Regular Issue

Empirical Analysis of Ethical Principles Applied to Different AI Uses Cases

Alfonso José López Rivero[1], M. Encarnación Beato[1], César Muñoz Martínez[2], Pedro Gonzalo Cortiñas Vázquez[2] *

[1] Computer Science Faculty. Universidad Pontificia de Salamanca, Salamanca (Spain)
[2] Universidad Nacional de Educación a Distancia, Madrid (Spain)

Received 9 May 2022 | Accepted 7 July 2022 | Published 28 November 2022

Abstract

In this paper, we present an empirical study on the perception of the ethical challenges of artificial intelligence groups in the classification made by the European Union (EU). The study seeks to identify the ethical principles that cause the greatest concern among the population, analyzing these characteristics among different actors. The main study analyses the difference between Information and Communications Technology (ICT) professionals and the rest of the population. Along with this study, we conducted a gender study; in addition, we studied differences between university students, classified as future professionals who can work in Artificial Intelligence, and other university students. We believe that this work is a starting point for an informed debate in the scientific community and industry on the ethical implications of artificial intelligence based on the classification of ethical principles made by the EU, which can be extrapolated to any analysis carried out on the use of data to apply them in algorithms based on Artificial Intelligence.

Keywords

Artificial Intelligence, Ethic, Ethics Of AI, Digital Ethics, Trust, Digital Transformation.

DOI: 10.9781/ijimai.2022.11.006

I. Introduction

Artificial Intelligence (AI) is positioning itself as a strategic technology for the development of society in general and the economy in particular. Its use is increasingly widespread, covering all areas from healthcare, manufacturing, and security to agriculture and leisure. It is a life-changing technology with great prospects to benefit society as a whole; however, this growing use and potential for growth have also brought a recent concern to make AI systems trustworthy systems.

Work has been ongoing in recent years to make AI a trustworthy system: as early as 2018, the European Union (EU) published a coordinated plan on AI [1] in which one of the four basic pillars is the development of ethical and trustworthy AI. The achievement of trusted AI is closely linked to ethical AI that is based on fundamental values and rights such as human dignity and the protection of privacy. To this end, the EU's coordinated plan promoted the formation of an independent group of experts from different fields (academia, business, and society) to establish ethical guidelines for the development and use of AI systems. In 2019, this High-Level Expert Group on artificial intelligence (HLEG) published the document "Ethics guidelines for trustworthy AI" [2], with a vision centered on ethics as a fundamental pillar to guarantee trustworthy AI that benefits both human prosperity at the individual level and the common good of society.

* Corresponding author.
E-mail addresses: ajlopezri@upsa.es (A. J. López Rivero), ebeatogu@upsa.es (M. E. Beato), cmunoz@cee.uned.es (C. Muñoz Martínez), pcortinas@cee.uned.es (P. G. Cortiñas Vázquez).

This concern for the ethics of AI is not only present in Europe, in this same area there are other private initiatives such as the "Beneficial AI" conference organized by the Future of Life Institute in Asilomar (USA) in 2017, which was attended by the world's leading AI researchers and entrepreneurs and which resulted in the so-called "23 Asilomar principles" [3], or the work of the AI4People group [4] where a review of all the ethical principles proposed by different public and private initiatives is carried out. They compiled 47 principles and found that they can be grouped into the four basic principles used in bioethics, adding a fifth additional principle which would be explicability which includes responsibility. In another similar initiative by Jobin et al. [5], following the line of the previous work, a review of the main principles and guidelines for the ethics of AI was carried out, finding that there is currently a global convergence around five ethical principles (transparency, justice, and fairness, non-maleficence, responsibility, and privacy).

Another notable initiative is that of the United Nations Educational, Scientific and Cultural Organization (UNESCO), which, in 2020, set up an Ad Hoc Expert Group to draft recommendations on the ethics of artificial intelligence [6]. This document came out the same year intending to be a normative instrument based on international law that, with a primary focus on human dignity and human rights, could provide responsible guidance for AI technologies. Or the initiative similar to the one undertaken by the Organization for Economic Co-operation and Development (OECD) [7].

Other studies take a more philosophical view of AI, M. Graves [8] and J. F. Calderero [9] examine AI spiritually to suggest possible directions for ethics IA development. O. Krüger [10] goes further in his research by discussing the emergence of a superhuman computer intelligence for solving humanity's problems.

This is the context for this recently updated study, in which we conducted an empirical study with the aim of evaluating the public's perception of the ethical issues raised by the use of artificial intelligence, based on the ethical principles of reliable AI identified by the EU expert group. This work has chosen to work in a differentiated way with different population groups: firstly, with Information and Communications Technology (ICT) professionals, i.e. people who are directly involved in the conception, design, development, and implementation of AI-based systems, and therefore have a direct influence on the ethics of these systems. ICT professionals are a key group in the ethics of AI systems, so if these professionals have a lax perception of, or reflect little concern for, issues related to the ethics of AI, this could in principle negatively influence the ethics of AI systems they implement. A second, separate study group is university students in directly technology-related careers who will in the future be the designers and developers of AI systems. Finally, a third group is made up of people with no connection to the world of technology, users in general who use or will use products or services that make use of AI systems. The main objective is to assess the concern for AI ethics of these groups in a comparative way to see if there are significant differences between them that could influence the development of these systems in the future. In addition, differences by gender, differences between ICT and non-ICT university students, and age groups are analyzed. It is important to highlight the importance of the analysis by gender, as recent studies [11] show that only 22% of AI professionals are women, similar percentages are found among ICT students, so this gap will remain in the future; therefore, analyzing whether there are differences by gender is an issue that cannot be considered minor.

An analysis of all the recent bibliography highlights the study carried by T. Hagendorff and K. Meding [12], they analyze the 11,000 main papers from the main AI conferences worldwide, highlighting that the ethical implications are quite unknown, and although collaboration in publications between the academic and business world is increasing, it also highlights the difference with the articles presented by industry, where the ethical implications are analyzed less frequently than academic articles.

Another study [13] answered the question of how should ethics be implemented in autonomous systems and AI in our lives? It posits that the solution lies in philosophical conceptualization as a framework for forming a model for practical implementation of AI ethics. In the study, they conducted Systematic Mapping (SMS) on keywords used in AI and ethics to help identify, challenge, and compare the main concepts used in the current discourse of AI ethics. They analyzed over 1000 articles and discovered 37 keywords, considered the first step to guide and provide a direction for future research in the field of AI ethics.

