

A metrological approach to ethical and legal issues in Artificial Intelligence

Alessandro Ferrero
DEIB
 Politecnico di Milano
 Milano, Italy
 0000-0002-6760-2925

Veronica Scotti
DEIB
 Politecnico di Milano
 Milano, Italy
 0000-0002-5582-0108

Abstract—Artificial Intelligence has developed in an impressive way during the recent years, and is now being applied to almost every field of human activities, slowly replacing human beings in operations whose level of required skills has significantly increased. Collaborative robots, or cobots, are a reality in industrial production, as well as virtual reality and robots driven by human motions from remote sites allow operators to control operations in dangerous areas. AI algorithms perform data searches and present the results in a very efficient way, so that they are helping decision makers in critical fields, such as medicine and justice. This poses new and somehow unforeseen ethical and legal problems that must be covered to avoid generating wrong or even illegal results. Some of these wrong results might be generated by the use of input data that might not be sufficiently accurate, especially when they are collected from the field, or whose limited accuracy is not properly considered when processing them. This paper aims at considering a possible, metrologically-sound approach to ethical and legal issues met in AI.

Index Terms—Artificial Intelligence, Big Data, Reliable data, Uncertainty, Ethics

I. INTRODUCTION

Nowadays, Artificial Intelligence (AI) is employed in almost every field of human activities and represents an extremely useful tool in assisting operators in activities in which some level of autonomy is required, and decisions have to be taken [1]. The name – AI – given to this specific branch of computer science assumes that it might assist, if not even replace, human intelligence somewhen in the future, although there is no general consensus upon this perspective [2].

As a matter of fact, the idea that human individuals might be replaced in some of their activities opens ethical problems, related not only to bad or good use of this technology (as the recent debate on ChatGPT has started), but also about the technology itself and how far it is supposed to go [3]. It opens also legal issues, related, on one side, to the legal entity that autonomous devices might assume and who is liable for possible damages [4] or violations [5] and, on the other side, to the direct use of AI in legal proceedings, in assisting the triers of fact [6].

While considering this issues, it is also worth considering the close connection between AI techniques and BigData [7], [8], since most of the incorrect results returned by the AI algorithms can be originated by incorrect or not fully characterized data used either in the training step or in the

subsequent validation step, as well as by not having considered some operative conditions that might become critical.

This paper aims at approaching these points under a metrological perspective, and showing how it can be usefully exploited to find a possible solution to ethical and legal issues. The following sections will, hence, propose how to frame ethical and legal issues under this perspective.

II. ETHICAL ISSUES IN AI

Ethics in science and technology has always been a controversial matter, has been influenced by religion and political ideologies, and has been approached under different perspectives, with authors involving moral issues and identifying it with moral [9] and others denying moral influence on ethics [10]. The debate is so far from a general consensus that there are authors that, in a rather provocative way, deny the possibility of even defining ethics, stating that this might lead us to “*run against the boundaries of language*” [11].

Up to a recent past, it was possible to assume that, when referring to scientific and technical activities, moral implications were more related to individuals, their behaviour and their actions, while ethics was rather applied to the categories of professionals and was implied in the profession practice, providing guidance to avoid harm to other people, also in a future perspective, for instance by protecting the existence of future generations not endangering the environment or wasting precious natural resources.

Anyway, regardless of the philosophical way this critical subject was approached, the focus has always been on human beings, their actions and possible consequences, since they were considered the sole entities with a developed conscience and, consequently, capable of foreseeing the adverse consequences of their actions not only on the individual who committed them, but also on the other individuals and, more in general, to the human kind they belong to.

As the body of knowledge increased, it became more and more evident that the possible outcome of a new theory, or a new device based on this theory, was also based on the available knowledge. It is also worth noting that, when an innovative theory and technology bursts out, the related available knowledge is, by its very nature, rather limited. In this respect, the conclusion drew by Johnson and Powers [9] when

discussing the “*moral agency of technology*” appears quite interesting, although rather alarming. They stated that “*An artifact ceases to be useful (or even sometimes safe) when its output is unpredictable (except, again, when unpredictability is the designed output), and this is exactly what happens when a user gets different outputs for the exact same input on different occasions*”.

