Investigate Ethical Challenges and Considerations in the Collection, Analysis, and Use of Data for IT Analytics, Addressing Issues Related to Privacy, Bias, and Responsible AI

Ranadeep Reddy Palle

Independent Researcher
Software Engineer

Abstract: *Ethical issues have turned to be a major shortcoming, especially when dealing with digital data and evidence. In terms of data collection and analysis, the usage of AI has played a vital role in mitigating the issues related to the presence of unethical practices. The study has provided an examination of the various ethical issues which can occur during the instances of data collection and its application in IT analytics. On the other hand, with the collection of data from 75 participants, SPSS was performed for understanding the core topic of assessment from the general public. Through the usage of the random sampling method, the responses were obtained from the participants, and tests such as multiple regression, Correlation test, demographic examination and descriptive analysis were performed.*

Keywords: IT Analytics, Digital Data, Privacy, Ethical Issues, Cyber Threat, Data Leakage, Data Collection, Data Analysis, Artificial Intelligence, AI

1. Introduction

Background of the study

Data accuracy and the maintenance of the integrity of digital evidence is significant for reaching the correct conclusions to the phenomenon being analysed. With the expansive usage of digital data after the digital disruption in the corporate grounds, an elaboration of data related issues such as data leakage and cyber threat is seen to occur. According to the comments by Nassar & Kamal (2021), increase of biased information causes the reaching of discriminatory conclusions, which hampers the overall transparency and visibility of the information. On the other hand, as depicted by Cheng, Varshney & Liu (2021), responsible usage of the data is necessary for meeting the rights of the individuals, and determining the fact that their rights are not unethically challenged.
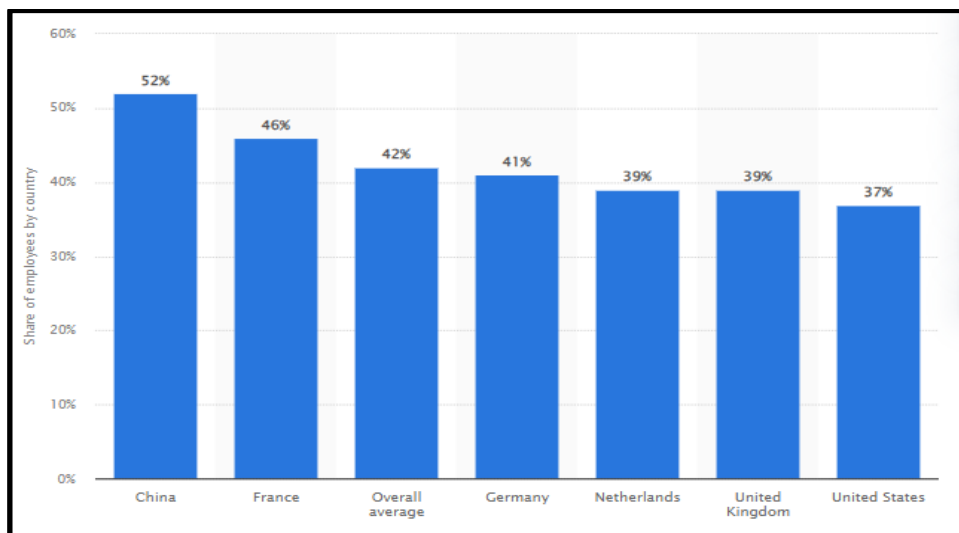

Figure 1: Ethical issues faced by employees of varying countries due to the usage of AI in their organisations (Source: Statista, 2022)

Because of such instances and shortcomings, the usage of responsible AI is important for increasing the privacy of the data and focusing the fact that informed consent is achieved.

2. Problem Statement

One of the main issues which are noted by the data analyst during the application of IT analytics on the digital data is the misuse of data, based on which the findings of the variables are based on. On the other hand, according to the

views of Richardson, Petter & Carter (2021), negative consequences can be obtained once there is the corruption of the personal and professional digital data present on the digital platform.