Studies similar to these, in which a theoretical study on AI ethics is conducted, can be found in the literature, however, we have not found any empirical studies like the work presented in this article, which is why our study is novel and can be considered a starting point for an informed debate in the scientific community and industry. The ethical implications of artificial intelligence based on the classification of ethical principles made by the European Union can be extrapolated to any analysis carried out on the use of data to apply them in AI-based algorithms.

This article is organized as follows: Section II details the materials and methods used to carry out the study; section III contains the results obtained in the different studies; and finally, section IV sets out the conclusions obtained.

II. Methods

A. Materials

To carry out the study, we have designed a questionnaire of 15 questions, each of them specifically focused on one of the 5 ethical principles mentioned above. This questionnaire was validated by an interdisciplinary group of experts in the fields of ethics, technological ethics, and technology, from the Faculties of Philosophy and Theology of the Pontifical University of Salamanca. Concerning ethical considerations, the review was carried out by professors Dr. Marceliano Arranz Rodrigo and Dr. Gonzalo Tejerina Arias. To this main block of questions, another first block of questions was added to control the demographic characteristics of the subjects surveyed, including age category, level of studies, gender, and degree of knowledge of the technology on a scale of 1 to 10. Immediately after this, the block of questions on the main ethical challenges posed by artificial intelligence today was asked. These questions have been drawn up by experts in ethics in general, in technological ethics and technology, based on the four ethical principles identified by the HLEG group of experts [2], with the addition of a fifth, privacy, based on the grouping made by Jobin, Ienca, and Vayena [5]. These four principles are as follows: 1) Respect for human autonomy 2) Prevention of harm 3) Fairness 4) Explicability, 5) Privacy. Although all the questions are grouped here according to the ethical principle to which they belong, in the survey they were grouped to avoid reducing the bias caused by questions on the same issue in a row.

All the questions related to ethical aspects are written according to the same scheme: They describe a situation of use of artificial intelligence that poses an ethical dilemma and indicates their degree of acceptability on a scale from 1 (not at all acceptable) to 10 (completely acceptable), based on their own ethical decisions:

The situations related to each of these principles that make up the questionnaire are as follows:

1. *Respect for human autonomy*

   V1. The use of artificial intelligence to generate lifelike images and/or videos, and distribute them on social networks to create currents of opinion.

   V2. The use of artificial intelligence to serve as an electoral propaganda mechanism for parties on social networks.

   V3. The use of artificial intelligence that seeks to modify the consumption habits of the population.

   V4. The use of artificial intelligence in games that learns about the behaviour of players to increase the time spent playing the game.

   V5. The use of human-like robots to care for the elderly, capable of adapting to their needs, which could create affective dependence on the person being cared for.

2. *Prevention of harm*

   V6. Operate an autonomous driving vehicle that has not been sufficiently tested.

   V7. The use of artificial intelligence to integrate it into lethal autonomous weapons.

3. *Fairness*

   V8. The use of artificial intelligence where it is known that the data to be used for its learning is not of sufficient quality, with the risk that it learns badly.

   V9. The use of artificial intelligence for personnel selection, without human intervention and therefore objective, but the data could be biased in favour of men over women.

V10. The use of artificial intelligence to propose sentences in trials (in an objective way) in which the data from which the artificial intelligence learns could disadvantage certain races, social classes or groups.

4. *Explicability*

V11. The implementation of an artificial intelligence system in which control over the system does not depend on the human factor.

V12. The use of artificial intelligence to decide workers' salary supplements, knowing that it will not be possible to trace the reasons that lead the system to make such a decision.

5. *Privacy*

V13. The use of artificial intelligence through facial recognition to identify, record and learn about people's consumption habits, to stimulate the purchase of certain products.

V14. The use of artificial intelligence for video-surveillance of the public that, with the installation of cameras in the streets and facial recognition techniques, can obtain information on citizens by identifying their movements.

V15. The use of artificial intelligence to gather information on the tastes of the inhabitants of a house, using listening to virtual assistants and using it to make music recommendations.

B. Participants

The survey was applied to 457 people who fall into one of the following groups: ICT professionals, ICT students, non-ICT students, and general users. The sample of ICT and non-ICT students was obtained from university students in the academic years 2020/21 and 2021/22, while for ICT professionals the survey was distributed by the Professional Association of Computer Engineers of Castilla y León and AETICAL (Association of Technology Companies of Castilla y León). The users who completed the survey were selected by random sampling and the questionnaire was sent to them by email. Participation in the study was voluntary with informed consent; all data collected is considered confidential and only used for the study. The data have been deleted after the extraction of the aggregated data presented in this article.

For the surveys conducted, n=456, considering the worst-case scenario; p=q=0.5, where we have a non-finite population scenario where the total population is unknown and a confidence level of 95%, the sampling error is determined. The equation to determine the sample size can be found in many studies, in this research we have used the analysis of Louangrath, P.I., [14].

$$n = \left(\frac{Z*\sigma}{E}\right)^2 \tag{1}$$

Where Z, is the critical value for the confidence interval, in the study we have considered 95%, σ is the standard deviation, obtained by:

$$\sigma = \sqrt{p*q*n} \tag{2}$$

For the survey sample size of 456, the total error in the study is E=4.6%.

The socio-demographic characteristics of the sample are shown in Table I. The sample sizes are considered sufficiently large to conclude with a high degree of reliability.

C. Tasks and Methods

1. Data Preprocessing

Before statistical analysis, data pre-processing needs to be performed to transform the raw data obtained into a "clean" and ordered dataset that allows for good statistical analysis.

TABLE I. Demographic Characteristics of the Sample

		n	%
Sexo	Male	288	63%
	Female	169	37%
Tipo	ICT professionals	94	20%
	ICT students	233	51%
	Non-ICT students	67	15%
	Others	63	14%
Edad	Under 30s	294	64%
	30-60 years old	102	22%
	Over 60 years old	61	14%

Data pre-processing is collected from different sources [15], [16], a procedure for data pre-processing is as follows:

- Data cleaning: This stage deals with missing data, noise, outliers, and duplicate or incorrect records, and minimizes the introduction of bias into the database.
- Data integration: The extracted raw data may come from heterogeneous sources or be in separate datasets. This step reorganizes the different raw data sets into a single dataset containing all the information needed for the desired statistical analyses.
- Data transformation: this step translates and/or scales the initial variables into formats more useful for the statistical methods the researcher wishes to use.
- Data reduction: Once the dataset has been integrated and transformed, this step removes redundant records and variables and reorganizes the data in a more efficient and orderly manner for analysis.

