# **Quarterly newsletter**



# **Ethics and Artificial Intelligence**



UNIVERSIDAD POLITÉCNICA DE MADRID



# Analysis of meta-trends

The iDanae Chair for Big Data and Analytics (where iDanae stands for intelligence, data, analysis and strategy in Spanish), created within the framework of a collaboration between Universidad Politécnica de Madrid (UPM) and Management Solutions, aims to promote the generation and dissemination of knowledge, the transfer of technology, and the furthering of R&D in the Analytics field.

One of the lines of work developed by the iDanae Chair is the analysis of meta-trends in the field of Analytics. A meta-trend

can be defined as a value-generating concept or area of interest within a particular field that will require investment and development from governments, companies and society in the near future<sup>1</sup>.

This report is focused on the role of ethics in the development of Artificial Intelligence models.

<sup>1</sup>For further information, please refer to iDanae, 2019.



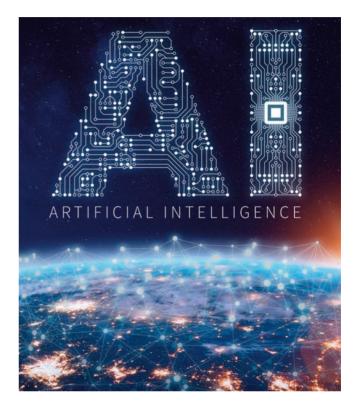
## Introduction

The rise of artificial intelligence (AI) is revolutionising all industry sectors, especially those that have digitally transformed their processes or have native digital processes. Systems that incorporate AI interact with people in decision-making (and even replace them on some occasions). It is therefore necessary that these decisions be made in accordance with the ethical standards established in society. For example, a program called COMPAS<sup>2</sup> has been in use since 1998 in the United States to predict the probability that a criminal will relapse in the future, using an algorithm that evaluates the person based on 137 parameters. This system, which is used as support, reviews the history of more than one million convicts, and has a 65% success rate<sup>3</sup>. These types of tools have been reviewed by different bodies. For example, the NCSC<sup>4</sup> has incorporated in its Guide to the Courts of Justice<sup>5</sup> a principle that deals with the use of these types of analysis tools. If one considers using software such as COMPAS to make decisions that will significantly affect people's lives, one must have an ethical framework to prevent bias.

The still open debate on how to incorporate ethical principles into the use of artificial intelligence in order to ensure consistency with human criteria is becoming increasingly important. This is due to both the greater use of AI and to the occurrence of significant errors of judgment resulting from the use of these decision-making systems. It therefore seems necessary to lay down an ethical hierarchy.

This debate is present in regulatory bodies, crystallising into normative proposals; in the business sphere, leading to organisational and governance changes; in the academic sphere, resulting in theoretical frameworks and academic studies; and, finally, in civil society as a whole through outreach activities, awareness-raising, etc.

The term Artificial Intelligence can be used with different meanings. This report focuses on weak AI, which refers to algorithms that perform specific tasks in a way that is far superior to the human ability, and that are especially effective in finding patterns and relationships between the data, but without the capability to include human criteria in their analysis if such criteria are not explicitly included<sup>6</sup>. It is in this case, where analytical capacity is separated from typically human criteria such as morality, wisdom or compassion, that ethical treatment becomes relevant in everything related to the data: from its collection to its use, which may include storage.



Other related concepts not addressed in this document are strong AI (referring to a machine that has an intelligence similar to that of humans<sup>7</sup>, including the ability to interpret feelings and emotions), and superior intelligence (known as "the singularity"<sup>8</sup>, an intelligence capable of learning by itself in a way that would surpass human intelligence).

Debates around this subject raise the question of whether it makes sense to speak of ethics in relation to artificial intelligence, under the hypothesis that an algorithm has no capacity to think<sup>9</sup>. So what exactly is the Ethics of Artificial Intelligence? Is the Ethics of Artificial Intelligence about what intelligent systems (algorithms, robots, etc.) should practice based on their values, or is it what human beings should apply

<sup>4</sup>National Center for State Courts (NCSC, 2011). <sup>5</sup>Ibídem.
<sup>6</sup>Coppin, 2004.

