TRANSLATION FROM THE ORIGINAL SUMMARY IN SPANISH

Seminar 'Digital Footprint: Servitude or Service?'

Data exploitation and scientific truth

Summary of the session of January 21, 2021

On January 21, the expert committee of the Seminar 'Digital Footprint: Servitude or Service?' held its eighth session by videoconference on the exploitation of big data and their relationship with scientific truth. Essential to the debate is the interdisciplinary approach of the Seminar, with the differences in approach that this implies.

The session began with a presentation by **Alfredo Marcos**, Professor of Philosophy of Science at the University of Valladolid. It continued with comments from **Sara Lumbreras**, researcher at the Institute of Technological Research of the Universidad Pontificia Comillas, and **Moisés Barrio**, Lawyer of the Council of State and Director of the Diploma of High Specialization in Legal Tech and digital transformation (DAELT) of the School of Legal Practice of the Complutense University of Madrid. This was followed by an extensive discussion involving all those present (list of participants attached).

With reference to the discussions in previous sessions and the general theme of the Seminar, Alfredo Marcos started from John Dewey's position on the interactive human experience: we leave our mark on digital technologies, and this process leaves its mark on us. The technique is not neutral in absolute terms: it is necessary and is part of human life. At each stage of development, we choose between lines of technological development, some better than others. Each new technology changes our capabilities, but it also changes our needs: servitude and service go hand in hand, and some servitudes are not acceptable. For the speaker, the answer lies in an attitude of serenity and a certain detachment towards the technological, rejecting those developments inspired by hypotheses and objectives that have devastating effects on human life.

Ontology and the denomination “Artificial Intelligence”

In 1956, computer scientist John McCarthy coined the term "Artificial Intelligence" (AI) at the famous Dartmouth conference. What that prophetic vision supposed in terms of calculation power, classification and analysis has been more than fulfilled. But is there something that can be strictly called "artificial intelligence"? In Alfredo Marcos's view, what is intelligent about AI is put by humans (it is not artificial), and what is artificial about it is not intelligent. Simulating intelligence is not being intelligent. In any case, we should talk about *AI systems*, of which those who design and use machines and their programs are part.

Today, several authors denounce an "AI myth," and science fiction theories about "strong AI" divert the debate from the current reality, in which functions are being delegated to machines, sometimes without the necessary controls.

The term AI can be misleading. Different names have been proposed to replace it, such as assisted or extended intelligence or human-centered artificial intelligence. Alfredo Marcos proposes to talk about *delegated control systems* (CoDe). Sara Lumbreras prefers to talk about *decision support systems*. Other Seminar participants refuse to enter into a discussion of terminology and ask for facts to be discussed.

Still, the question is not only one of language, the philosophers insist, but of the conception of reality: the machine, without a human person to interpret it, is nothing more than matter undergoing physical changes. The confusion that the term AI can lead to has been fueled by the media, contributing to the technology's mythologization based on AI's promises and not so much on its current applications.

If the concept of intelligence is taken as the capacity to understand and solve problems, it must be accepted that the term applied to a machine is nothing more than an approximation: devices are not capable of understanding or perceiving problems; the problems to be solved are always human.

For those who criticize the use of the term AI, this nomenclature devalues intelligence and generates a dualistic conception, as if in the human condition, intelligence could be separated from the person. Furthermore, categorizing a system as "smart" tends to justify the delegation of responsibility. In other words, if it is the "intelligence" of the algorithm that has made a decision, there will be a tendency to attribute it consequences that are considered inevitable. If, on the contrary, it is understood that the machine does not have a real experience of the data that have been introduced and

that these data have no value outside of human intelligence, the necessary conclusion is reached that AI cannot be anything other than a help, an additional support to take the adequate decision.

Take, for example, the case of numbers: machines cannot add two plus two since they do not have the notion of number or quantity. A few decades ago, engineers learned to program from the basics; thus, they understood that in a machine adding one to a given number is achieved by changing its last bit (if it is 1, it is set to 0, and if it is 0 it is set to 1, in the first case the previous bit is changed recursively). This essential operation takes infinitely complex forms, but it is vital to understand that there are no numbers in machines, only ordered memory locations that store ones and zeros. That is what bolsters the computer science edifice. But programmers and data scientists who currently use AI models for decision-making have not always known the most basic level of these models and use instead software that simplifies processes. For this reason, it is sometimes more difficult for them to see that inside the machine, there is no knowledge, no concepts. Having this clear is the basis for knowing what to expect and what can be delegated to AI models. AI does not understand data; it just manipulates them. To say that it understands them would be like defending that a cockatoo, which repeats sounds that mimic human speech, is truly capable of communicating a piece of knowledge, a feeling, or a thought of its own.