During pre-processing, care must be taken not to accidentally introduce bias by modifying the dataset in a way that affects the outcome of statistical analyses, and to avoid arriving at statistically significant results by "trial and error" analysis on different versions of a pre-processed dataset.

In the study carried out, the following considerations have been taken into account concerning the points described above:

- Data cleaning: In the study that has been carried out, the ID has been checked to analyze duplicate data and, as for possible outliers of the variables analyzed, the first step was to check if there was a difference between the mean and median to detect possible values, in all cases the values were close; in addition, it was verified, by performing a scatter plot, if there were data out of range, in no case were outliers obtained.
- Data integration: In the analysis that has been carried out we have homogeneous sources of data as we have stratified data for the different groups that are analyzed: gender, workers/users, and ICT/non-ICT specialists.
- Data transformation: From the initial results obtained none of the variables followed a normal distribution, so the variables had to be transformed into normal distributions. This transformation is detailed in the following section.
- Data reduction: In the study presented, being a previously validated analysis, there are no redundant variables, and the data are ordered for the analysis.

2. Data Transformation. Normalisation of Variables

The quantitative data used for this study were obtained from a survey, collecting 15 quantitative variables that we wanted to measure concerning the perspective on ethics in artificial intelligence held by different actors. In the first analysis of these variables, it was verified that they did not follow a normal distribution, as can be seen in Fig. 1 and Fig. 2, for the variable V1.

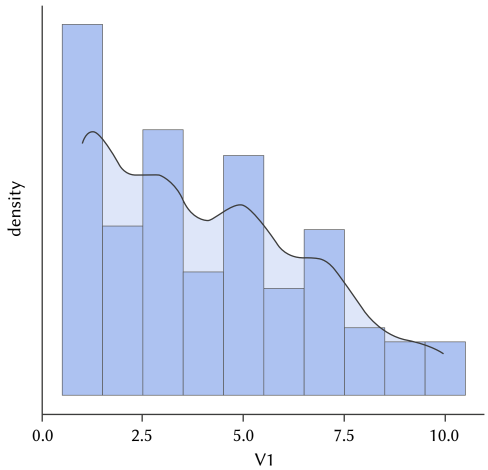

Fig. 1. Histogram for untransformed variable.

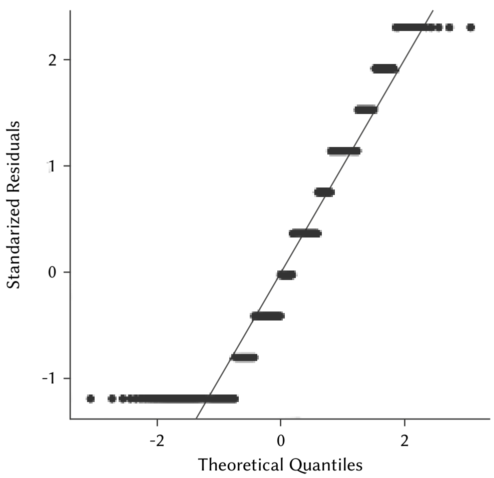

Fig. 2. Q-Q graphic diagram for untransformed variable.

In the histogram, it can be seen that the variable does not follow a normal distribution and if the Q-Q, Quantile-Quantile graph is performed, it is verified that the data do not follow a normal distribution function.

A similar process can be found in "Evaluating the normality of various quantitative data transformation procedures" by Dago et all [17], for which the following transformations are used: Z-score, Logarithmic, Adaptive Gamma Distribution, and Box-Cox.

3. Z-score Standardisation

In Z-transform data normalization, the x_i values of a variable V_i are normalized based on the mean and standard deviation of V [18]. Z-scores, or standard scores, indicate how many standard deviations and observations are above or below the mean. These scores are a useful way of putting data from different sources on the same scale. A value x_i of V is normalized to x_i' by calculating:

$$x_i' = \frac{x_i - \bar{x}}{\sigma_i} \tag{3}$$

Where $\bar{x}$ y σ_i are the mean and standard deviation of the variable V_i.

4. Logarithmic Transformation

The logarithmic transformation is very useful when results are influenced by independent factors, and in particular for transforming distributions with positive skewness, [19], the logarithmic transformation formula is given by:

$$x_i'' = \log x_i' \tag{4}$$

This logarithmic transformation is only valid for numbers greater than zero, to avoid having numbers less than zero when normalizing the variable, a constant is added to each value, in the study carried out the average value of the data has been added.

$$x_i'' = \log(x_i' + \bar{x}) \tag{5}$$

5. Adaptive Gamma Distribution

When the response variable is assumed to follow a parametric probability distribution, gamma in this study, the parameters of this distribution can be modeled, each independently, following linear, non-linear, non-parametric functions. This versatility makes the gamma function a suitable tool for modeling variables that follow a whole range of distributions: non-normal, asymmetric, or with non-constant variance [20].

6. Box-Cox Transformation

The Box-Cox transformation is a family of potential transformations of non-normal dependent variables, which allows transforming the variable to normal density functions, as this feature is a statistical assumption that allows using many more techniques for the analysis of the variables [21]. They take the idea of having a range of power transformations instead of the classical square root, logarithm, and inverse, available to improve the efficiency of normalization and variance equalization for variables with positive and negative skewness [22], [23]. The variable transformation has the form, where x are the original values we have in the sample, alpha is a transformation constant to avoid negative values when calculating the logarithm and lambda is the transformation parameter:

$$x_t^{\lambda} = \begin{pmatrix} \frac{(x_t)^{\lambda}-1}{\lambda} & if\ \lambda \neq 0 \\ \log(x_t + \alpha) & if\ \lambda = 0 \end{pmatrix} \tag{6}$$

In the study presented, the statistical analysis has been carried out with the R-based software package, JAMOVI [24], for the transformation of the variables the following equation has been followed, where K is a constant, the value 5 was taken as the initial scale is from 0 to 10, V_i is each of the 17 variables that have been analyzed, $\bar{V}$, σ_{v_i} are, respectively, the mean and the standard deviation of each variable.

$$V_i' = BOXCOX(GAMMA(LN\left(\frac{V_i - \bar{V}}{\sigma_{v_i}} + K\right)) \tag{7}$$

Fig. 3 and Fig. 4 show the result of the transformation performed for the first variable, V1, whose distribution function follows a normal density function. From the Q-Q plot it is found that the data fit the function.

If we consider the value of the Shapiro-Wilk statistic, we have that the transformed variable has a p-value of 0.762, which implies that the values fit a normal distribution.

III. Results

Firstly, we carried out a descriptive analysis of each of the variables related to ethical perception, grouped by ethical principle. This analysis is shown in Fig. 5, where each of the variables studied, grouped by ethical principle, is represented by a box plot.