- <sup>7</sup>Searle, 1980.
- <sup>8</sup>Chalmers, 2010. <sup>9</sup>A different ques

<sup>9</sup>A different question, outside the scope of this article, is whether computers can think, whether they can have consciousness and, depending on the answer to these questions, whether they should have an ethic.

<sup>&</sup>lt;sup>2</sup>COMPAS, 2015. <sup>3</sup>Sam Corbett-Davies, 2016.



when developing and using intelligent systems? Who will be responsible for mistakes made by a machine, or for decisions taken by a human based on an intelligent system?

This report focuses on weak AI and on ethics with actual or potential legal consequences. In particular, the process of developing an AI project is analysed in order to understand the responsibilities throughout the project, considering the relationship that this responsibility may have with legal consequences. For example, when leveraging on an automatic system in a loan admission process, in the advice for a pension fund, or in the decision on a person's suitability for a job, the following question arises: who is responsible for a wrong decision? Different responsibilities could arise: (i) the person who follows the recommendation from the system; (ii) the person who decided to acquire and use the system; (iii) the person who sells the system; (iv) the person who programmed the algorithm; (v) the person who decided to use data that might be wrong; etc.

In some cases there may be a legal consequence, even if the developer has not thought of it, and it may have an uncalculated economic impact.

These systems are currently in use. The development of digitalization increases the number of actions and, therefore, of users who may be potentially affected by incorrect decisions. Therefore, the following question needs to be answered: who is responsible?

Answering this question requires an analysis of all development phases in an AI project as well as an understanding of the ethical aspects that may arise in each phase:

- 1. **Objective setting**. At this stage it is necessary to analyse whether the objectives are ethical (which is something that is common to all projects, not exclusive to the field of artificial intelligence).
- 2. **Data collection**. This requires analysing the data collection process, including data ownership and privacy aspects in addition to data collection itself and data use.
- 3. **Model construction**. In this phase, several aspects have to be analysed again::
  - a. Possible model bias.
  - b. Model error.
  - c. Model transparency and interpretability.
- 4. **Responsibility in the implementation and use of models**. Finally, the responsibility involved in the implementation and communication of models must be analysed with a view to their subsequent use.

With the aim of providing an overview of the context of Ethics in AI, this article reviews the four areas outlined above under the perspective of ethics with legal considerations (without addressing other possible aspects of ethics), and ends with a view of the initiatives and regulatory frameworks that attempt to address them.



## Ethics in the development of artificial intelligence projects

#### **Project objective**

Determining the objective of an artificial intelligence project is one of the points at which ethical dilemmas arise, especially in those areas where AI replaces human judgment. Concerns in this regard arise not only because of the potential errors these systems can make, which will be discussed later, but also because of a reduction in the human ability to make decisions in areas where there is no clear ethical way of solving problems. A typical case arises in the generation of artificial intelligence systems used in disease diagnosis or the analysis of drug interactions, where it must be decided whether the system should replace the doctor's diagnosis or prescription, or whether it should only be used as a support guide and the responsibility should remain with the physician. Another wellknown example is the potential prioritization of casualties by the operating system of a self-driving vehicle<sup>10</sup>.

#### Data collection

Data ethics refers to the branch of ethics that deals with assessing moral issues in connection with data, algorithms and This area of ethics includes the analysis of the moral dimensions of information<sup>12</sup>: privacy, anonymisation, transparency, truthfulness, and accountability in data-related processes (capture, transformation, analysis and use)<sup>13</sup>. It therefore focuses on the ethical issues that arise in the collection and analysis of large data sets and on issues ranging from the use of big data in biomedical research and social sciences<sup>14</sup>, to advertising<sup>15</sup>, among others.

Another issue of interest under analysis is whether a data ethic should be developed to determine if the current concepts of privacy<sup>16</sup>, fair and non-discriminatory treatment, and traceability (with regulations such as GDPR), as well as any other applicable rules, such as those relating to honour or the right to one's own image, are sufficient. For example, Strava made public some geolocation data from devices worn by soldiers during their training routines which revealed the position of military bases in Iraq and Afghanistan<sup>17</sup>. Article 22 of the GDPR states<sup>18</sup>: 'The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'. Therefore, the

current legislation not only includes elements relating merely to the collection and storage of data, but also alludes to their modelling and use. There are also data that, although private or intimate in nature, may be available for modelling purposes. Is it ethical to use such data to make decisions that could vary the treatment of an individual? For example, in line with the above, the Council of Europe issued a recommendation that includes a proposal to prohibit insurance companies from calculating the future risk posed by their customers based on genetic tests<sup>19</sup>.