While such a statement may be viewed as plausible and attractive, it appears also to limit the development of new technologies, such as AI, whose output has some natural degrees of unpredictability, but also, if it is strictly applied, it might restrict the use of well known devices, such as the measuring instruments, since, due to measurement uncertainty, they seldom return the same value when the *exact same input* is given to them. Moreover, assuming that an artifact may cease to be safe when its output is not fully predictable might also raise unjustified alarm, especially in laypersons, about new technologies, thus preventing their study and, consequently, a deeper knowledge of their effects and a careful weighing of the possible adverse effects against the beneficial ones.

An alluring and more modern response to such concern is Miké’s approach [12] to ethics in science and technology: it is recognized that uncertainty is always implied in every scientific and technical activity and, therefore, ethics cannot be based on a fully certain ground, but only on the available evidence. It is then possible to refer to the *ethics of evidence* [12] and consider that, when practising a profession, an individual must refer to the available knowledge and evidence, and should not disregard any of the pieces of information that available knowledge and evidence may provide him or her.

It is worth noting, once again, that the focus is still on the human beings, although it opens a rather interesting question on what should be considered as *knowledge* and *evidence* and whether it can be considered reliable or not.

There is little doubt that, nowadays, the major source of information are the so-called Big Data and that the largest body of knowledge related to the different fields of human knowledge is represented by Internet. The Big Data are also the source of information for the most modern application of AI, such as the various *chatbots*, including ChatGPT, that are capable, in response to a question formulated in natural language, of searching the relevant data of interest, process them and provide a generally exhaustive answer in different languages, showing an appropriate choice of terms, correct grammar and syntax.

The way the answer is formulated, in a semantically correct way, makes it appear credible, regardless of its trueness, thus transferring some of the typical ethical issues related to human beings to these new entities created by the AI. Therefore, considering that the correctness of AI results is largely based on its ability to search the Big Data and find useful correlations, the importance of *feeding* AI algorithms with correctly validated data and its implications with ethics become quite evident, and the ethics of evidence might represent the correct approach to deal with it.

Here is where metrology comes into play. Metrology is the science of measurement and covers two equally important points: the analysis and development of measuring methods and instruments, on one side, and the characterization of the obtained measurement results, in terms of *measurement uncertainty*, on the other side [13]. This is a critical point since it is widely recognized that, when all of the known or suspected components of error have been evaluated and the appropriate corrections have been applied, there still remains an uncertainty about the correctness of the stated result, that is, a doubt about how well the result of the measurement represents the value of the quantity being measured [14]. Measurement uncertainty represents the universal tool to evaluate and express such a doubt, according to the recommendations given by the Guide to the expression of Uncertainty in Measurement (GUM) [14].

It can be readily understood that any measurement result becomes useless if such a doubt is not quantified [13], and hence the relevant information carried by measurement uncertainty shall be declared, according to Miké’s theory [12], to avoid hiding an important part of the available evidence. Not doing so might cause wrong conclusions and wrong decisions, when the decisions are based on measurement results [15].

It is therefore imperative that AI applications consider only data that have been validated by declaring uncertainty and that no data are included in the Big Data without a proper metrological validation [7]. While this is a quite obvious conclusion when data are obtained through experimental activities involving measurements, it may appear quite difficult – if not impossible – to achieve when the considered data do not come from experimental activities.

It is worth noting that the method proposed by the GUM to evaluate and express uncertainty is *universal* [14] and can be applied also to evaluate the reliability of data obtained by means other than experiments [16], [17]. In particular, the evaluation of the source reliability and the identification of possible, often hidden correlations between the different sources can be performed by borrowing the methods employed in metrology, that can indeed become a universal method to estimate the reliability of all considered data.

Moreover, the method recommended by the GUM [14] to combine the different contributions to uncertainty can be usefully extended to evaluate uncertainty on any result returned by AI applications, including the answers provided by the more recent and evolved chatbots. If such methods are implemented inside the AI applications, so that they can self-evaluate uncertainty of the provided result, they will give the final user a terrific tool to evaluate the reliability of the obtained result: from the perspective of the ethics of evidence [12], this appears as the most effective way to ensure that the whole available evidence has been considered, so that unethical use of these same results can be avoided.

III. LEGAL IMPLICATIONS OF AI

As mentioned in Section I, AI implies also legal issues, both in a direct and indirect way. The direct implications are those

related to possible privacy violations [5] and those related to assessing the liability for damages, or even deadly accidents caused by devices or machines – such as the autonomous vehicles – driven by AI algorithms [4].

