Artificial Intelligence Applied to Whistleblowing Channels

André Alves[1] (ID)
Rui Mansidão[2] (ID)

Abstract: *The present study addresses the implications of using AI – Artificial Intelligence – and iAP – Public Administration Interoperability – in whistleblowing channels, to improve the effectiveness of these tools for reporting irregularities in organizations.*

To this end, a brief literature review was conducted on whistleblowing channels, AI, and iAP. Subsequently, the aim was to assess the current level of citizens' knowledge about AI and iAP, as well as their levels of trust and perceived usefulness regarding these technologies.

Citizens show an average positive level of knowledge about AI and iAP, with no significant differences between these variables. Similarly, citizens consider these technologies to be trustworthy and useful for application in whistleblowing channels, which constitutes an important starting point for their widespread implementation.

1. INTRODUCTION

Fraud and corruption are complex problems with multiple causes, requiring integrated solutions to address them. Recently, there has been an effort to create mechanisms for preventing fraud and corruption in both public and private organizations. With the publication of the General Regime for the Prevention of Corruption in 2022, it became mandatory for certain entities to develop corruption risk prevention plans and related infractions, codes of conduct, training programs, and the implementation of whistleblowing channels. Regarding the whistleblowing channels, the General Regime for the Protection of Whistleblowers of Infractions states that entities covered by this regime must provide whistleblowing channels that ensure the integrity, confidentiality, and anonymity of both the reports and the whistleblowers (Articles 9 and 13).

Whistleblowing channels are important tools in the fight against fraud and corruption as they serve as a means of communicating irregularities within each organization, which must, in turn, be addressed according to their specific nature, while also considering the whistleblower's rights, as provided in the RGPDI. However, these channels also act as entry points for other types of submissions beyond whistleblowing and are frequently used to file complaints. Inappropriate use of whistleblowing channels results in wasted time and resources during the analysis, processing, and referral stages, which could be avoided if there were a way to guide the potential whistleblower when formulating their submission. From the perspective of organizations, the tasks of analyzing, processing, and referring whistleblowing reports are currently done on an *ad hoc* basis for each report, making them very demanding in terms of human resources and time, and lacking in automation.

There are several whistleblowing channel solutions available in the market for companies and public administration bodies, which comply with the current legal requirements. However, there is

[1] Escola Superior de Tecnologia de Setúbal, Campus do IPS, Estefanilha, 2914-508 Setúbal, Portugal
[2] Escola Superior de Tecnologia de Setúbal, Campus do IPS, Estefanilha, 2914-508 Setúbal, Portugal

currently no solution in Portugal that incorporates AI to, on the one hand, help the whistleblower better formulate their report, in Portuguese or other languages, and on the other hand, assist in the analysis and processing of these reports by each organization. Additionally, there are no whistle-blowing channels that use the interoperability of public administration data.

This study aims to analyze the degree of citizens' knowledge about AI and interoperability, as well as their level of confidence in these technologies. Additionally, it seeks to assess the perceived usefulness of applying AI and interoperability in whistleblowing channels from the citizens' perspective.

2. WHISTLEBLOWING CHANNELS, ARTIFICIAL INTELLIGENCE AND INTEROPERABILITY

2.1. Whistleblowing Channels

The identification of corruption cases and their correct reporting is not always an easy task. The creation of whistleblowing channels has allowed for better communication of irregularities, making them important tools in the fight against corruption.

Article 5, paragraph 1 of the General Regime for the Prevention of Corruption establishes the obligation for the covered entities to adopt and implement a compliance program that includes, at a minimum, a corruption risk prevention plan and related infractions, a code of conduct, a training program, and a whistleblowing channel, in order to prevent, detect, and sanction acts of corruption and related infractions, carried out against or through the entity.

Figure 1. Compliance program
Source: Own research

The whistleblowing channel to be adopted by each entity must comply with the provisions of the **General Regime for the Protection of Whistleblowers of Infractions** Articles 9 and 13 outline the characteristics that whistleblowing channels must possess, namely that they should ensure exhaustiveness, integrity, the confidentiality of the whistleblower's identity or anonymity, the confidentiality of the identity of third parties mentioned in the report, and prevent access by unauthorized persons.