Figure 2: Ethical shortcomings related with digital data
(Source: Balasubramaniam et al. 2020)

With the inauguration of cyber threats on the data, there can also be negative implications on the evidence, which drastically declines the usage of the information, leading to corruption. In addition to such, as per the notions by Balasubramaniam et al. (2020), there can also be the generation of conflict of interest because of the reduced credibility of the information, and hampers the overall objectivity of the evidence. Hence, such potential notions are to be taken care of by the IT data analytics during their collection and utilisation of the digital information.

3. Research Aim and Objectives

Research aim
The study aims to draw an extensive assessment of the different ethical challenges and the considerations to be made during data collection and analysis, alongside the usage of data for IT analytics through AI application.

Research objectives
The objectives developed for the study are as follows:

RO1: To inspect the varying factors impacting the ethical notions of data collection and analysis for IT analytics with AI
RO2: To examine the significance of increasing ethical considerations of data protection in IT analytics with AI usage
RO3: To analyse the issues related to privacy and bias of data in case of IT analytics through the application of AI

RO4: To assess the tactics important for heightening ethical considerations and data protection in terms of IT analytics with AI

Research questions
The questions to be answered with the completion of the study are as follows:
RQ1: What are the different factors impacting the ethical notions of data collection and analysis for IT analytics with AI?
RQ2: In which way is the betterment of ethical considerations of data protection in IT analytics significant, especially with the usage of AI?
RQ3: What are the different shortcomings related to privacy and bias of data in case of IT analytics with AI?
RQ4: What are the necessary strategies significant for improving ethical considerations and data protection in terms of IT analytics with AI?

Significance of the study
With the performance of the study, the different ethical issues which occur during the collection of the digital data, and their subsequent interpretation with the application of IT analytics will be known. Alongside such, the usage of responsible AI for increasing privacy and reduction of bias, will also be known through the performance of the study.

4. Literature Review

Critical examination of the factors affecting the ethical aspects of digital data collection
According to the workings by Mikalef et al. (2022), the collection of data in an ethical manner from the users, by giving them a strict information of where the data would be used, and how they would be protected, ought to be taken into consideration. On the other hand, as per the views of Anagnostou et al. (2022), the accuracy of data inspection with the aid of responsible AI for amalgamating data ethics in its examination also acts as a factor of data governance and ethics. Based on the notions of digital governance, data ethics reflects on the moral valuations related with the digital data present. Hence, such aspects are to be taken into consideration during the collection and interpretation of the information.

Theoretical assessment of Theory of Data Ethics
The Theory of Data Ethics deals with the aspects of data governance where the digital ethics and digital regulations amalgamate with each other. On the other hand, as per the thoughts by Tarafdar et al. (2020), the determination of the degree to which the evidence needs to be transparent and the control processes ought to be visible to the users falls under the digital regulations. Hence, with the adherence with digital governance, the protection of ethics by the IT analysts in terms of digital data collection and inspection can be known
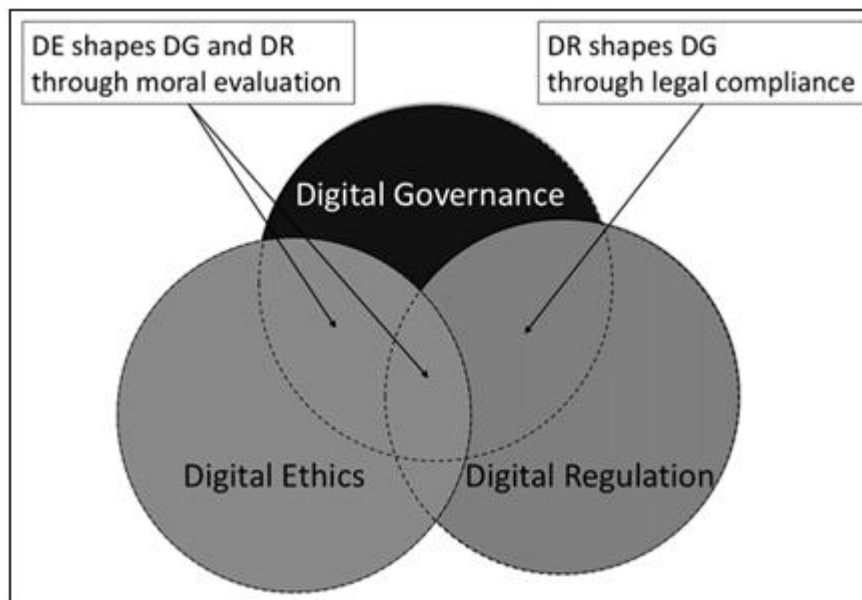
Figure 4: Elements of the Theory of Data Ethics
(Source: Tarafdar et al. 2020)

As noted in the above figure, the legal and the moral compliances adhered with the digital governance can be integrated into the known digital processes.