The indirect implications are those related to the use of AI algorithms in courtrooms [6], both in helping the trier of fact to find similar cases and pertaining jurisprudence, and expert witnesses in ascertaining facts, or assessing liability, based on scientific evidence.

In both cases, the validity of the data processed by the AI applications to provide the desired solution takes great importance. In this case, the always present ethical implications go side by side with typical legal issues. It can be proven that, if the answer provided by the AI algorithms is based on data whose uncertainty has not been properly considered, these same answer can be seen in a similar way as false testimony rendered by an eyewitness who is, even in perfect good faith, hiding part of the facts he or she has witnessed [18]–[20].

Therefore, once again, a metrological analysis of the considered data may greatly help the trier of fact in evaluating the “reliability” of the the results provided by the AI application, in the same way as he or she can evaluate, based on his or her experience, the reliability of an eyewitness [21].

When AI is used in courtrooms [6], checking the reliability of the sources and whether they have been brought up-to-date is of utmost importance. While this might be quite easy when the pertaining jurisprudence is searched and presented to judges and lawyers, since the original documents can be accessed and checked, it might become a more complex and delicate process if AI applications are employed to interpret scientific results.

In order to provide an example of the present limitations of AI in this field, a real case was presented to ChatGPT and the received answer was compared with the analysis conducted by the expert witness called by the defence and the verdict rendered by the court. The considered case is an Italian DUI (Driving Under the Influence) case, discussed in front of Vicenza’s criminal court, in which the defendant blood alcohol concentration (BAC) was measured through a breath alcohol concentration (BrAC) measurement and found to be 1.62 g/l, while the law limit above which the most severe sanction is applied is 1.5 g/l.

The expert witness analyzed all uncertainty contributions, including the possible breathalyzer drift, and proved that there was a 50.5 % probability that the actual BAC was below the 1.5 g/l limit [22]. The trier of fact recognized the validity of this analysis and applied the less severe sanction provided by the law when the measured BAC lies in between 0.8 g/l and 1.5 g/l.

Let us now see how ChatGPT answered to the following question, that could have been formulated, for instance, by the prosecutor to check whether the defence expert’s considerations were acceptable: “Knowing that the maximum allowed value for BAC is 1.5 g/l, a driver tested with BAC of 1.62 g/l, measured with Lion INTOXILZER breathalyzer shall be

sanctioned for committing DUI crime beyond any reasonable doubt?”

Here is ChatGPT unedited answer¹: *In many countries, including the United States, a blood alcohol concentration (BAC) of 0.08 % or higher is considered illegal for drivers. The exact legal limit may vary depending on the country and the state or province within that country.*

In your question, you mention a maximum allowed value for BAC of 1.5 g/l, which is equivalent to 0.15 % BAC. If this is the legal limit in the relevant jurisdiction, then a driver who tested with a BAC of 1.62 g/l would be above the legal limit and could face DUI (driving under the influence) charges.

Regarding the use of the Lion INTOXILZER breathalyzer, it is important to note that the accuracy and reliability of breathalyzers can be influenced by a number of factors, including calibration, usage, and the presence of certain substances in the person’s breath. It is also possible for the results of a breathalyzer test to be challenged in court.

In summary, if the legal limit for BAC in the relevant jurisdiction is 1.5 g/l, and the driver in question tested with a BAC of 1.62 g/l, this would likely be considered evidence of DUI and could result in legal consequences. However, the accuracy of the breathalyzer results and other factors may also be taken into consideration.

It can be immediately perceived that this answer is very generic and it does not help in quantifying how the considered number of factors influence the accuracy of the breathalyzer. Therefore, the question was focused on accuracy and reliability and reformulated in this way: “I’m interested in your comment about accuracy and reliability of breathalyzers. How can they be quantified?”

Here is the answer: *The accuracy and reliability of breathalyzers can be quantified through various metrics and tests. Here are a few examples:*

- 1) *Calibration: Regular calibration of the breathalyzer is necessary to ensure accurate readings. Calibration involves adjusting the device to a known standard, usually using a gas mixture of a known alcohol concentration. The accuracy of the device can be assessed by comparing the readings it provides with the known standard. If the device consistently produces readings that are within an acceptable range of the known standard, then it is considered to be accurately calibrated.*
- 2) *Precision: The precision of a breathalyzer refers to its ability to produce consistent readings over multiple tests. This can be assessed by conducting repeated tests on the same subject, using the same device and conditions, and comparing the results for consistency.*
- 3) *Interference testing: Interference testing is used to determine whether substances other than alcohol in a person’s breath can interfere with the accuracy of the breathalyzer. For example, certain mouthwashes or medications may contain alcohol or other substances*

¹The bold characters are used by the Authors to highlight a significant part of the answer.

that can give false positives on a breathalyzer test. Interference testing involves exposing the device to these substances and measuring the extent to which they affect the readings.