Depending on the specificities of each organization, it may be necessary to implement both an internal whistleblowing channel (Article 8 of the RGPDI) and an external whistleblowing channel (Article 12 of the RGPDI), although both channels may be similar in terms of the fields required for submitting a report. According to Article 9, paragraph 2, internal whistleblowing channels are operated internally for the purposes of receiving and following up on reports by people or services designated for that purpose, and must ensure independence, impartiality, confidentiality, data protection, secrecy, and the absence of conflicts of interest in the performance of their duties. In the case of external whistleblowing channels, according to Article 13, paragraph 2, the competent entities designate employees responsible for handling reports, which includes: a) Providing all

interested parties with information about the reporting procedures, ensuring the confidentiality of the advice and the identity of the persons involved; b) Receiving and following up on reports; c) Providing substantiated information to the whistleblower regarding the measures planned or adopted to follow up on the report and requesting additional information if necessary.

In terms of form and admissibility, internal whistleblowing channels allow the submission of reports, either in writing or verbally, by workers, anonymously or with identification of the whistleblower (Article 10, paragraph 1). The report may be submitted using electronic authentication methods, such as the citizen's card or mobile digital key, or through other electronic identification methods issued in other Member States and recognized for this purpose under Article 6 of Regulation (EU) No 910/2014 of the European Parliament and Council, dated July 23, 2014, as long as these means are available (Article 10, paragraph 3). Regarding external whistleblowing, the form and admissibility are similar, although electronic authentication is not provided for in this case (Article 14).

Each organization must choose the whistleblowing channel to implement and designate staff to perform the functions mentioned above. Currently, there are several solutions on the market that offer whistleblowing channels in compliance with the legal requirements in force. However, it is not known whether there is any solution in Portugal that incorporates AI in the analysis and processing of reports, as well as enabling data interoperability between entities.

In the literature, we find some references related to whistleblowing channels and its technological developments. Silva and Sousa (2017) provide an analysis of the influence of reporting channels on the decision to report financial irregularities. The authors concluded that the presence of an anonymous whistleblowing channel increases the likelihood that employees with permanent contracts will report financial misconduct. Bazart et al. (2019) investigate the relative effectiveness of whistleblower-based audits versus random audits, concluding that whistleblower-based audits are more efficient and have a stronger deterrent effect, promoting higher levels of compliance.

Noor and Mansor (2019) analyze the implementation of whistleblowing channels in the Malaysian public sector using artificial intelligence. The results indicated that the existence of whistleblowing channels strengthens the internal control system and increases the transparency and accountability of the public sector. Gao (2020) highlights the use of artificial intelligence in these channels and compares various means of submitting reports. The results showed no significant differences in terms of the perception of anonymity between online whistleblowing channels and telephone reporting.

 Zouvia et al. (2020) identify the potential of AI applied to whistleblowing channels, particularly in reporting through virtual assistants, live translation, more efficient verification, and proactive monitoring. Putra et al. (2022) discuss various issues, including how the existence of whistleblowing channels and big data influence fraud prevention, and how fraud prevention is interconnected with internal auditing, risk management, whistleblowing channels, and big data analysis to prevent criminal behavior. Mrowiec (2022) presents a literature review on the factors that influence internal whistleblowing in companies, which can be grouped into the following areas: ethics, leadership, policies and procedures, retaliation and safeguards, social climate, organizational justice, training, whistleblowing channels, communication, additional motivations, organization size and structure, and audit committees.

Previtali and Cerchiello (2022) analyze organizational factors that influence whistleblowing behavior, such as the implementation of reporting procedures and anti-corruption training programs.

The authors suggest that, in addition to anti-corruption laws, it is essential to foster a cultural shift within organizations, supported by training and procedures that protect whistleblowers, in order to preserve organizational integrity.

Kagias et al. (2023) present a maturity model for whistleblowing practices, which can be used to strengthen organizations' reporting systems, thereby providing a more robust control environment. The model also allows for internal and external benchmarking to assist organizations in measuring their progress in relation to compliance and the effectiveness of their reporting channels.

Kang (2023) presents a literature review on whistleblowing in the public sector, with the primary objective of analyzing definitions, theories, research methods, and factors related to the intention and behavior of reporting wrongdoing. The study concludes that whistleblowing is a critical practice for promoting transparency and administrative efficiency.

The Report to the Nations study (ACFE, 2024) indicates that the most effective method for detecting occupational fraud is "Whistleblowing," which accounts for 43% of the recorded cases. Occupational fraud is an offense related to corruption, characterized by an unlawful act in which an individual exploits their position or role to gain undue advantages and personal enrichment.

In summary, whistleblowing channels are crucial tools for detecting and preventing corruption, and AI and data interoperability present future opportunities for enhancing their effectiveness.