Finally, there are three aspects of data that can be confused but have different nuances: ownership, collection, and use of data.

- ► Data ownership: regardless of where data is stored, data ownership rights may belong to different agents. The GDPR regulation positions itself in favour of the users rather than the companies that make use of the data.
- **Collection:** refers to the forms of data collection. regardless of who owns the data or whether users make their data public. A question that is often asked in this respect is the following: is it ethical for companies to collect data if the user does not want it, even if the data are public? Typical examples are browsing data or information shared on social networks.
- Data use: refers to any uses arising from the exploitation of data, including the potential value generated for businesses and governments. The questions that arise are similar to those in the previous point. For example, is it necessary to detail how a dataset will be used? An example is the experiment conducted by the Public Health Service of the United States in collaboration with Tuskegee University in Alabama on untreated syphilis, where African-American men were only told that they would receive free

<sup>15</sup>Hildebrandt, 2008.

<sup>&</sup>lt;sup>10</sup>Jean-François Bonnefon, 2016.

<sup>&</sup>lt;sup>11</sup>Floridi, L., y Taddeo, M. (2016).

<sup>&</sup>lt;sup>12</sup>Some authors speak of the 5 C's: consent, clarity, consistency, control (and transparency), and consequences (and harm) (DJ Patil, 2018). <sup>13</sup>Íbidem.

<sup>&</sup>lt;sup>14</sup>Mittelstadt BD, 2016.

<sup>&</sup>lt;sup>16</sup>Including the concept of measuring privacy on the web, understood as the observation of websites and services to detect, characterize and guantify behavior affecting privacy (Steven, E., y Narayanan, A. 2016). <sup>17</sup>DJ Patil, 2018.

<sup>&</sup>lt;sup>18</sup>GDPR, 2018.

<sup>&</sup>lt;sup>19</sup>European Council, 2016.



medical care<sup>20</sup>. To mitigate this type of situation, the European Commission<sup>21</sup> indicates that a research participant must be informed of the specific use to be made of his or her data. Another matter of interest is whether the value generated by the data should be passed on to the users or can be retained by the companies and agents that use them. For example, the governor of the state of California proposed, in his inaugural speech, a law under which companies whose business model is based on the collection and use of data would have to pay a dividend to the people whose data are used<sup>22</sup>.

#### **Model construction**

#### Potential bias

There is potential bias<sup>23</sup> in the construction and use of an artificial intelligence system. This bias may arise from multiple sources, including: (i) the existence of training data that are either incomplete (because there are no population attributes or characteristics relevant to the model) or unbalanced (because the attributes or predicted variable are not suitably represented), or (ii) potential bias that may result from specific techniques being used in the model training process. If no corrective treatment is applied, the model outcome may be biased, yielding false positives or false negatives that affect subsequent decision-making. This can occur in different types of algorithms, such as facial recognition models, algorithms for recruitment within the Human Resources function, or performance evaluation. One of the problems caused by bias is that its treatment may be incompatible with obtaining a model that optimizes predictive power for a given sample. This is why addressing model bias is fundamental to avoid loss of predictive power on the one hand, and discrimination that might be unnecessary or violate people's rights on the other.

An example that serves to illustrate aspect (i) above is the model developed by Amazon to hire new employees. The model was found to be biased, as it penalised women's candidacies. This happened because the model was trained with the curricula received by Amazon in the previous ten years, mostly submitted by men, which made the model interpret that being a woman was an unfavourable characteristic when looking for a technical profile<sup>24</sup>. This is an error arising from unbalanced training data, in this case because it contains a historical bias.

An example of aspect (ii) could arise following the ruling by the European Court of Justice<sup>25</sup> that differentiation of insurance premiums for men and women, solely on grounds of gender, is incompatible with the EU Charter of Fundamental Rights. However, since in the case of motor insurance women statistically have fewer accidents than men, a machine learning model could end up discriminating between men and women, even if the training data did not specify the customers' gender, through the use of gender-correlated variables, which would result in bias when processing the data.