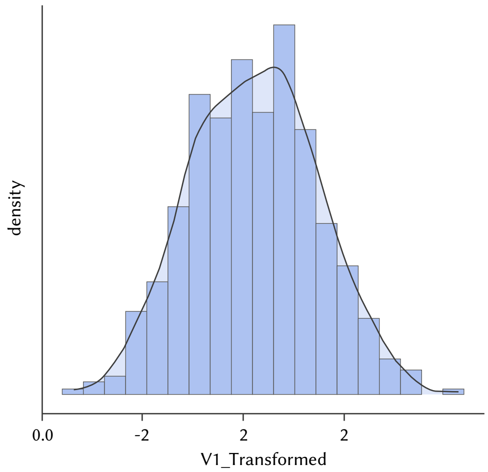

Fig. 3. Histogram for the transformed variable.

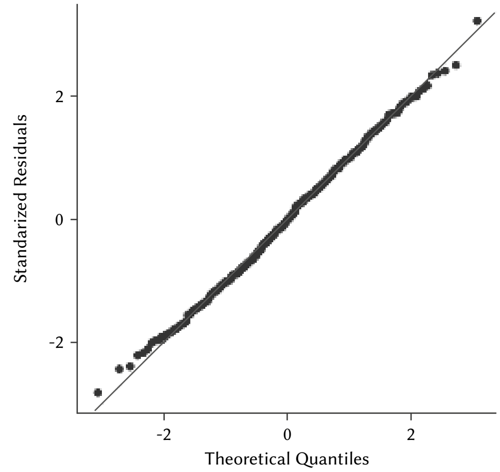

Fig. 4. Q-Q plot for the transformed variable.

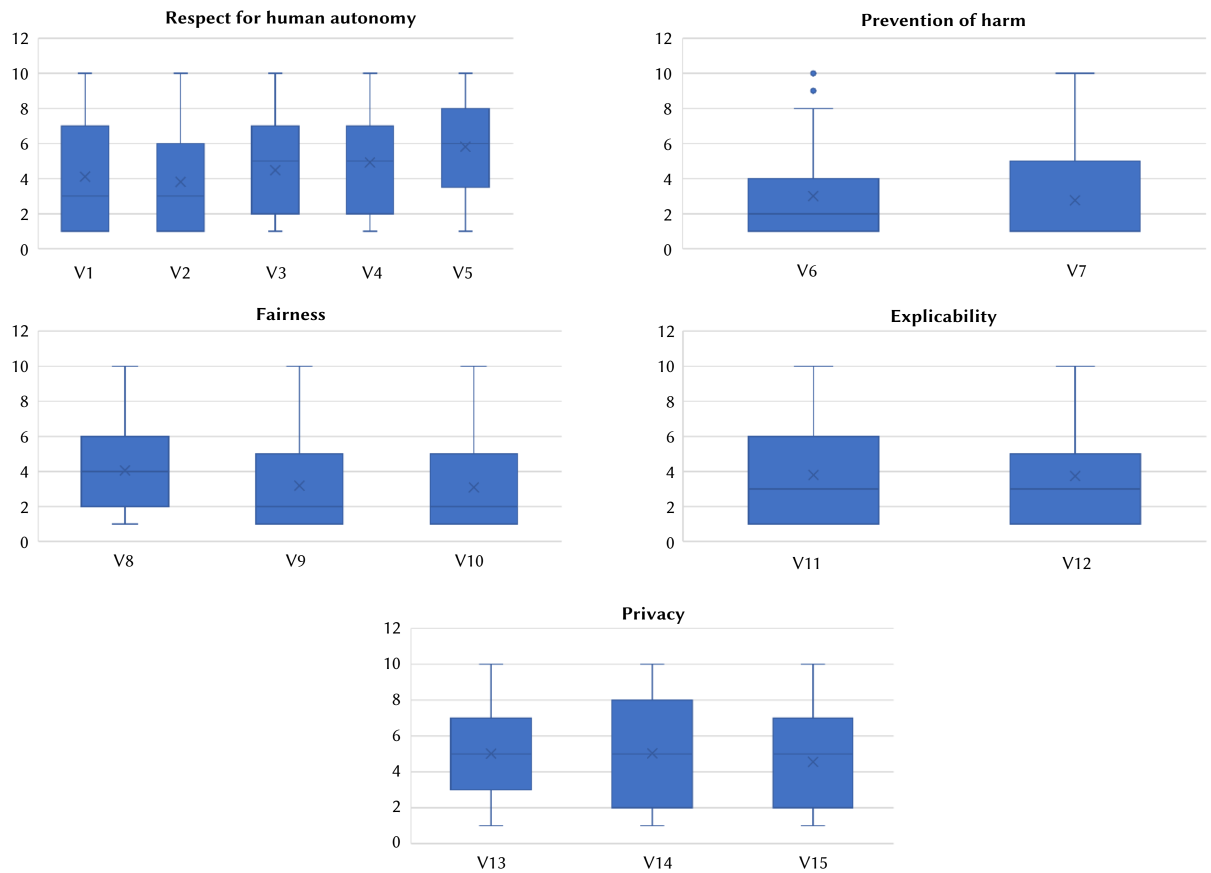

Fig. 5. Box plots showing the distribution of values for each variable and ethical principle.

Subsequently, an inferential analysis based on the p-value has been carried out. Before showing the results of these analyses, we will explain how to interpret the p-value and the effect size used.

The p-value measures the compatibility of a sample with a hypothesis, not that the hypothesis is true [25], as the p-value is a probability statement about the observed sample in the context of a hypothesis, not about the hypotheses being tested. This is why the American Statistical Association (ASA) [26], published an article in 2016 highlighting the misuse of p-values in statistical inference. Professor Cobb highlighted why we use p=0.05 as the cut-off value, citing numerous cases in which it is not used correctly. These errors in considering the interpretation of the p-value alone led Professor Nuzzo to publish an article on statistical errors in the scientific method [27].

Therefore, to test the significance of the hypotheses, in addition to the p-value, other parameters must be considered, and in this analysis, we have considered the effect size, a concept introduced by Cohen [28].

Effect sizes, Cohen's d, are the most important result of empirical studies [29], used to describe the standardized mean difference of an effect. This value can be used to compare effects across studies, even when the dependent variables are measured in different ways. A commonly used interpretation is to refer to effect sizes as small (d = 0.2), medium (d = 0.5) and large (d = 0.8) according to the benchmarks suggested by Cohen [28].