Gap in literature
One of the gaps which have been observed in the past set of literature is the lack of exact methods which can be deployed by the organisations for the mitigation of the ethical issues arising due to the usage of digital information. As per the views of Felländer et al. (2022), responsible AI has been able to increase the reliance of the digital evidence, and enable the AI to decrease their self-harm on the digital information. However, certain issues with empathy and the transparency of the digital data with the application of responsible AI have yet been observed. Therefore, the study has provided an elaboration of the perseverance of responsible AI and the positive notions of software usage for increasing the credibility and the security of the information from cyber threats and issues.

5. Methodology

Research methodology consists of the different stages and the steps important for analysing the data being gathered, from its varying sources (Pandey & Pandey, 2021). For the completion of the study, the collection of the primary quantitative had been achieved from 75 participants. The respondents were chosen with the help of random sampling, and the positivism research philosophy was integrated for analysing the factual data obtained. Two different research approaches, namely the inductive and the deductive research approaches can be applied in a study, based on the needs and demands of the research. In this case, there was the integration of the deductive research approach, where the hypotheses developed were analysed. The similarities and the patterns identified within the variables is enabled with the integration of the deductive research approach. On the other hand, as per the point of view of Mohajan (2018), the examination of the primary data helps in providing numerical results with the aid of statistical analysis. In such

a manner, the expansion of statistical understanding of the relationship between the elements of a study can be enabled. For the study, the collected data were passed through the primary quantitative software of SPSS or "Statistical Package for the Social Sciences". Tests such as demographic analysis, descriptive analysis, Pearson's correlation test and multiple regression were conducted to draw the results for the developed research questions.

Hypotheses development
The hypotheses developed for the study are as follows:
- *Alternative hypothesis (H1):* There is a strong relationship between biased data collection and ethical and data collection considerations
- *Alternative hypothesis (H2):* There is a heavy linkage between lack of privacy and ethical and data collection considerations
- *Alternative hypothesis (H3):* There is a sturdy association between accountability of responsible AI and ethical and data collection considerations
- *Alternative hypothesis (H4):* There is a strong bond between lack of transparency and ethical and data collection considerations

6. Findings

Demographic analysis

Age of the respondents

Table 1: Age of the respondents
What is your age?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	30	39.5	40.0	40.0
	2	21	27.6	28.0	68.0
	3	15	19.7	20.0	88.0
	4	9	11.8	12.0	100.0
	Total	75	100.0	100.0	

From the above table, it can be noted that the majority of the participants who opted for the survey belonged to the age range of 18-25 years, as 39.5% was obtained for such a category.
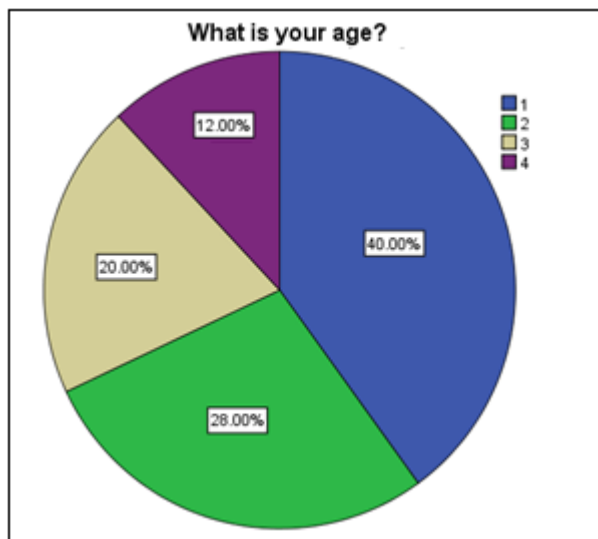

Figure 5: Age of the respondents

From figure 5, the blue segment, indicating option 1 of the survey, where the participants were between the ages of 18 to 25 were present, were seen to respond to the survey.