#### Errors

Artificial intelligence models have an inherent error rate. When using these models in decisions of high importance, e.g. as support in judicial or health and safety decisions, and where it is not possible to correct the error a posteriori, is it ethical to use these models, even knowing that they make mistakes?

On the one hand, it is necessary to analyse whether these algorithms should be used, even when they improve the success rate compared to traditional systems, given that they can hide higher error rates for certain subpopulations, with the consequent unfair treatment for them. On the other hand, renouncing their use may generate an opportunity cost. Although there are formulas for reviewing decisions in both the traditional and the machine learning systems, some of the final decisions will necessarily be false positives or negatives in either case. A paradigmatic example of this dilemma can be found in the field of health: on the one hand, a false positive can cause unnecessary treatment, leading to a painful process for the patient and an expense for the system; on the other hand, a false negative can delay necessary treatment and reduce its success rate. Moreover, implementing these models, which require large amounts of data in order to learn and find patterns, is problematic in situations where the data are scarce, as is the case with rare diseases. However, the analysis of errors made by algorithms and the appropriateness of their implementation also needs to be made taking into consideration whether these errors are lesser than the ones made by humans when carrying out the same tasks. Moreover, it is important to question whether it is necessary for the model to have a 100% success rate or whether it is enough if it improves on the human success rate, allowing for the existence of errors.

<sup>&</sup>lt;sup>20</sup>Brandt, 1978

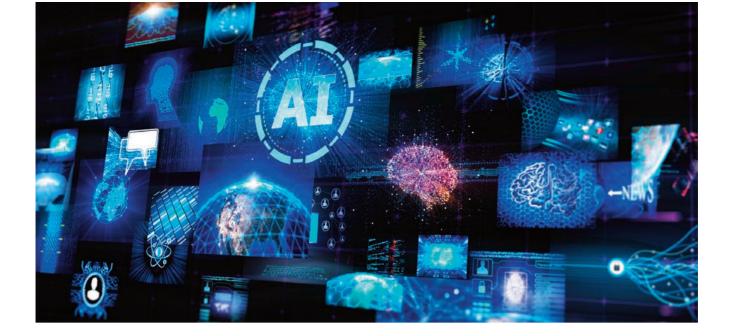
<sup>&</sup>lt;sup>21</sup>European Commision, 2018

<sup>&</sup>lt;sup>22</sup>Newsom, 2019.

<sup>&</sup>lt;sup>23</sup>Bias means a disproportionate weight in favour of or against one thing, person or group in comparison with another, generally in a manner that is considered unfair.

<sup>&</sup>lt;sup>24</sup>Dastin, 2018.

<sup>&</sup>lt;sup>25</sup>European Commission, 2012.



#### Transparency and traceability

The fact that many artificial intelligence models are highly complex and can be used either autonomously or as a support in decision-making, means it is important to understand how models work and what they should be used for. It is also essential to interpret and monitor how models works in order to determine whether they are ethical (particularly for nonreplicable models), in terms of not only their outcome, but also of how this outcome is arrived at<sup>26</sup>. On the other hand, transparency about how models work can determine possible limitations in model construction or implementation and, therefore, model use.

However, the right of access to information on models is subject to certain legal limits, which may be an impediment to achieving interpretability or transparency. There are different reasons why these limits exist: to protect professional secrecy, intellectual property or national security, to guarantee the confidentiality required in certain processes such as judicial or health related processes, or not to hinder the investigation of crimes. In this context, some initiatives have addressed the possibility of restricting access to information and have raised whether the information should be made the responsibility of specific regulatory agencies in cases where its relevance is justified. In this case, both the degree of understanding that can reasonably be expected from the agents involved and the instruments that a regulatory agency may use to carry out its supervisory responsibility are fundamental.

# Responsibility in the implementation and use of models

As the use of Al increases, there are more situations that require determining the responsibility for decisions taken. It is necessary to incorporate ethical values into technological developments in artificial intelligence, as well as to address how technology should respond to issues that have an ethical impact<sup>27</sup>.

This, in turn, is a driving force for this responsibility to be explicitly established in both legal frameworks and commercial agreements. Al can make mistakes that cause damage, or it can be used maliciously despite its proper functioning, therefore a system of clear responsibility is required in both cases. This need is especially relevant in situations where the use of Al has a great impact, such as when it is applied in the health, financial or judicial spheres.