TABLE II. Averages. Users (1) Vs. ICT Professionals (2)

		Mean		Median		Standard deviation	
		1	2	1	2	1	2
Respect for human autonomy	V1	4.58	2.29	5	1	2.87	2.19
	V2	4.17	2.39	4	1	2.76	2.26
	V3	4.84	3.09	5	2	2.69	2.50
	V4	5.14	4.02	5	3.5	2.95	2.80
	V5	5.74	6.06	6	7	2.77	2.73
Prevention of harm	V6	3.09	2.69	2	2	2.59	2.41
	V7	2.95	2.04	1	1	2.69	2.05
Fairness	V8	4.25	3.28	4	3	2.57	2.50
	V9	3.29	2.82	2	1	2.68	2.69
	V10	3.21	2.57	2	1	2.70	2.49
Explicability	V11	3.86	3.54	3	3	2.70	2.60
	V12	3.83	3.38	3	3	2.49	2.66
Privacy	V13	5.35	3.71	5	3	2.64	2.71
	V14	5.30	4.00	5	3	2.87	2.84
	V15	4.86	3.40	5	2.5	2.89	2.65

TABLE III. Users in General Vs. ICT Professionals

		p	Effect Size	95% Confidence Interval	
				lower	upper
Respect for human autonomy	V1	**0.025**	0.2603	0.03250	0.4877
	V2	**0.023**	0.2631	0.03532	0.4906
	V3	**0.001**	0.3827	0.15392	0.6110
	V4	0.561	0.0673	-0.15962	0.2941
	V5	**0.037**	-0.2423	-0.46960	-0.0146
Prevention of harm	V6	0.645	-0.0534	-0.28022	0.1735
	V7	0.502	0.0778	-0.14919	0.3046
Fairness	V8	0.769	0.0340	-0.19287	0.2608
	V9	0.445	0.0884	-0.13858	0.3153
	V10	0.384	0.1009	-0.12614	0.3277
Explicability	V11	0.740	0.0385	-0.18841	0.2653
	V12	0.412	0.0951	-0.13188	0.3220
Privacy	V13	**0.042**	0.2360	0.00838	0.4633
	V14	0.372	0.1035	-0.12352	0.3304
	V15	**0.016**	0.2802	0.05231	0.5078

TABLE IV. Average. Men (1) Vs Women (2)

		Mean		Median		Standard deviation	
		1	2	1	2	1	2
Respect for human autonomy	V1	4.09	4.14	3.00	4	2.98	2.76
	V2	3.84	3.75	3.00	3	2.77	2.76
	V3	4.42	4.59	5.00	5	2.76	2.72
	V4	5.52	3.87	6.00	4	2.93	2.69
	V5	6.29	4.99	7.00	5	2.58	2.87
Prevention of harm	V6	3.24	2.62	2.00	1	2.65	2.35
	V7	3.07	2.25	1.00	1	2.77	2.17
Fairness	V8	4.23	3.75	4.00	3	2.51	2.68
	V9	3.58	2.52	2.50	1	2.80	2.34
	V10	3.39	2.56	2.00	1	2.81	2.33
Explicability	V11	4.21	3.08	4.00	2	2.68	2.54
	V12	4.04	3.22	4.00	2	2.55	2.42
Privacy	V13	5.28	4.57	5.00	4	2.80	2.58
	V14	5.10	4.91	5.00	5	3.00	2.75
	V15	4.67	4.37	5.00	4	2.97	2.76

TABLE V. Difference by Gender Males Vs. Females

		p	Effect Size	95% Confidence Interval	
				lower	upper
Respect for human autonomy	V1	0.429	-0.0767	-0.267	0.1134
	V2	0.857	-0.0175	-0.207	0.1724
	V3	**0.036**	-0.20401	-0.3945	-0.0132
	V4	0.559	0.05667	-0.1334	0.2466
	V5	**<.001**	0.36639	0.1738	0.5583
Prevention of harm	V6	0.771	0.02820	-0.1618	0.2181
	V7	**0.014**	0.24028	0.0491	0.4310
Fairness	V8	0.132	0.14613	-0.0443	0.3363
	V9	0.085	0.16746	-0.0231	0.3577
	V10	0.654	0.04342	-0.1466	0.2333
Explicability	V11	0.267	0.10769	-0.0825	0.2977
	V12	0.371	0.08679	-0.1033	0.2768
Privacy	V13	0.178	0.13063	-0.0597	0.3207
	V14	0.983	-0.00209	-0.1920	0.1878
	V15	0.690	0.03872	-0.1513	0.2286

Concerning the analyses carried out, Table II shows the average scores for each question for ICT professionals, people who are working in this sector, compared to the rest of the respondents (General users: 1, professionals: 2).

With these data, we have analyzed the differences in ethical perceptions in each of the groups. The result of these differences for each of the ethical principles is shown in Table III. In this table, and all the equivalent tables in this section, we have marked in black the questions where we found significant differences between the groups.

In the second analysis, we have carried out a gender study to see if there are differences between men and women concerning the perception of ethical principles. The results of the averages obtained and the statistical analysis of differences are shown in Table IV and Table V, respectively.

Next, we have carried out an analysis to determine whether differences are found between students in careers directly related to technology versus non-technology students.

Bearing in mind that technology students will be future ICT professionals, the aim is to analyze whether their perception of the ethics of AI is different from that of other university students in non-technology-related subjects. The results of the analysis are shown in Table VI and Table VII.

Finally, we have carried out an analysis by age, and we have considered whether there are differences between young people

TABLE VI. Average NON-ICT University Students (1) Vs. ICT Students (2)

		Mean		Median		Standard deviation	
		1	2	1	2	1	2
Respect for human autonomy	V1	5.03	4.85	5	5	2.65	2.95
	V2	4.42	4.54	4	5	2.55	2.80
	V3	5.30	4.94	5	5	2.55	2.71
	V4	4.90	5.88	5	6	2.74	2.82
	V5	4.52	6.37	4	7	2.73	2.61
Prevention of harm	V6	2.73	3.45	1	2	2.42	2.74
	V7	2.15	3.54	1	2	2.07	2.91
Fairness	V8	4.48	4.51	5	4	2.59	2.52
	V9	2.55	3.87	2	3	2.03	2.88
	V10	3.22	3.57	2	2	2.71	2.81
Explicability	V11	3.52	4.37	3	4	2.39	2.75
	V12	3.46	4.33	3	4	2.34	2.51
Privacy	V13	5.54	5.80	5	6	2.40	2.59
	V14	5.87	5.43	6	6	2.91	2.93
	V15	5.12	5.06	5	5	2.83	2.96