Gender of the respondents

Table 2: Gender of the respondents
What is your gender?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	37	49.3	49.3	49.3
	2	25	33.3	33.3	82.7
	3	13	17.3	17.3	100.0
	Total	75	100.0	100.0	

As seen from table 2, male participants were seen to be in a larger number, as compared to the females, as 49.3% of the total responses were seen to be provided for the first option. 17.73% of the total participants were seen to opt for the third option, where they did not prefer to reveal their genders for the survey.
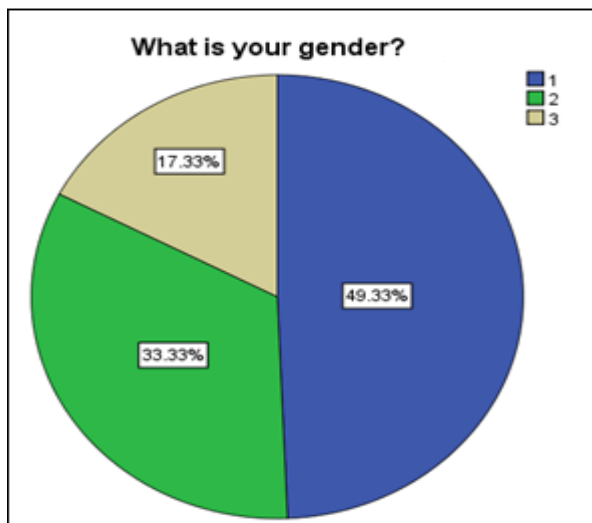

Figure 6: Gender of the respondents

As noted in figure 6, the first option covered the majority of the pie chart, indicating that the largest section was the blue coloured area, with 49.3%. Hence, it can be determined that the responses for the survey were based on the experience and notions of males.

Years of IT experience of the respondents

Table 3: Years of IT experience of the respondents
What is your years of IT experience

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	25	33.3	33.3	33.3
	2	14	18.7	18.7	52.0
	3	16	21.3	21.3	73.3
	4	20	26.7	26.7	100.0
	Total	75	100.0	100.0	

Based on table 3, the largest number of years, in terms of experience from the participants was less than 1. This is due to the fact that 33.3% of the total respondents chose the first option which was for less than a year. Such an aspect correlates with the age group of the respondents, where the majority was seen to be between 18 to 25 years.

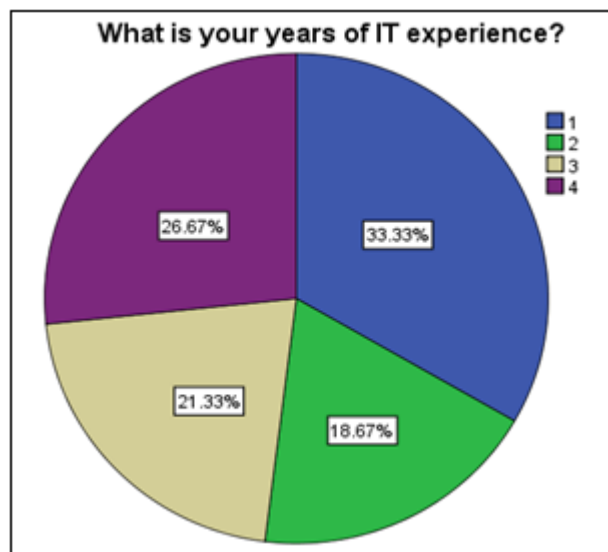**Figure 7:** Years of IT experience of the respondents

As seen in figure 7, the section covered in blue is seen to occupy the maximum, indicating that the first option was chosen by the greatest number of participants. Hence, it can be deduced that the years of experience for the participants who had worked in the IT industry was less than 1 year.