An example that illustrates the above is the traffic accident that produced the first fatal victim involving an autonomous vehicle . A debate arose as to who should be responsible for the accident: whether the car manufacturer, the company that had developed and installed the autonomous driving system or the person who supervised the operation of the system and who should have braked.

As the above example shows, responsibility may arise in any of the earlier described AI model development phases, and in some cases it may even derive from decisions at several of these stages.

Although these types of ethical challenges are not new, and are therefore not a consequence of using algorithms in decisionmaking, there is a new aspect of analysis arising from the need in some cases for the programming to be explicit as to what decisions would be taken in extreme situations, as well as from the large-scale application of a single criterion as opposed to individualised decisions.

<sup>26</sup>For further information, please refer to iDanae, 2019.
 <sup>27</sup>Dignum, 2018.
 <sup>28</sup>Laris, 2018.

# **Regulatory frameworks**

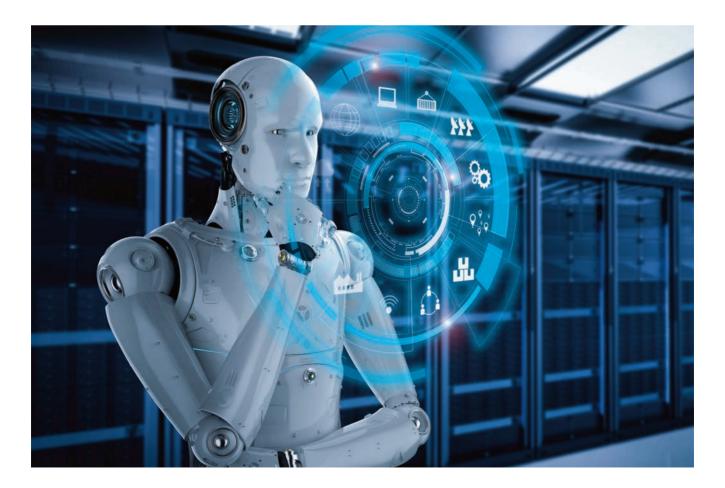
In view of the above, the different regulators, as well as civil society and other stakeholders involved are considering laying down ethical principles in their respective spheres of action to govern both the way in which these algorithms are developed and used.

At present, different frameworks and proposals coexist in this regard. From the most basic ones, where these principles are addressed as an extension of privacy or as a set of fair treatment criteria, to others based on principles that all these algorithms must comply with.

#### Public and private initiatives

Within the field of ethics in artificial intelligence, a number of initiatives from both business and government have been aimed at regulating some issues, or at least establishing

certain principles that will underpin regulations on the use of Al in decision making. From the private initiatives, companies are opting to reinforce the model validation, audit and risk control functions, placing special emphasis on interpretability and explainability throughout the entire model life cycle (development, implementation and use) and data life cycle (origin, storage and use), as well as user perception. The same approach is true for corporate governance, which has led to a reinforced control function through stronger request for information as well as greater responsibility on the part of senior management, and has included independent people for dealing with Al issues in committees. All this allows companies to be more aware of and responsible for both fully automated decisions and those that are aided by the use of machine learning algorithms.





As early as 1993, the Association for Computing Machinery<sup>29</sup> published a code of ethics and professional conduct aimed at anyone who uses computer technologies in a way that has an impact (professionals, instructors, students, etc.). This code of ethics incorporates general ethical principles (honesty, reliability, non-discrimination, privacy, etc.), as well as a set of principles and responsibilities of computer professionals.

Three of the most outstanding current private initiatives are (i) The Asilomar Principles<sup>30</sup>, developed in January 2017 by the Future of Life Institute together with the Asilomar conference<sup>31</sup>. These principles establish guidelines for the development of an ethical AI, tackling problems that are usually encountered in research (its purpose, who finances it, cooperation between research groups and avoiding competitiveness between teams), ethical issues (such as security, transparency, responsibility, privacy, human control, respect for society and traceability) and long-term issues (constraints, potential risks, capabilities, AI importance and limitations on self-improving AI); (ii) the IEEE<sup>32</sup> has developed an initiative on the ethics of autonomous and intelligent systems; or (iii) the Partnership on Al<sup>33</sup> aims at establishing good practices and educating the general public on Al.

