ARTIFICIAL INTELLIGENCE IN SCIENCE: SHUT UP AND CALCULATE

Ramón Alberto Carrasco, Mario Arias-Oliva, Jesús Serrano-Guerrero, Francisco Chiclana

Universidad Complutense De Madrid (Spain), Universidad Complutense De Madrid (Spain), Universidad Castilla-La Mancha (Spain), De Montfort University Leicester (United Kingdom)

ramoncar@ucm.es; mario.arias@ucm.es; jesus.serrano@uclm.es; chiclana@dmu.ac.uk

EXTENDED ABSTRACT

Quantum mechanics is astonishing, both in terms of its accuracy and its interpretation. Not even geniuses like Albert Einstein could imagine the underlying reality within this theory, which, on the other hand, has formulations that are fulfilled with very high precision. In order to reach a consensus on its interpretation, the best physicists came together between 1925 and 1927, giving rise the so-called "Copenhagen interpretation". However, for many authors, this interpretation implies a refusal to grasp the truth. Thus, David Mermin (1989) coined the phrase "shut up and calculate!" to summarize this Copenhagen interpretation or, rather, attitude. This sentence summarizes the approach of many scientists who apply this accuracy theory mechanically, without considering anything else regarding its interpretation.

On the other hand, we are living a boom of the artificial intelligence (AI) that lives its "golden spring". This is mainly due to the advances of the machine learning algorithms, which learn from vast amounts of data, produce at very high velocity, in various structured and non-structured forms (including audios, videos, images, natural language, etc.). Among these algorithms, the ones that do the best job are the so-called black box algorithms, which are not interpretable by humans, including deep learning based on artificial neural networks (Das et al., 2020). Therefore, the situation we are facing involve the delegation of decision-making in favour of AI, even although these decisions are based on algorithms that are non-compressible for us (Marín & Carrasco, 2021). The use of these black-box algorithms is spreading to all fields: military, healthcare, education, finance, business, marketing, science, etc.

Indeed, the scientist world is not an exception and, more and more this kind of algorithms are being used in researches. The reflection is as follows: Are scientists in general, particularly those not specialized in AI, aware of the non-interpretability of the outcomes generated by this type of AI?

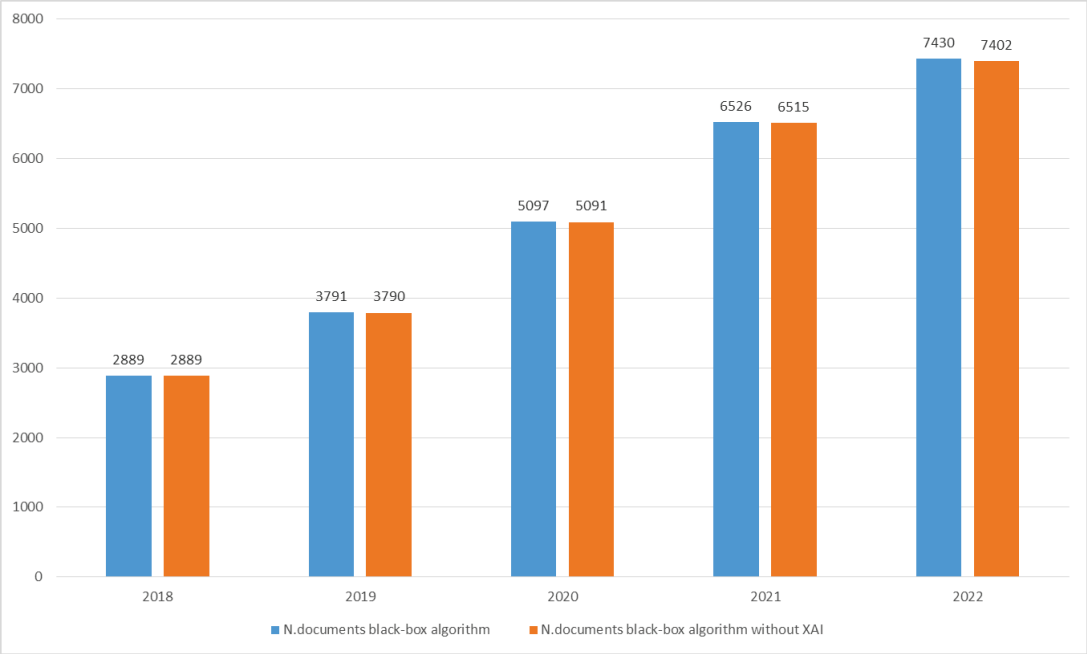
The field of explainable artificial intelligence (XAI) emerged with the aim of making the results of these black-box algorithms more interpretable (Monje et al., 2022; Marín et al., 2022; Gunning & Aha, 2019). So, we could rephrase the previous question: Is the scientific community employing XAI to enhance the interpretability of black-box algorithms?

To help clarify this question, we are relying on the documents published in the Web of Science (WoS) database. While exploring just applications of black-box algorithms, we conduct a query exclusively within the social science domains in the core collection of WoS. In other words, this approach aims to exclude studies within the AI field itself. We are going to conduct the exploration through two paths. Firstly, let us obtain documents that use black-box algorithms in social sciences. In the second place, we add that they do not use XAI techniques. Both queries are displayed in Table 1 and the results are showed in the Figure 1.

Table 1. WOS queries description (*year* between 2018 and 2022).