2.2. Artificial Intelligence

Artificial Intelligence (AI) refers to the ability of computer programs to perform tasks that normally require human intelligence. Instead of being explicitly programmed to carry out a specific task, these programs are trained to learn and improve their performance over time. AI integrates sciences, theories, and techniques (such as mathematical logic, statistics, probability, computational neurobiology, and computer science) to mimic the cognitive abilities of a human through a machine (AMA, 2022).

AI encompasses a variety of approaches and techniques, including Machine Learning, Neural Networks, Natural Language Processing, Computer Vision, and Recommendation Systems.

The anticipated benefits of applying AI to whistleblowing channels align with practical examples of AI use in the Administrative Management sector, namely: "Provision of automated conversational interfaces with virtual assistants to streamline citizen and business support scenarios; Enhancement of security and privacy in IT systems" (AMA, 2022). Additionally, Noor and Mansor (2019) highlight that the use of AI in whistleblowing channels improves responses to corrupt practices, advocating for its implementation across all public sector entities.

2.3. Interoperability in Public Administration

The iAP aims to provide an easy and integrated method for offering cross-sector electronic services. It consists of the ability of organizations to interact and act for common benefits through communication and the sharing of information and knowledge. The iAP is a service-oriented central platform, that intends to provide the Public Administration (PA) with a shared tool that allows for the interconnection of PA systems and the composition and provision of multichannel electronic services that are closer

to the needs of citizens and businesses through four independent services: Integration, Messaging, Payments, and Authentication. This service is based on the concept of providing shared services among various entities, to simplify the availability of these services to the public. The iAP provides online services and the respective management, secure communication between Information Systems, online payments, and messaging services, among other functionalities. Interoperability in the Public Administration has gained relevance and importance regarding the creation and evolution of digital public services, to combat fragmentation and promote the creation of a single, unified ecosystem.

According to the European Interoperability Framework, created by the European Commission and transmitted to all member countries, the interoperability model is composed of four layers of governance, which must be integrated into the development of digital public services:

- **Legal interoperability**: Enables organizations operating under different legal frameworks, policies, and strategies to work together. It may require that the legislation does not hinder the establishment of European public services within and between Member States, and clear agreements on how to handle cross-border legislative differences, including the option of enacting new legislation, are necessary.
- **Organizational interoperability**: Refers to how entities align their processes, responsibilities, and expectations to achieve a common goal and benefits around interoperability. In practice, organizational interoperability assumes the documentation and integration of business processes where relevant information is shared.
- **Semantic interoperability**: Ensures that the format and understanding of shared data and information are preserved and understood throughout all exchanges. In practical terms, it ensures that "what is sent is what is understood." In addition to promoting the alignment of the semantic concepts of services for the citizen, semantic interoperability also promotes the alignment of syntax, i.e., the format of the information and the grammar used.
- **Technical interoperability**: The layer that allows for the technological connection between different systems and services, such as interfaces, tools, back-end and front-end systems, data platforms, payment systems, notification systems, and more. In Public Administration, technical interoperability is ensured by the iAP, which enables the exchange of information between different entities and the composition and provision of multichannel electronic services that better meet the needs of citizens and businesses in an agile manner and with economies of scale.

Almeida Prado Cestari et al. (2020) emphasize the importance of interoperability for efficiency and collaboration among public sector entities. Additionally, they propose a diagnostic model, the Public Administration Interoperability Capability Model (PAICM), designed to assess and enhance the interoperability of public organizations in terms of information exchange, processes, and systems.

In the study by Džino et al. (2023), the challenges and opportunities associated with digitalization and interoperability in the public sector of Bosnia and Herzegovina are presented. According to this study, digitalization and interoperability are essential for enhancing public administration by increasing productivity and creating service opportunities. Furthermore, the adoption of these technologies offers benefits not only for institutions but also for the economy and citizens.

There are few studies in the literature that reference the use of data interoperability in whistleblowing channels. Only Putra et al. (2022) identifies that the integration of AI and interoperability in the Public Administration can improve whistleblowing channels, increase efficiency, save costs, and enhance competitiveness.

3. HYPOTHESIS, METHODOLOGY AND RESULTS

The main research question inherent to this work is to understand whether citizens consider AI and interoperability useful when applied to whistleblowing channels. To do this, it is necessary to assess the level of knowledge about these technologies and the level of confidence in them.

Surveys were conducted with 50 randomly selected citizens, asking them about their level of knowledge regarding AI and iAP. On a scale of 0-10, where 0 = "Non-existent knowledge" and 10 = "Very high knowledge", the average knowledge levels were 6.654 and 5.340, respectively. Subsequently, a paired T-student test was calculated for the two samples, and it was concluded that there is not enough evidence to reject the null hypothesis, meaning there is no significant difference between the level of knowledge about AI and iAP (p-value = 0.110).