- 4) *Comparison to other methods:* Another way to assess the accuracy of a breathalyzer is to compare its results to those obtained from other methods, such as blood tests or gas chromatography. If the results of the breathalyzer are consistent with those obtained from other methods, then it is considered to be accurate and reliable.

It's worth noting that even with proper calibration and testing, breathalyzers are not infallible and can sometimes produce inaccurate readings. Factors such as environmental conditions, operator error, and **physiological differences between individuals** can all affect the accuracy of breathalyzer results. As a result, it is important to use breathalyzer results as one piece of evidence in a DUI case, rather than relying solely on them for conviction.

Once again the answer is very inconclusive and also incorrect when it associates an adjustment operation to calibration. However, it mentioned, at least, the *physiological differences between individuals* that are the origin of the definitional uncertainty contribution, which represents the major contribution to uncertainty in BAC tests [22]. Therefore, an additional question was asked, in order to get some more details about this point: "What about the foundational validity of the BrAC methods? Is the conversion factor to BAC unique for every individual and reliable?"

The answer was: *Breath alcohol concentration (BrAC) testing is a widely used method for determining a person's level of intoxication. BrAC testing relies on the assumption that the concentration of alcohol in a person's breath is proportional to the concentration of alcohol in their blood. This assumption is based on Henry's law, which states that the concentration of a gas in a liquid is directly proportional to the partial pressure of that gas above the liquid.*

The conversion factor used to convert BrAC to blood alcohol concentration (BAC) is generally considered to be reliable and consistent across individuals, assuming that the device used for testing is properly calibrated and maintained. The conversion factor is based on a number of factors, including the person's body weight, gender, and the time elapsed since their last drink. These factors have been studied extensively, and the conversion factor is widely accepted as a reliable estimate of a person's BAC.

However, it's worth noting that the conversion factor may not be exact for every individual. Factors such as lung capacity, lung function, and even the shape of a person's airway can affect the accuracy of BrAC testing. Additionally, certain medical conditions, such as gastroesophageal reflux disease (GERD), can cause false positives on a breathalyzer test.

Overall, while the conversion factor used to convert BrAC to BAC is generally considered to be reliable, it's important to remember that breathalyzer results should be considered

in conjunction with other evidence in DUI cases, and that individual variations can affect the accuracy of these tests.

At this point it becomes evident that this chatbot is unaware of the relevant literature on BAC test methods and can provide only very general considerations found in survey papers rather than in the pertaining scientific literature. It is also apparent that it can provide significant answers only if the question is formulated by someone who already knows the answer, otherwise the answer remains at a very superficial level.

It is worth noting that the chatbot used in these tests – ChatGPT – is a general purpose chatbot and we cannot exclude that more sophisticated tools are under development for legal applications. However, this does not change the already mentioned problem about the validity and exhaustiveness of the data on which the AI bases its results and this appears to be the greatest limitation of AI, at least in legal applications, because it cannot be up-to-date with all new findings, that remain the prerogative of human intelligence.

IV. CONCLUSIONS

While AI applications have significantly developed in the recent years and have achieved impressive semantic capabilities, the results they can provide are still largely based on the availability of data on which they can be trained and they can search to formulate their response to the problem they have been presented.

Under this assumption, the paper has proposed to consider the involved ethical issues in the perspective of the ethics of evidence and has shown that a strict metrological approach to the validation of the considered data can provide a sound quantitative estimation of the reliability of the obtained results.

Reliability of the AI-provided results is critically important also when they are used to help solving legal cases. As an example of the results that can be obtained with the presently available AI tools, the scientific evidence available in a real DUI case was presented to ChatGPT and the obtained answers discussed and compared to the analysis provided by an expert witness, proving that this AI tool would have been of little help in this specific case.

It can be concluded that AI is a very promising tool also in the legal field, and its development should be definitely encouraged. On the other hand, at least in the present stage of development, it is strongly influenced by the availability and validity of the data considered to devise a solution to the submitted problem. Therefore, it becomes ethically important to give evidence that all relevant pieces of information have been considered, including those related to their uncertainty.