TABLE VII. NON-ICT Vs. ICT University Students

		p	Effect Size	95% Confidence Interval	
				lower	upper
Respect for human autonomy	V1	0.609	-0.07102	-0.3427	0.20122
	V2	0.300	-0.14399	-0.4163	0.12935
	V3	0.839	0.02817	-0.2437	0.29981
	V4	**0.048**	-0.27480	-0.5495	0.00191
	V5	**0.003**	-0.41330	-0.6925	-0.13117
Prevention of harm	V6	0.944	-0.00975	-0.2814	0.26199
	V7	0.889	0.01933	-0.2525	0.29097
Fairness	V8	0.551	-0.08271	-0.3545	0.18967
	V9	0.072	-0.25059	-0.5247	0.02537
	V10	0.455	-0.10377	-0.3757	0.16890
Explicability	V11	0.716	0.05041	-0.2216	0.32206
	V12	**0.020**	0.32495	0.0465	0.60105
Privacy	V13	0.555	-0.08186	-0.3536	0.19051
	V14	0.725	0.04886	-0.2232	0.32051
	V15	0.645	0.06389	-0.2083	0.33557

TABLE VIII. Averages Over 60 Years Old (1) Vs. Under 30 (2)

		Mean		Median		Standard deviation	
		1	2	1	2	1	2
Respect for human autonomy	V1	2.33	4.36	1	4.00	2.34	2.93
	V2	2.43	3.97	1	3.00	2.47	2.75
	V3	2.97	4.70	2	5.00	2.52	2.66
	V4	3.98	4.93	3	5.00	2.95	2.92
	V5	5.97	5.60	7	6.00	2.74	2.77
Prevention of harm	V6	2.62	2.93	1	2.00	2.36	2.49
	V7	2.08	2.63	1	1.00	2.25	2.48
Fairness	V8	3.20	4.09	3	4.00	2.55	2.61
	V9	2.95	2.95	1	2.00	2.85	2.48
	V10	2.59	3.00	1	2.00	2.64	2.61
Explicability	V11	3.26	3.73	3	3.00	2.48	2.66
	V12	3.28	3.68	2	3.00	2.65	2.51
Privacy	V13	3.69	5.11	3	5.00	2.79	2.66
	V14	3.77	5.17	3	5.00	2.72	2.84
	V15	3.13	4.79	2	5.00	2.72	2.85

TABLE IX. Older Vs. Younger

		p	Effect Size	95% Confidence Interval	
				lower	upper
Respect for human autonomy	V1	0.340	-0.1344	-0.4106	0.1430
	V2	0.090	-0.2390	-0.5170	0.0410
	V3	**<.001**	-0.5285	-0.8179	-0.2350
	V4	0.882	0.0209	-0.2549	0.2966
	V5	0.055	0.2707	-0.0103	0.5496
Prevention of harm	V6	0.183	0.1877	-0.0909	0.4647
	V7	0.604	0.0730	-0.2034	0.3488
Fairness	V8	0.327	0.1380	-0.1394	0.4143
	V9	0.858	0.0252	-0.2507	0.3009
	V10	0.251	-0.1616	-0.4382	0.1163
Explicability	V11	0.787	-0.0381	-0.3137	0.2379
	V12	0.919	-0.0143	-0.2900	0.2615
Privacy	V13	0.112	-0.2241	-0.5018	0.0555
	V14	0.168	-0.1944	-0.4715	0.0843
	V15	**0.013**	-0.3520	-0.6335	-0.0678

(under 30 years old) and older people (over 60 years old). We have carried out the analysis with this casuistry to eliminate the bias that could be introduced by ICT professionals, all of whom are middle-aged so that professionals are excluded from this analysis by age. The results of this study are shown in Table VIII and Table IX.

IV. Discussion

We will now analyze the results obtained. First, we will focus on the descriptive analysis of the different situations on technological ethics. As indicated, the situations presented all pose some ethical challenge related to AI and the respondent is asked to rate them from 1 (not at all acceptable) to 10 (totally acceptable) using their own ethical decisions. The numerical scale of 1 to 10 has been chosen because of its similarity to the ratings, so we can understand that values lower than 5 indicate non-acceptance of the situation presented and higher than 5 indicate acceptance, although the numerical value will give us an idea of the magnitude of acceptance.

If we look at Fig. 1 and analyze the data represented there descriptively, we can see that in the analysis of averages there are only three of the 15 situations presented that obtain an average score higher than 5, these being variables V5, V13, and V14, with V5 standing out in particular with an average of 5.81. Variable V5, which falls within the ethical principle of respect for human autonomy, is related to the use of robots with a human appearance for the care of the elderly. The use of these robots for the care of the elderly means that it is valued positively and very highly with respect to the rest, despite the fact that, as reflected in the situation presented, it may create affective dependence in a particularly sensitive group such as the elderly. On the other hand, variables V13 and V14, related to the ethical principle of privacy, reflect situations in which facial recognition techniques are used, on the one hand, to find out consumer habits and stimulate purchases (V13) and, on the other, to identify citizens' movements for street surveillance (V14). In principle, this reflects the fact that citizens are not as concerned about their privacy as they are about the data protection and privacy regulations that exist in all countries.

On the other side of the spectrum, the variables V7 (2.77), V6 (3.01), V10 (3.08), and V9 (3.19), all with values close to or below 3, stand out for their low average scores. The scores show concern for the safety/reliability of these systems when they may affect human life, as they are situations of AI use in autonomous weapons and self-driving vehicles, although it should be noted that there may have been a bias due to the rejection of anything related to weapons. On the other hand, situations V10 and V9 are found within the principle of fairness and are closely linked to the rights to non-discrimination, solidarity, and justice, the second principle that causes the most misgivings. It reflects situations in which AI is used for decision-making in the knowledge that bias in the data may disadvantage women against men or certain races (if used to determine convictions in trials). However, when the situation raised about fairness is more general, i.e., in V5, where it is only mentioned that the data used is not of sufficient quality without indicating what problems this may entail, the average score of the respondents is significantly higher than 4.05, perhaps because many of the respondents do not manage to assess the problems that the lack of quality may bring.