Variable related analysis
Descriptive analysis

Table 4: Descriptive statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
DV	75	1	5	2.79	1.719	.128	.277	-1.758	.548
IV1	75	3.00	9.00	6.3867	1.91645	-.443	.277	-.843	.548
IV2	75	2.00	9.00	5.4000	2.42732	-.047	.277	-1.212	.548
IV3	75	1	5	3.01	1.428	-.024	.277	-1.411	.548
IV4	75	2.00	10.00	6.0267	1.87424	-.229	.277	.317	.548
Valid N (listwise)	75								

The above table 4 shows the different numerical frequencies of the varying elements of the study. As per the opinions of Pallant (2020), the inspection and the overview of the different numerical functions such as the mean, median, mode and others, can be enabled with the help of the test of descriptive statistics.

Validity test

Table 5: Validity test

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.423
Bartlett's Test of Sphericity	Approx. Chi-Square	96.161
	df	10
	Sig.	.000

As per the validity test, the value of df or degree of freedom is seen to be 10. From such a valuation, it can be noted that there can be 10 other independent instances which can be used for the calculation of the statistics.

Pearson's correlation test

Table 6: Pearsons's correlation test

Correlations

		DV	IV1	IV2	IV3	IV4
DV	Pearson Correlation	1	-.381[**]	.306[**]	.070	.505[**]
	Sig. (2-tailed)		.001	.008	.548	.000
	N	75	75	75	75	75
IV1	Pearson Correlation	-.381[**]	1	.146	.062	.217
	Sig. (2-tailed)	.001		.210	.596	.061
	N	75	75	75	75	75
IV2	Pearson Correlation	.306[**]	.146	1	.574[**]	.108
	Sig. (2-tailed)	.008	.210		.000	.358
	N	75	75	75	75	75
IV3	Pearson Correlation	.070	.062	.574[**]	1	.277[*]
	Sig. (2-tailed)	.548	.596	.000		.016
	N	75	75	75	75	75
IV4	Pearson Correlation	.505[**]	.217	.108	.277[*]	1
	Sig. (2-tailed)	.000	.061	.358	.016	
	N	75	75	75	75	75

According to the above table, there is a negative valuation of 0.381 of the elements of biased data collection and ethical challenges and data collection considerations. This is due to the fact that the expansion of biassed data collected decreases the ethical considerations made in terms of data collection and its interpretation.

Multiple regression

Table 7: Model summary

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.653[a]	.427	.394	1.338	.427	13.035	4	70	.000	1.381

Table 8: ANOVA

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	93.313	4	23.328	13.035	.000[b]
	Residual	125.273	70	1.790		
	Total	218.587	74			

As observed in table 7 and 8, the value of significance is noted to be 0.000. Such an aspect indicates that the set of alternative hypotheses developed for the study have been met.

Table 9: Coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.240	.907		1.366	.176	-.570	3.050
	IV1	-.314	.089	-.350	-3.527	.001	-.492	-.136
	IV2	.240	.088	.339	2.720	.008	.064	.416
	IV3	.049	.154	.041	.316	.003	-.259	.356
	IV4	.350	.098	.381	3.569	.001	.154	.545

Seen in table 9, the value of significance for all the elements of biassed data collection, lack of privacy, accountability of responsible AI and lack of transparency with ethical and data collection considerations have been less than 0.05. Such an aspect indicates that the hypotheses have been met.

7. Discussion

Data protection is a vital factor governing the usage of data in various fields, in a safe and secure manner. In the case of IT analytics, the inclusion of data privacy and reduction of data leakage is necessary for reducing the two major issues of bias and discrimination. According to the views of Rodrigues (2020), with the presence of biassed data collection tools and tactics, the generation of issues linked with ethical discrimination can occur. As stated by Felländer et al. (2022), responsible AI has been involved in the preservation of data from such ethical grounds, and helped in increasing safety and trustworthiness of the measures. As noted in the findings, the valuations of other elements, apart from the presence of biased information at -0.381, are positive, meaning that an increase of such elements increases the ethical constraints. On the other hand, as per the comments by Mikalef et al. (2022), once the value of significance is lower than 0.05, the respective hypothesis is met. Hence, all the developed hypotheses, from H1 to H4, have been met. With the further investment into responsible AI, a greater generation of data ethics have been seen to be maintained, as the H3 with greater accountability at 0.003 have been proven. According to the comments by Du & Xie (2021), responsible AI is developed in a potential manner to deploy ethical measures during the gathering and the analysis of the digital data.

