **Jamila El Alami,** Director of the National Center for Scientific and Technical Research (CNRST), Professor, Morocco;

● **Azzeddine El Midaoui,** President of Ibn Tofail University, Morocco;

● **Mohammed Rhachi,** President of Mohammed V University, Morocco;

● **Bouchta Moumni,** President of Abdelmalek Essaadi University, Morocco;

● **Leila Hanafi,** Professor at George Washington University, USA;

● **Narjis Hilale,** Professor at the International University in Geneva, Author, Member of Morocco's Special Commission on the Development Model, Switzerland;

● **Younes Alami,** Vice-President of the National Federation of Electricity, Electronics and Renewable Energies FENELEC, Morocco;

● **Mohamed Douyeb,** President of the Digital Act Collective (NGO), Morocco;

● **Raja Bensaoud,** Professor of Universities, co-founder of Digital Act, Morocco;

● **Saida Belouali,** Professor, ENSAO, Mohammed Premier University, Morocco;

● **Salah Baina,** Professor, ENSIAS, Mohammed V University, Morocco;

● **Abdelmajid Elouadi,** Professor, ENSAK, Ibn Tofail University, Morocco;

● **Jean-Pierre Noël Llord,** Professor, Ecole Centrale de Casablanca, Centrale Supélec Paris, Morocco, France;

● **Malik Boumediene,** Professor, UM6P, Morocco;

● **Bouchentouf Toumi,** Professor, Mohammed Premier University, Morocco;

● **Mounir Bensalah,** Secretary General of the National Human Rights Council (CNDH Morocco);

- VI -
APPENDIX

Scan a QR code to access the seminar's panelists presentations

Jamila El Alami
Bibliometric studies

Khouloud Abejja
Digital Development in Morocco

Emmanuel Goffi
Cultural Pluralism as an Essential Part of the Ethical
Assessment of AI

Scan a QR code to access the seminar's panelists presentations

Narjis Hilale
The evolution of AI and its impact on Women

Workineh Kelbessa
Technology, the Environment and Ethics

Alina Bargaoanu
Disinformation 2.0 protecting the integrity of the
information eco-system in a technologized world

Scan a QR code to access the seminar's panelists presentations

Peter-Paul Verbeek
COMEST UNESCO a comprehensive global
standard-setting instrument

John C. Havens
IEEE Global Initiative Responsible innovation and
ethically aligned design

Elizabeth D. Gibbons
In search of accountability, A human rights critique of
Guidelines for 'Ethical AI'

المجلس الوطني لحقوق الإنسان
ⵜⴰⵎⴰⵡⴰⵙⵜ ⴰⵏⴰⵎⵓⵔ ⵉ ⵉⵣⵔⴼⴰⵏ ⵏ ⵓⴼⴳⴰⵏ
National Human Rights Council

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS

Organization, ethics and guiding principles
an international benchmark

Proceedings of the International Seminar
Organized December 3rd, 2021