Query description	**Query**
Documents using of black-box algorithms in social sciences	TS = (Neural Networks OR Deep Learning OR Support Vector Machine* OR Random Forests OR Gradient Boosting Machine* OR Extreme Learning Machine* OR Kernel Machin* OR Kernel-based Learning Machine* OR Long Short-Term Memory OR Ensemble Classification Models OR Emsemble model) AND PY=(*year*) AND WC=("Agricultural Economics & Policy" OR Anthropology OR "Area Studies" OR Art OR "Asian Studies" OR "Behavioral Sciences" OR Business OR "Business, Finance" OR Classics OR "Construction & Building Technology" OR "Criminology & Penology" OR "Cultural Studies" OR Dance OR Demography OR "Development Studies" OR Economics OR "Education & Educational Research" OR "Education, Scientific Disciplines" OR "Education, Special" OR "Environmental Studies" OR "Ethnic Studies" OR Folklore OR Geography OR History OR "Humanities, Multidisciplinary" OR "Information Science & Library Science" OR "International Relations")
Documents using of black-box algorithms in social sciences without XAI	TS=(Neural Networks OR Deep Learning OR Support Vector Machine* OR Random Forests OR Gradient Boosting Machine* OR Extreme Learning Machine* OR Kernel Machin* OR Kernel-based Learning Machine* OR Long Short-Term Memory OR Ensemble Classification Models OR Emsemble model) NOT TS = (XAI OR “Explainable Artificial Intelligence”) AND PY=(*year*) AND WC=("Agricultural Economics & Policy" OR Anthropology OR "Area Studies" OR Art OR "Asian Studies" OR "Behavioral Sciences" OR Business OR "Business, Finance" OR Classics OR "Construction & Building Technology" OR "Criminology & Penology" OR "Cultural Studies" OR Dance OR Demography OR "Development Studies" OR Economics OR "Education & Educational Research" OR "Education, Scientific Disciplines" OR "Education, Special" OR "Environmental Studies" OR "Ethnic Studies" OR Folklore OR Geography OR History OR "Humanities, Multidisciplinary" OR "Information Science & Library Science" OR "International Relations")

Figure 1. Documents using of black-box algorithms in social sciences.



Source: self-elaboration based on WOS (2023)

These results indicate that most of the research in social science using black-box AI does not consider the interpretability of this AI. The primary purpose of utilizing this AI is predicting certain variables. However, few inquire about the interpretation of these black models. In addition, as expected, it is possible to deduce from Figure 1 that the usage of this type of AI is increasing. Many scientists are convinced of the effectiveness of these algorithms primarily due to the accuracy of the predictions but they do not worry excessively about the interpretation. Once again in the history of science, we are moving towards the "shut up and calculate!" approach.

We must emphasize that the non-interpretability of AI could represent a neglect of another important aspect related to AI research: responsibility, bias, fairness, ethical considerations, etc (Giovanola & Tiribelli, 2023; Yapo & Weiss, 2018). To the extent that these aspects are essential, in many cases, within the field of science, we must be demanding in our endeavor to comprehend this black AI. Put more pragmatically, in the utilization of XAI techniques. While these techniques are not flawless, they at least assist us in approaching the interpretation of the previously mentioned aspects.

KEYWORDS: XAI, black-box algorithms, interpretable AI, ethical AI.

ACKNOWLEDGMENTS: The co-author Ramón Alberto Carrasco González was funded by the Madrid Government (Comunidad de Madrid-Spain) under the Multi-annual Agreement with Universidad Complutense de Madrid in the line Excellence Programme for university teaching staff, in the context of the V PRICIT (Regional Programme of Research and Technological Innovation).

This work was also supported by the Project PID2019-103880RB-I00 funded by FEDER funds provided in the National Spanish projects, the project PID2022-139297OB-I00 funded by MCIN / AEI / 10.13039/501100011033 and by "ERDF A way of making Europe".

REFERENCES

Das, S., Agarwal, N., Venugopal, D., Sheldon, F. T., & Shiva, S. (2020, December). Taxonomy and survey of interpretable machine learning method. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 670-677). IEEE. Retrieved from https://drsaikatdas.com/papers/taxonomy.pdf

David Mermin, N. (1989). What's wrong with this pillow? Physics Today, 42(4), 9-11. Retrieved from https://doi.org/10.1063/1.2810963

Giovanola, B., & Tiribelli, S. (2023). Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. AI & society, 38(2), 549-563. https://doi.org/10.1007/s00146-022-01455-6

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. AI magazine, 40(2), 44-58. https://doi.org/10.1609/aimag.v40i2.2850

Marín, G. & Carrasco, R. A. (2021). What Do Machines Think About? In [New] Normal Technology Ethics: Proceedings of the ETHICOMP 2021 (pp. 129-133). Universidad de La Rioja. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=7977270

Marín, G., Galán, J. J., & Carrasco, R. A. (2022). XAI for Churn Prediction in B2B Models: A Use Case in an Enterprise Software Company. Mathematics, 10(20), 3896. https://doi.org/10.3390/math10203896

Monje, L., Carrasco, R. A., Rosado, C., & Sánchez-Montañés, M. (2022). Deep learning XAI for bus passenger forecasting: A use case in Spain. Mathematics, 10(9), 1428. https://doi.org/10.3390/math10091428

Yapo, A., & Weiss, J. (2018). Ethical implications of bias in machine learning. Retrieved from https://scholarspace.manoa.hawaii.edu/bitstreams/d062bd2a-df54-48d4-b27e-76d903b9caaa/download