In the following, we will analyze the results obtained from the comparative analyses. The main idea behind this study is to assess whether there are differences between the ethical perception of people who are working in technology (professionals) and the rest. We believe that the ethical perception of professionals can influence the ethics of the AI products they develop and, therefore, it is important that their concern for ethical issues is high. If we analyze the results obtained in Table III, we observe that all the differences found between the two groups are within the ethical principle of respect for autonomy, strongly associated with the right to human dignity and liberty, specifically in variables V1, V2, V3, and V5. In the first three, situations are presented that make use of AI for different types of manipulation. Thus, in V1, AI is used to create real-looking images or videos that can create currents of opinion, in V2, AI is used by political parties for their propaganda, and in V3, AI is used to modify the population's consumption habits. The effect sizes (Cohen's d) in them (0.26, 0.26, and 0.38) indicate a small but significant effect size. Furthermore, if we analyze the sign and mean scores of these two groups on each variable (Table II), we can observe that the mean scores on these variables are much lower in the case of professionals. In other words, professionals are less permissive on issues related to human autonomy. Except in variable V5 which has to do with the use of IA for the care of the elderly, where the score is reversed, with professionals being more permissive. In this variable, the effect size is -0.24.

If the descriptive data collected in Table II are analyzed in general, it can be seen that for all the variables analyzed except for the care of the elderly, the professionals are on average less permissive than the rest, which would indicate that their concern for the ethics of AI is greater.

We have also carried out a comparative analysis by gender, men, and women, in order to analyze whether there are differences between them in terms of concern for the ethics of AI. The results of the analysis are reflected in Table V. If we look at the table, we find differences in variables V3 and V5 related to human autonomy and variable V7 related to prevention of harm. However, although the order of magnitude of the effect size is similar, the sign is not the same. Thus, in V3 where AI is used to modify the consumption habits of the population, women are more permissive than men (Cohen's d -0.20), while in the other two variables where there are differences, V5 (use for elderly care) and V7 (to integrate it into autonomous weapons), women are less permissive than men, reflecting in both scores lower than those of men. Thus, the average score (Table IV) for women in the question on the use of AI in weapons (V7) is 2.25 compared to 3.07 for men, being in both cases far from approval, which would be 5. And in the question on the use of AI for the care of the elderly that can cause emotional dependence (V5), a lower score is also obtained in the case of women, the average for them is 4.99 compared to 6.29 for men. In any case, this is the question that practically passes in both cases.

Once the gender analysis and the analysis of professionals versus the rest were done, an analysis was carried out to look for differences between university students in ICT-related degrees and other university students in other fields. The idea of this analysis was to see whether ICT students were, already in their university years, more aware of the ethical issues involved in the use of AI. The results of this analysis are reflected in TABLE VI and TABLE VII. If we look at TABLE VI, we see that there are significant differences in only three variables (V4, V5, and V15), with the sign the negative effect in all of them, which indicates that ICT students value these situations with higher scores, i.e., they are more permissive or have fewer ethical problems with these situations. The results for these three variables do not ratify the initial hypothesis that ICT students were more aware of issues related to the ethics of AI. Specifically, the situations raised refer to the use of AI in games to increase playing time (V4 human autonomy), that of caring for the elderly, and finally V15 in which the use of AI by virtual assistants to make recommendations about music (privacy) is raised. Analyzing these results and looking at the percentage of men and women in both categories, we find that in the case of ICT students the percentage of women is 25% while among non-ICT students it is 63%. To see if this fact could have an influence, we carried out a gender analysis only with the students, only finding differences in favor of women (less permissive) in question V5, the one related to caring for the elderly. Therefore, we could indicate that only variables V4 and V15 have a different behavior between ICT and non-ICT students, with NON-ICT students being less permissive in the use of AI in the aforementioned situations.

Finally, we have carried out an analysis by age, in this case, we have carried out an analysis of young people (under 30) versus those over 60, in order to see if age is a factor that can influence the perception of ethics. Middle-aged subjects have been excluded from this analysis in order to eliminate the effect, already studied at the beginning of this section, of ICT professionals, as they all belong to this intermediate age group. The results of this analysis are reflected in TABLE VIII and TABLE IX, in which differences are only found in variables V3 (seeks to modify consumption habits) and V15 (use of AI by virtual assistants for music recommendations), with the assessment of the over-60s being more restrictive in both variables, with an average of 2.97 compared to 4.70 for young people for variable V3 and 3.13 compared to 4.79 for variable V15. However, we did not find any difference in the rest of the variables studied, in which age has no influence, so it is not considered a determining factor.

V. Conclusion

Following the field study in which we evaluated the population's perception of the ethics of AI systems, the first conclusion we draw from this study is that, in general, the population does not consider the use of AI to be acceptable in situations in which an ethical problem arises and, therefore, we can affirm that there is a high level of concern among the population about the ethics of these systems. This affirmation is corroborated by the average obtained in the responses, taking into account all the variables, which is 3.79, well below the pass mark. It should be remembered that in the situations presented, an ethical problem related to one of the ethical principles identified by the EU's high-level expert group on artificial intelligence [2] is presented and the degree of acceptability is indicated, on a scale from 1 Not at all acceptable to 10 Completely acceptable.

On the other hand, if we go into the concern of each of the five ethical principles, we find that the ones that raise the most concern are those of prevention of harm and fairness, the first of which is strongly linked to the protection of physical integrity, so the result is to be expected. The second, fairness, is related to ensuring that individuals and groups are free from unfair bias, discrimination, and stigmatization. The study also highlights the good reception of AI when it is used for the care of the elderly, even though it may create emotional dependence, with an average score of 5.81, as well as the average scores above 5 in the situations raised concerning video surveillance, in which people's privacy may be violated.

As far as ICT professionals are concerned, their concern about the ethics of AI is found to be higher than the rest of the population, indicating that they are even less permissive than the general population. This is a promising scenario as they are the first guarantors of the ethics of AI systems and project good prospects for the future in this area.

As for ICT university students, we found hardly any differences concerning other university students, which indicates that, although professionals in these fields are more aware of the ethics of AI, this is not the case when they are in their university years, where we found no differences.

Finally, we would like to point out that this is a field study carried out in Spain, so future lines of work will focus on extending it to other countries, as well as carrying out longitudinal work that will allow us to monitor and analyze the evolution of concern for the ethical issues of AI as the technology evolves and becomes more present in society.

Acknowledgment

We would like to thank the experts in ethics at the Pontifical University of Salamanca, Dr. Marceliano Arranz Rodrigo, retired professor of the Faculty of Philosophy and Dr. Gonzalo Tejerina Arias, professor of the Faculty of Theology, for their contributions and comments on the design of the survey for data collection.

References

[1] European Commission. Artificial Intelligence for Europe, 2018. Accessed: Jun. 24, 2022. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN.

[2] European Commission, Directorate-General for Communications Networks, Content and Technology. Ethics guidelines for trustworthy AI, 2019, Accessed: Jun. 24, 2022. [Online]. Available: https://data.europa.eu/doi/10.2759/346720.