Although the level of knowledge in AI is slightly higher than the level of knowledge in iAP, this difference is not statistically significant, suggesting that, in general, the respondents have similar levels of knowledge on both topics. The distributions show that participants tend to have a higher level of knowledge in AI, but with less variation, while the knowledge in iAP is more dispersed.

Table 1. Descriptive statistics and hypothesis testing for the level of knowledge about AI and iAP

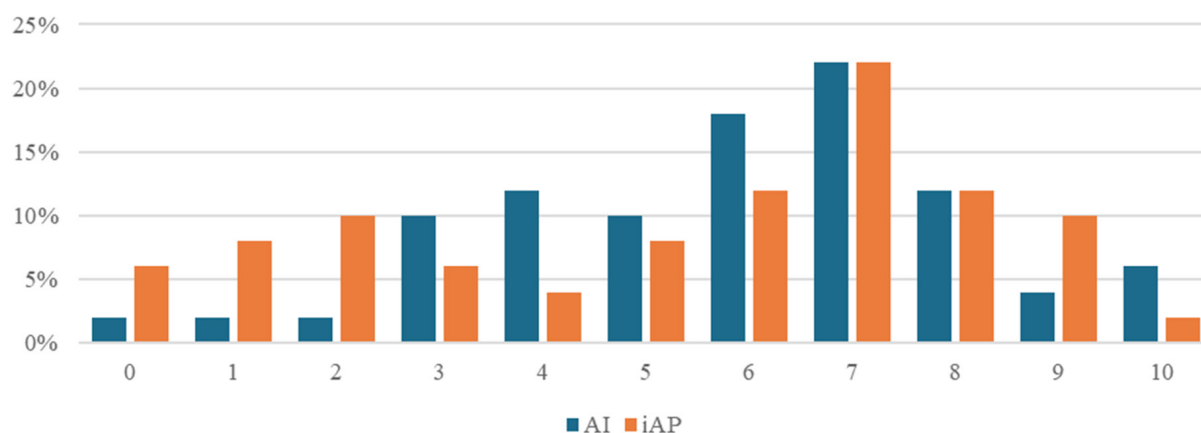|	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis	T-student
Level of AI knowledge	1	10	6,654	2,121	-0,164	-0,371	1.627
Level of iAP knowledge	1	10	5,340	2,861	-0,456	-1,015	(0.110)

Source: Own calculations


Figure 2. Level of knowledge about AI and iAP
Source: Own calculations

Regarding the level of confidence in use these technologies, the average values were 6.4 for AI and 7.042 for iAP. Similarly to the level of knowledge, a paired t-test was calculated, resulting in a p-value of 0.09, indicating that there is not enough evidence to reject the null hypothesis that there is a significant difference between the confidence levels in using a whistleblowing channel equipped with AI and iAP. Overall, both AI and iAP are viewed positively by the respondents, with iAP being considered slightly more trustworthy.

Table 2. Descriptive statistics and hypothesis testing for the level of confidence about AI and iPAQ

	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis	T-student
Level of AI confidence	0	10	6,400	2,491	-0,685	-0,021	1.729
Level of iAP confidence	0	10	7,042	2,713	-1,007	0,382	(0,090)

Source: Own calculations
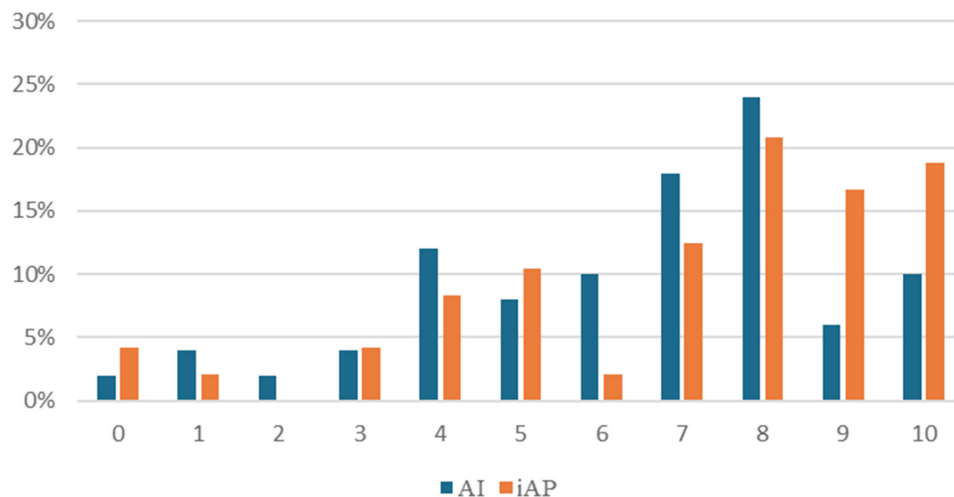
Figure 3. Level of confidence about AI and iAP