8. Conclusion

Hence, the study focused on the different types of ethical issues which can arise from the collection of digital data and the application of data analytics, for the examination of the same. The two major issues, as identified in the study, related to the ethical notions of data were seen to be the presence of biases during data interpretation and the lack of privacy of the evidence. Therefore, the study would prove to be important for both academic and the corporate sector, especially of the IT field. With the inclusion of the strategies for the improvement of the AI usage to increase privacy, reduce bias and provide grounds for the responsible utilisation of AI, the discussion was laid down in the study.

References

[1] Anagnostou, M., Karvounidou, O., Katritzidaki, C., Kechagia, C., Melidou, K., Mpeza, E., ... & Peristeras, V. (2022). Characteristics and challenges in the industries towards responsible AI: a systematic literature review. *Ethics and Information Technology*, *24*(3), 37.

https://link.springer.com/article/10.1007/s10676-022-09634-1

[2] Balasubramaniam, N., Kauppinen, M., Kujala, S., & Hiekkanen, K. (2020). Ethical guidelines for solving ethical issues and developing AI systems. In *Product-Focused Software Process Improvement: 21st International Conference, PROFES 2020, Turin, Italy, November 25–27, 2020, Proceedings 21* (pp. 331-346). Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-030-64148-1_21

[3] Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible ai algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, *71*, 1137-1181. http://www.jair.org/index.php/jair/article/view/12814

[4] Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, *129*, 961-974. https://www.sciencedirect.com/science/article/pii/S0148296320305312

[5] Felländer, A., Rebane, J., Larsson, S., Wiggberg, M., & Heintz, F. (2022). Achieving a data-driven risk assessment methodology for ethical AI. *Digital Society*, *1*(2), 13. https://link.springer.com/article/10.1007/s44206-022-00016-0

[6] Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. *European Journal of Information Systems*, *31*(3), 257-268. https://www.tandfonline.com/doi/abs/10.1080/0960085X.2022.2026621

[7] Mohajan, H. K. (2018). Qualitative research methodology in social sciences and related subjects. Journal of economic development, environment and people, 7(1), 23-48. http://ojs.spiruharet.ro/index.php/jedep/article/view/571

[8] Nassar, A., & Kamal, M. (2021). Ethical Dilemmas in AI-Powered Decision-Making: A Deep Dive into Big Data-Driven Ethical Considerations. *International Journal of Responsible Artificial Intelligence*, *11*(8), 1-11. https://neuralslate.com/index.php/Journal-of-Responsible-AI/article/view/43

[9] Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw-hill education (UK). https://books.google.com/books?hl=en & lr= & id=CxUsEAAAQBAJ & oi=fnd & pg=PP1 & dq=spss & ots=n40ABHJ61S & sig=vglonwtVAB6kHqQ46OVH_Xn2hUM

[10] Pandey, P., & Pandey, M. M. (2021). Research methodology tools and techniques. Bridge Center. http://dspace.vnbrims.org:13000/jspui/bitstream/123456789/4666/1/RESEARCH%20METHODOLOGY%20TOOLS%20AND%20TECHNIQUES.pdf

[11] Richardson, S. M., Petter, S., & Carter, M. (2021). Five ethical issues in the big data analytics age. *Communications of the Association for Information Systems*, (1), 18. https://drive.google.com/file/d/1W4oFPuFB-IowtLJ1yzD7zOd1RPF9t0tz/view

[12] Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, *4*, 100005. https://www.sciencedirect.com/science/article/pii/S2666659620300056

[13] Statista. (2022). *Share of employees who believe to have expierenced use of AI by their organization that resulted in ethical issues.*Statista. Retrieved from: https://www.statista.com/statistics/1037914/employees-concerned-ethical-issues-artificial-intelligence/ on 24/12/2023

[14] Tarafdar, M., Teodorescu, M., Tanriverdi, H., Robert, L., & Morse, L. (2020). Seeking ethical use of AI algorithms: Challenges and mitigations. https://deepblue.lib.umich.edu/handle/2027.42/162590

[15] Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: key problems and solutions. *Ethics, Governance, and Policies in Artificial Intelligence*, 97-123. https://link.springer.com/chapter/10.1007/978-3-030-81907-1_8