[3] Future of Life Institute. ASILOMAR principles. Principles developed in conjunction with the 2017 Asilomar conference, 2017. Accessed: Jun. 24, 2022. [Online]. Available: https://futureoflife.org/2017/08/11/ai-principles/

[4] L. Floridi, J. Cowls, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U, Pagallo, F. Rossi, B, Schafer, P. Valcke, and E. Vayena, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and machines*, vol. 28 no. 4, pp. 689-707, 2018.

[5] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, vol. 1 no. 9, pp. 389-399, 2019.

[6] UNESCO Ad Hoc Expert Group (AHEG). First Draft of The Recommendation on the Ethics of Artificial Intelligence, 2020. Accessed: Jun. 24, 2022. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000373434

[7] OCDE. Recommendation of the Council on Artificial Intelligence. 2019. Accessed: Jun. 24, 2022. [Online] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449?_ga=2.248246209.790951165.1621927053-1872220198.1621927053

[8] M. Graves, "Emergent Models for Moral AI Spirituality," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7 no. 1, pp. 7-15, 2021.

[9] J. F. Calderero, "Artificial Intelligence and Spirituality," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7 no. 1, pp. 34-43, 2021.

[10] O. Krüeger, "The Singularity is near!. Visions of Artificial Intelligence in Posthumanism and Transhumanism," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7 no. 1, pp. 16-23, 2021.

[11] S. Duke, "Will AI make the gender gap in the workplace harder to close?" World Economic Forum. 2018. Accessed: Jun. 24, 2022. [Online] https://www.weforum.org/agenda/2018/12/artificial-intelligence-ai-gender-gap-workplace/

[12] T. Hagendorff and K. Meding, "Ethical considerations and statistical analysis of industry involvement in machine learning research," *AI & SOCIETY*, pp. 1-11, 2021.

[13] V. Vakkuri and P. Abrahamsson, "The Key Concepts of Ethics of Artificial Intelligence," in *IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, 2018, pp. 1–6.

[14] P. Louangrath, "Sample Size Calculation for Continuous and Discrete Data," *International Journal of Research & Methodology in Social Science*, vol. 5 no. 4, pp. 44-56, 2019.

[15] B. Malley, D. Ramazzotti and J. Wu, "Data Pre-processing,". *Secondary Analysis of Electronic Health Records*, pp. 115-141, 2016.

[16] X. Chu, I. F. Ilyas, S. Krishann and J. Wang, "Data Cleaning: Overview and Emerging Challenges," in *Proceedings of the 2016 International Conference on Management of Data, (SIGMOD)*, 2016, pp. 2201–2206.

[17] D. N. Dago, G. A. J. Kablan, K. A. Alui, H. D. Lallié, D. Dagnogo, N. Diarrassouba and M. Giovanni, "Normality Assessment of Several Quantitative. Data Transformation Procedures," *Biostatistics and Biometrics*, vol. 10 no. 3, pp. 51-65, 2021.

[18] D L. A. Shalabi, Z. Shaaban and B. Kasasbeh, "Data Mining: A Preprocessing Engine," *Journal of Computer Science*, vol. 2 no. 9, pp. 735-736, 2006.

[19] R. M. West, "Best practice in statistics: The use of log transformation," *Annals of Clinical Biochemistry*, vol. 59 no. 3, pp. 162-165, 2022.

[20] A. A. Al-Babtain, I. Elbatal, C. Chesneau and F. Jamal, "Box-Cox Gamma-G Family of Distributions: Theory and Applications," *Mathematics,* vol. 8 no. 10, pp. 1801, 2020.

[21] G. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological),* vol. 26 no. 2, pp. 211-243, 1964.

[22] J. Osborne, "Improving your data transformations: Applying the Box-Cox transformation," *Practical Assessment, Research, and Evaluation,* vol. 15 no. 1, pp. 12, 2010.

[23] N. Ishak and S. Ahmad, "Estimating Optimal Parameter of Box-Cox Transformation in Multiple Regression with Non-normal Data," in *Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, Singapore, 2018, pp. 1039–1046.

[24] The jamovi project (2021). *jamovi* (Version 2.2) [Computer Software]. Retrieved from https://www.jamovi.org

[25] N. Altman and M. Krzywinski, "Points of significance: interpreting P values," *Nature Methods,* vol. 14 no. 3, pp. 213-215, 2017.

[26] R. L. Wasserstein and N. A. Lazar, "The ASA statement on p-values: context, process, and purpose," *The American Statistician,* vol. 70 no. 2, pp. 129-133, 2016.

[27] R. Nuzzo, "Statistical errors," *Nature,* vol. 506 no. 7487, pp. 150-152, 2014.

[28] J. Cohen, *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum Associates, Publishers, 1988.

[29] D. Lakens, "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs," *Frontiers in psychology,* vol. 4, pp. 863, 2013.

Alfonso José López Rivero

Alfonso José López Rivero. PhD in Computer Science. He is a professor, since 1996, and member of the research group GESTICON (Ethical and Technological Management of Knowledge) at the Pontifical University of Salamanca, UPSA, (Spain). He is a member of the organizing and scientific committee of several international symposia and co-author of articles published in several recognized journals, workshops and symposia. He has been Dean of the Faculty of Computer Science and Director of the Office for the Transfer of Research Results at UPSA.

M. Encarnación Beato

Mª Encarnación Beato (PhD.). Received a PhD. in Computer Science from the University of Valladolid in 2004. She is professor at the Universidad Pontificia de Salamanca (Spain) since 1997. At present she is a member of the MARATON (Mobile Applications, inteRnet of things, dAta processing, semanTic technologies, OpeN data) research group at the Universidad Pontificia de Salamanca. She has been a member of the organizing and scientific committee of several international symposiums and has co-authored papers published in a number of recognized journals, workshops and symposiums.

César Muñoz

César Muñoz. PhD in European Union. He is an associate professor at the National Distance Education University (Spain). His research area focuses on transport and tourism economics. He's especially interested in exploring the application of ICT in these sectors. He has published several articles in scientific journals, including, International Journal of Tourism Research or Tourism Geographies.

Pedro Gonzalo Cortiñas Vázquez

Pedro Cortiñas Vázquez holds a PhD in Economics and Business Administration and is Professor of Applied Economics and Statistics at the Universidad Nacional de Educación a Distancia. Master in Commercial Management and Marketing by ESIC, University Expert in Multivariate Analysis and in Market Research and Opinion Polls, Specialist in Macroeconomics by The MIT Sloan School of Management.