REFERENCES

- [1] M. Polit, T. Talbert, B. Lopez, and J. Melendez, *Artificial Intelligence Research and Development*, ser. Frontiers in Artificial Intelligence and Applications 146. IOS Press, 2006.
- [2] B. C. Smith, *The Promise Of Artificial Intelligence: Reckoning And Judgment*. The MIT Press, 2019.
- [3] V. Scotti, "Artificial intelligence," *IEEE Instrumentation & Measurement Magazine*, vol. 23, no. 3, pp. 27–31, 2020.
- [4] —, "Can instruments replace humans also as liable entities? [legal metrology]," *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 5, pp. 11–12, 2015.

- [5] —, “Big data or big (privacy) problem?” *IEEE Instrumentation & Measurement Magazine*, vol. 20, no. 5, pp. 23–26, 2017.
- [6] T. Sourdin, *Judges, Technology and Artificial Intelligence*. Elgar Law, Technology and Society, 2021.
- [7] D. Petri, “Big data, dataism and measurement,” *IEEE Instrumentation & Measurement Magazine*, vol. 23, no. 3, pp. 32–34, 2020.
- [8] L. Mari and D. Petri, “Measurement, dataism and post-truth ideology: The good, the bad and the ugly,” *IEEE Instrumentation & Measurement Magazine*, vol. 25, no. 7, pp. 3–6, 2022.
- [9] D. G. Johnson and T. M. Powers, “Ethics and technology: A program for future research,” in *Encyclopedia of Science Technology and Ethics*, C. Mitcham, Ed. Farmington Hills, MI, USA: Thomson Gale, 2005, vol. 1, pp. xxvii – xxxv.
- [10] S. Blackburn, *Ethics: A very short Introduction*. New York, NY, USA: Oxford University Press, 2001.
- [11] L. Wittgenstein, “A lecture on ethics,” in *Moral Discourse and Practice: Some Philosophical Approaches*, S. Darwall, A. Gibbard, and P. Railton, Eds. New York, NY, USA: Oxford University Press, 1996, pp. 65–70.
- [12] V. Miké, “The ethics of evidence: A call for synthesis,” in *Encyclopedia of Science Technology and Ethics*, C. Mitcham, Ed. Farmington Hills, MI, USA: Thomson Gale, 2005, vol. 1, pp. lii – lx.
- [13] A. Ferrero, “The pillars of metrology,” *IEEE Instrumentation & Measurement Magazine*, vol. 18, no. 6, pp. 7–11, 2015.
- [14] JCGM 100:2008, *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement*, (GUM, originally published in 1995), Joint Committee for Guides in Metrology, 2008.
- [15] A. Ferrero and V. Scotti, *Uncertainty and Conscious Decisions*. Cham: Springer International Publishing, 2022, pp. 115–124. [Online]. Available: https://doi.org/10.1007/978-3-031-14619-0_8
- [16] A. Ferrero, D. Petri, P. Carbone, and M. Catelani, *Fundamentals of Hard and Soft Measurement*. Hoboken, NJ: Wiley-IEEE Press, 2015, pp. 203–262.
- [17] D. Petri, L. Mari, and P. Carbone, “A structured methodology for measurement development,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 9, pp. 2367–2379, 2015.
- [18] A. Ferrero and V. Scotti, “Forensic metrology: a new application field for measurement experts across techniques and ethics,” *Instrumentation Measurement Magazine, IEEE*, vol. 16, no. 1, pp. 14–17, 2013.
- [19] —, “Forensic metrology: when measurement science meets ethics,” in *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering*, 2014, pp. 1–6.
- [20] —, *Forensic Science: When Science Enters the Courtroom*. Cham: Springer International Publishing, 2022, pp. 27–40. [Online]. Available: https://doi.org/10.1007/978-3-031-14619-0_3
- [21] D. L. Schacter, *The seven sins of memory: How the mind forgets and remembers*. Houghton Mifflan, New York, NY, USA, 2021.
- [22] A. Ferrero and V. Scotti, *Breath and Blood Alcohol Concentration Measurement in DUI Cases*. Cham: Springer International Publishing, 2022, pp. 127–150. [Online]. Available: https://doi.org/10.1007/978-3-031-14619-0_9