# Considerations of a transparent and explainable AI

The need for an ethical AI requires a multitude of components, as in this working group it has been highlighted, and in my opinion the requirements of transparency and explainability transcend all of them. The industry is the one that has built this new "science" and that, through it, has set the standards that must regulate it. And the absence of transparency and explainability has taken us to the point of legal construction in which we now find ourselves.

However, I believe that science will also regulate itself for greater effectiveness of its development. Thus, and serve as a representative example, in 2015 the Monte Sinai hospital in New York developed an artificial intelligence model called Deep Patient that predicted diseases, in some cases almost infallibly. But since the predictions were not understood, they could not justify them to patients. That is, although it worked, the model could not be trusted in.

Circumstances like this are causing some companies to be already developing the explanation of how their AI work. Thus, for example, Neftlix publishes on its website how its recommendation system works; or Disney, as also said in this forum, is using algorithms to analyze and plot possible biases in its scripts.

However, it is necessary to remember that we already have a right not to be the subject of a decision based solely on automated processing (art. 22 GDPR). And that, in addition, the European Parliament Resolution of 16 February 2017, with recommendations to the Commission on civil law standards on robotics (2015/2103 (INL)), states that the principle of transparency:

"Is that it must always be possible to justify any decision that has been taken with the help of artificial intelligence and that may have a significant impact on the life of one or several people; believes that it should always be possible to reduce the calculations of the artificial intelligence system to a form understandable to humans "

However, this should be a starting point in the legislative development that I intend to propose, focusing only on section 4 of the text of the High Level Expert Group, on transparency. And this analysis must be connected with being aware that, within the race for the hegemony of AI, a restrictive regulation would reduce the competitive capacity of companies in the European Union. So, trying to equalize that weak balance, the previous criteria that should prevail, I believe may be the following:

- To establish a prudent period of regulatory adaptation (between one and two years).

- The requirement of the AI regulations established must be preceded by the compliance made by the Public Administrations and Institutions. They should set the standard on which the private sector should work.

- To define legally what is transparent AI and explainable AI.

- To establish qualified transparency and explainability requirements whose voluntary compliance is encouraged by various means: tax bonus, requirement to collaborate with public administration and others that may be considered.

- To demand compliance within a shorter period (the first year, for example) to companies that receive subsidies or collaborate / work for public administration, as well as for the public sector.

In a second aspect, two regulation modalities would have to be differentiated, depending on whether the algorithms are public or private sector. And within the private sector, the algorithms will have to be categorized by scope of decision, modulating the requirements accordingly because, as practice is demonstrating, design is largely the expression of intentions.

The group of experts has already established the basis for these categories by indicating that an impact-based approach is adopted and that the less human supervision, the more evidence and stricter the governance should be.

However, and as a first step, the requirements that the public sector must meet should be established in a closed way, among which the following would be highlighted:

- ✓ Public libraries of algorithms that are easily accessible to all.

- ✓ Subject to audit of the models and data used for automated decisions.

- ✓ Accessibility by citizens to the mechanisms of operation of an algorithmic system.

- ✓ Source code advertising.

- ✓ Administrative transparency, that is, the right to citizen explanation in the non-public grant of rights. The doctrine on this subject exposes, among others, the so-called counterfactual explanations. For example, if I am denied public aid or subsidy, the system should motivate the requirements met and those not met.

- ✓ Judicial transparency, so that judicial decisions must be motivated, explained and auditable.

- ✓ Legal control of algorithms evaluating the need, where appropriate, for a specific public supervisory body.

And in a third section, all the public and private sector have to follow a series of criteria such as:

- An AI diverse in terms of gender, ethnicity, religion, age or any other personal or social circumstance.
- Availability for the user of the type of personal and non-personal data used by AI systems.
- Right to know how patterns are determined by AI.
- The access to the purpose for which this data is used.
- To ensure a level of understanding of how the conclusions reached by the AI have been reached.
- When an error is verified in the AI, the transparency in the access to the establishment of the models that produced it.
- Side with all of the above, that the success or otherwise of the decision can be specified or measured with some variable. All this is a development of the transparency of the business model required by the High Level Expert Group´s document.
- And finally, as also mentioned in other forums, that the objectives of the creators of the algorithms are aligned with the objective of the customers.

Being aware of the complexity of the analysis of the black boxes of AI and that statistical correlations do not reflect causes and effects, we must be aspirational of how far we can go so that the law can define the framework in this revolution that is already so present in our lives.

Yours sincerely, Antonio J. Cabrera

CABRERA RODRIGUEZ ANTONIO JOSE - 74928077B

Firmado digitalmente por CABRERA RODRIGUEZ ANTONIO JOSE - 74928077B
Fecha: 2019.11.30 19:15:11 +01'00'

Independent Researcher for Achievement of Human Rights and Artificial Intelligence

antonioj.cabrera@icamalaga.org

Spain