How machines work

Beyond the terminology, it is undoubtable that there are in fact systems, known as AI, which are used as management tools. To decide rightly in automation, one needs to understand how these systems work, what kind of help they provide in decision making, and most importantly: establish their application limits. In short, AI algorithms perform their functions by detecting lines of correlation between vast amounts of data. But they do not discover causal relationships. That is, they do not allow us to understand or explain the phenomenon. The processes of human intelligence can be categorized into induction and deduction, to which Charles S. Peirce (1839-1914) added the category of abduction, that is, the creative step of formulating conjectures and hypotheses. The algorithm proceeds by induction, not by deduction, and even less by abduction.

Algorithms work thanks to vast amounts of data; in the data history, algorithms find patterns that apply to new scenarios on which to decide. It would not be surprising if, in

some instances, depending on the data composition, the machine reaches results which are apparently correct, but really false. It is as if a student were learning only through examples of solved problems, in which she finds a logic that she later applies to solve new problems. A typical situation of induction problems occurs, in which all the premises are true, but the conclusion may be false because the context stability rule ("rebus sic stantibus") is not fulfilled. As with any diagnostics, clinical judgment is required to know whether or not a conclusion is valid. Therefore, in changing reality, all models have a degree of uncertainty. The algorithm cannot predict (speaking in the future tense), it can only establish possible models (speaking in the conditional). The algorithm proposes what something could be if certain conditions are met, based on previous experience.

Although, in many cases, AI works correctly and is a useful tool, its way of proceeding can present several typical problems. Among these problems is possible *overfitting*. This occurs when the provided data is not sufficient for it to be generalized; then the algorithm is forced and extracts patterns that cannot be extrapolated. It is as if the student has too few examples to make generalizations and to understand a whole new context. This error, difficult to detect, is salvageable when the rules that the algorithm has derived can be accessed and audited. But this is not the situation in all cases, machines often work as a "black box": not only does the AI not understand the conclusion it has reached, but it also does not allow us to know how it has arrived. That is to say, although it is possible to examine the initial code with which it works, the path followed by the algorithm to make a decision is not understood, it remains hidden, often for commercial or intellectual property reasons. In these cases, reliance on AI as a decision-making tool becomes complicated: how can we trust something that we are not capable of understanding? Something that hides its way of proceeding from us? To these questions from some participants, others opposed that it is also not fully understood how the human brain proceeds, despite many centuries of increasingly deep psychological and neurological research: this by itself does not detract from the legitimacy of human decisions.

Another recurring problem in AI is *algorithmic bias*. The issue of bias appears when the algorithm is based on variables considered inadequate. The best-known example of algorithmic bias issues is that of COMPAS (Correctional Offender Management Profiles for Alternative Sanctions), a black box-type algorithm used in the United States to predict criminal recidivism in prisoners. This algorithm, developed by Northpoint, Inc., has been used for years to decide whether or not inmates are granted parole. After a detailed analysis of its predictions, the ProPublica news agency found that the algorithm

used the prisoners' race as one of the main variables. Thus, African American prisoners, regardless of their personal history, received worse predictions than whites. This stemmed from the database used to nurture the algorithm, in which prisoners of color had a worse recidivism rate. All this results in a case of racial injustice.

Examples like this there are many. The solution to these problems could perhaps come if the algorithms allowed us to see the rules on which they infer, if they stopped being black boxes. If the variables are transparent and AI application is clear, there is less danger of falling into these types of problems, both in overfitting and algorithmic bias. But it is not easy to apply this idea. In the real experience of data scientists, only complex models, which the user cannot understand, are those that have the possibility of working adequately in real situations. On the other hand, said the operators, if we try to audit and explain the systems, they lose their effectiveness and their reason for being, since all the time that is intended to be saved would be lost again. AI models are developed to fulfill specific tasks, and if they had to be made entirely transparent, auditable, or explainable, they would be unable to achieve those objectives.

The debate does not provide any definitive answer on this. If many existing algorithmic processes cannot be audited and made transparent, it should be possible, however, to project new models that seek to be understandable from the start. To do this, developers should consider transparent variables and have interdisciplinary teams to achieve a genuinely context-appropriate design. This is already a fact in many fields of AI application: for example, when developing algorithms for medical diagnostics, no one thinks of having only engineers, you need to have medical specialists in the branch in question. The data patterns are only correlations; only an expert can relate them to known phenomena and new testable hypotheses. Transparency and intelligibility of processes do not refer to the algorithm only; above all, it is about making clear the aims to be achieved and the data with which one works. The darkness that remains around AI has its roots, not so much in the necessity of the processes themselves, but rather in commercial strategies of a monopolistic nature - those of a few large companies in the United States - or in political strategies - in the opinion of the speaker, maybe those of the Chinese government.