For their part, governments and civil society are proposing various initiatives that are based on issues such as approval before implementation, auditing by external bodies or limiting the use of models in certain areas.

More specifically, government initiatives are being developed by i) the United States<sup>34</sup>, covering all types of impacts of AI on society: in innovation, in industry and for workers; ii) United Kingdom<sup>35</sup>, where a Data Ethics framework has been defined and good governance of technologies based on the use of

data is being developed; or iii) European Commission<sup>36</sup>, where principles have been defined for the processing of information in research proposals that include the use of personal data, as well as for approaching IA based on respect for human rights and the control of processes that yield AI systems' decisions. Other bodies, such as the World Economic Forum<sup>37</sup>, have also looked into other issues like the potential for mass unemployment, the biases mentioned above or potential cybersecurity problems. Of particular interest is the proposal for a future European Agency for Robotics and Artificial Intelligence<sup>38</sup>, which would be responsible for identifying areas of work where there are potential risks, such as those dealing with health-related or public interest issues.

#### <sup>29</sup>ACM, 2018.

<sup>30</sup>Asilomar, 2017.

<sup>37</sup>Bossman, 2016.

<sup>&</sup>lt;sup>31</sup>Asilomar has held several conferences on ethical standards of various scientific disciplines, such as on the design of genes and living organisms in 1975 or on the future risks of AI in 2009.

<sup>&</sup>lt;sup>32</sup>The IEEE is a professional technical organization dedicated to the advancement of technology for the benefit of humanity. Please read IEEE, 2019.

<sup>&</sup>lt;sup>33</sup>This consortium is made up of a multidisciplinary group of researchers and academics, and companies such as Apple, Amazon, Accenture, Baidu, Facebook, Google, IBM, Intel, McKinsey, Microsoft, Nvidia, PayPal, Salesforce, Sony, Samsung and Unicef, among others. See Partnership on AI, 2016. <sup>34</sup>USA Government, 2019.

<sup>&</sup>lt;sup>35</sup>UK Government, 2018, y UK Government, 2019.

<sup>&</sup>lt;sup>36</sup>European Comission, 2018, y European Comission, 2019.

<sup>&</sup>lt;sup>38</sup>European Parliament, 2016.

# **Conclusions**

In this report, ethics with actual or potential legal consequences has been analysed. The advent of AI calls for the need to raise and address ethical problems that arise from its development and use, in order to find possible solutions. This need is not covered at this moment in time, partly due to the difficulty this entails, and partly due to the pace of AI development. However, since AI systems are going to have an influence on people's lives and behaviour, decisions that are the output of AI algorithms need to be made in accordance with the ethical rules established by society. The digitalisation of all systems and the possibility of capturing data from all processes will exponentially increase the effect of decisions based on AI systems.

It is also important to emphasize that the interpretability of the models obtained with Al has an impact in the use cases.

Therefore, the interpretability of the models and the possible use cases are directly related.

Among the most important dilemmas at present are those related to the purpose of AI, the use of data, the emergence of bias and error, as well as the problems that arise when it comes to understanding the decision process and the responsibility for the resulting decisions. It is therefore necessary to establish ethical principles and laws that govern the way these algorithms are developed and used. Responsibility for this development must be shared by individuals and companies as well as by governments and regulators.



### References

**ACM. (2018).** Code of Ethics and Professional Conduct. Association for Computing Machinery Council. Retrieved from https://www.acm.org/code-of-ethics

**Asilomar. (2017).** Future of Life. Retrieved from Future of Life: https://futureoflife.org/ai-principles/

**Bossman, J. (2016).** World Economic Forum. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2016/10/top-10-ethicalissues-in-artificial-intelligence/

Brandt, A. M. (1978). Racism and Research: The Case of the Tuskegee Syphilis Study. The Hastings Center Report, 21-29.

**Chalmers, D. (2010).** The singularity: a philosophical analysis. Journal of Consciousness Studies .

**COMPAS. (2015).** Practitioner's Guide to COMPAS Core. Retrieved from Practitioner's Guide to COMPAS Core: http://www.northpointeinc.com/downloads/compas/Practition ers-Guide-COMPAS-Core-\_031915.pdf

Coppin, B. (2004). Artificial Intelligence Illuminated .