On the current use of AI

The approach to an ethical stance towards the development and application of AI models will be based on the development by public administrations of practices and legislation that cover the aspects mentioned here. From this perspective, Moisés Barrio proposes a positive approach to AI in an "algorithmic rule of law," in which the technological automation tools would be adopted only when they represent an improvement and when they do not affect the legitimacy of democracy.

Although it sometimes seems like a project for the future, the debate about the use of AI is a current issue. Various tasks and functions are delegated to AI models daily. Such is the case of the stock market, in which almost all investment decisions have already been delegated to algorithms. Another example is what happens with each person and their mobile phone. Many decisions have already been delegated, for instance: to take routes on maps' apps or restaurant's choices. It is challenging to be contrary to AI when deciding in many aspects, this can already be seen on a day-to-day basis. Hence, personal training in virtues is essential, something that goes much further than complying with deontology, that is, with rules that people do not usually make their own, but which they apply without the intervention of their freedom and conscience.

Through an ethic of virtues, and not a deontology of standards for a passive user, one can save society from the bondage of technology and understand the proper service of tools, such as AI. As has already been said, all technology changes our capacities and needs; thus, servitude and service go hand in hand. But it is necessary to analyze this in a dynamic context: the service and the servitude do not always affect simultaneously, and in many cases, those who benefit are not the same as those who bear the servitude. An "algorithmic rule of law" also implies access for all to digital media, as was discussed in the previous session, and this is far from being the case in today's society. Therefore, it is necessary to apply a realistic moral discernment, not taking as necessary everything that is technically possible and seeking, however, to put the available means at the service of human development.

In other words, a critical judgment is necessary to demystify AI; to not underestimate or overestimate it, since we cannot renounce it. We must select those technologies that improve rather than undermine human life.

Attendees:

1. **Albert Cortina**, Lawyer, Expert in Transhumanism Director of the DTUM study
2. **Alfonso Carcasona**, CEO, AC Camerfirma
3. **Alfredo Marcos Martínez**, Professor of Philosophy of Science, Universidad de Valladolid
4. **Ángel Gómez de Agreda,** Colonel Chief, Geopolitical Analisis Area, DICOES/ SEGENPOL
5. **Ángel González Ferrer**, Executive Director, Digital Pontificial Council for Culture
6. **Carolina Villegas**, Researcher, Iberdrola Financial and Business Ethics Chair, Universidad Pontificia de Comillas
7. **David Roch Dupré**, Professor, Universidad Pontificia Comillas
8. **Diego Bodas Sagi**, Lead Data Scientist – Advanced Analytics, Mapfre España
9. **Domingo Sugranyes**, Director, Seminario de Huella Digital
10. **Esther de la Torre**, Responsible Digital Banking Manager, BBVA
11. **Francisco Javier López Martín**, Former Secretary-General, CCOO Madrid
12. **Gloria Sánchez Soriano**, Transformation Director, Legal Department, Banco Santander
13. **Guillermo Monroy Pérez**, Professor, Instituto de Estudios Bursátiles
14. **Javier Camacho Ibáñez**, Director of Ethical Sustainability and professor at ICADE and ICAI
15. **Javier Prades**, Dean, Universidad Eclesiástica San Dámaso
16. **Jesús Avezuela**, General Director of the Pablo VI Foundation
17. **José Luis Calvo**, AI Director. SNGULAR
18. **José Luis Fernández Fernández,** Director of the Iberdrola Chair of Economic and Business Ethics ICADE
19. **José Ramón Amor,** Coordinator, Bioethics Observatory of the Pablo VI Foundation
20. **Juan Benavides**, Professor of Communications, Universidad Complutense de Madrid
21. **Julio Martínez s.j.**, Dean, Universidad Pontificia Comillas
22. **Moisés Barrio**, Lawyer. Council of State and Director of Diploma of High Specialization in Legaltech and digital transformation - Escuela de Práctica Jurídica - Universidad Complutense de Madrid
23. **Paul Dembinski**, Director de Observatoire de la Finance (Ginebra)

24. **Richard Benjamins**, Data & IA ambassador, Telefónica
25. **Sara Lumbreras**, Deputy Director of Research Results, Associate Professor, Institute for Technological Research, ICAI, Universidad Pontificia Comillas