**Dastin, J. (2018).** Reuters. Retrieved from Reuters: https://www.reuters.com/article/us-amazon-com-jobsautomation-insight/amazon-scraps-secret-ai-recruiting-toolthat-showed-bias-against-women-idUSKCN1MK08G

**Dignum, V. (2018).** Ethics in artificial intelligence: introduction to the special issue. Ethics Inf Technol 20, 1–3. doi:10.1007/s10676-018-9450-z

DJ Patil, H. M. (2018). Ethics and Data Science. O'Reilly Media, Inc.

**European Comission. (2012).** EU rules on gender-neutral pricing in insurance industry enter into force. European Comission.

**European Comission. (2018).** Ethics and data protection. European Comission.

**European Comission. (2019).** Policy and investment recommedations for trustworthy Al. European Comission.

**European Council. (2016).** Council of Europe. Retrieved from Council of Europe:

https://search.coe.int/cm/Pages/result\_details.aspx?ObjectId=0 9000016806b2c5f

**European Parliament. (2016).** Directorate-General for Internal Policies. Retrieved from Directorate-General for Internal Policies: https://www.europarl.europa.eu/RegData/etudes/STUD/2016/5 71379/IPOL\_STU%282016%29571379\_EN.pdf

Floridi, L., y Taddeo, M. (2016). What is data ethics? Phil. Trans. R. Soc. A374: 20160360. Retrieved from http://dx.doi.org/10.1098/rsta.2016.0360

**GDPR.** (2018). GDPR. Retrieved from GDPR: http://www.privacy-regulation.eu/es/22.htm

Hildebrandt, M. (2008). Profiling and the identity of the European citizen. Springer.

iDanae. (2019). iDanae. Retrieved from iDanae: https://blogs.upm.es/catedra-idanae/newsletter-trimestral-3t19/

**IEEE. (2019).** IEEE. Retrieved from IEEE: https://ethicsinaction.ieee.org/

Jean-François Bonnefon, A. S. (2016). The Social Dilemma of Autonomous Vehicles. Science.

Laris, M. (2018). The Washington Post. Retrieved from The Washington Post:

https://www.washingtonpost.com/transportation/2018/12/20/ nine-months-after-deadly-crash-uber-is-testing-self-drivingcars-again-pittsburgh-starting-today/

**Mittelstadt BD, F. L. (2016).** The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Science and Engineering Ethics.

NCSC. (2011). NCSC. Retrieved from NCSC: https://www.ncsc.org/~/media/Microsites/Files/CSI/RNA%20Gu ide%20Final.ashx

**Newsom, G. (2019).** Office of Governor. Retrieved from Office of Governor: https://www.gov.ca.gov/2019/02/12/state-of-the-state-address/

**Partnership on Al. (2016).** Partnership on Al. Retrieved from Partnership on Al: https://www.partnershiponai.org/about/

Sam Corbett-Davies, E. P. (2016). The Washington Post. Retrieved from The Washington Post: https://www.washingtonpost.com/news/monkeycage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysisis-more-cautious-than-propublicas/

Searle, J. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences.

**Steven, E., y Narayanan, A. (2016).** Online Tracking: A 1-million-site Measurement and Analysis. SIGSAC.

**UK Government. (2018).** Data Ethics Framework. Department for Digital, Culture Media & Sports.

**UK Government. (2019).** Centre for data ethics and innovation. Retrieved from Centre for data ethics and innovation: https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation

**USA Government. (2019).** White House. Retrieved from White House: https://www.whitehouse.gov/ai/

#### **Authors**

Ernestina Menasalvas (UPM) Alejandro Rodríguez (UPM) Manuel Ángel Guzmán (Management Solutions) Daniel Ramos (Management Solutions) Carlos Alonso (Management Solutions) David Castaño (Management Solutions)



UNIVERSIDAD POLITÉCNICA DE MADRID

The Universidad Politécnica de Madrid is a public-law organization of a multisectoral and multidisciplinary nature that is engaged in teaching, research, as well as science and technology development activities.

www.upm.es



Management Solutions is an international consulting firm whose core mission is to deliver business, risk, financial, organizational and process-related advisory services, with operations in more than 40 countries and a multidisciplinary team of 2,500 professionals working for over 900 clients worldwide.

www.managementsolutions.com

### For more information, visit **blogs.upm.es/catedra-idanae/**