Source: Own calculations

Participants were also asked about the usefulness of integrating:
- AI in a whistleblowing channel, to assist the whistleblower in reporting the complaint;
- AI in a whistleblowing channel, as a way to assist in the processing of each complaint;
- iAP in a whistleblowing channel, to improve data transfer between entities.

Table 3. Descriptive Statistics and Hypothesis Testing for the level of usefulness in integrating AI and iAP into a whistleblowing channel

	Min.	Max.	Mean	Std. Dev.	Skewness	Kurtosis	ANOVA
Level of usefulness in integrating AI for reporting	0	10	6,100	2,794	-0,563	-0,398	1.538 (0.218)
Level of usefulness in integrating AI for treatment	0	10	6,680	2,766	-0,798	-0,338	
Level of usefulness in integrating iAP	0	10	7,265	2,699	-1,097	0,723	

Source: Own calculations

The three variables present relatively close averages, all between 6 and 7, indicating that the respondents consider both AI for assisting in reporting and processing, as well as iAP, to be useful. Although iAP has the highest average (7.265) and AI for reporting has the lowest (6.100), the differences are not large enough to be considered significant, as confirmed by the result of the ANOVA test.

The negative skewness in all variables suggests that respondents tend to assign higher values of usefulness, with interoperability being the most favored. The kurtosis values range from slightly platykurtic to leptokurtic, indicating that while some distributions are more flattened, the distribution for interoperability is more concentrated around the mean.

In conclusion, the three variables are considered similarly useful by respondents, with interoperability being slightly more valued. However, these differences are not statistically significant, as indicated by the p-value from the ANOVA.
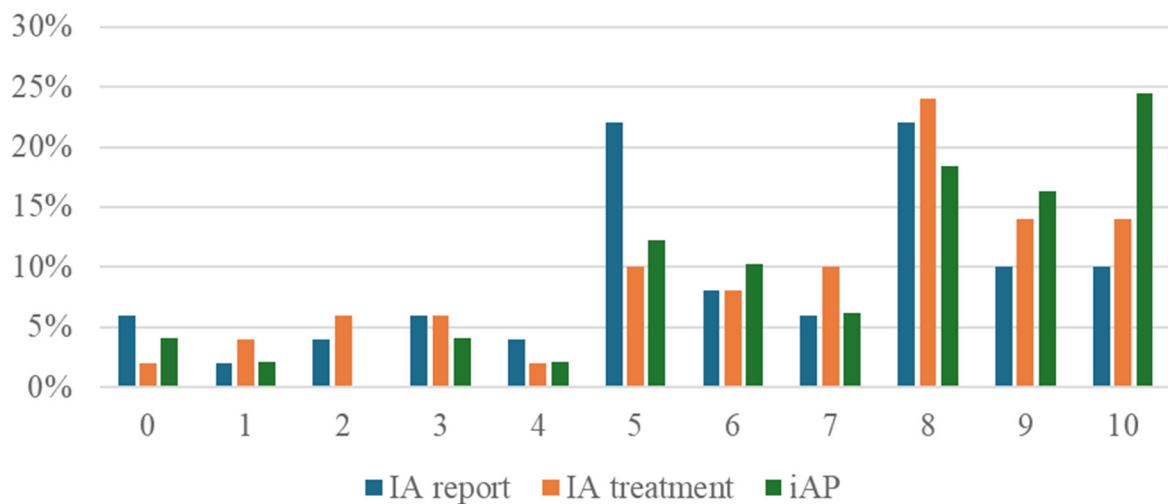

Figure 4. Level of usefulness of AI and iAP

Source: Own calculations

4. FUTURE RESEARCH DIRECTIONS

Based on the results, which indicate that AI and iAP are technologies considered useful within the scope of whistleblowing channels, it may be important to analyze other factors that influence perceived usefulness, such as the user experience with channels that incorporate this technology compared to traditional channels.

Once data on the use of these technologies is available, longitudinal studies on the application of AI and iAP in whistleblowing channels can be conducted, as well as comparative studies across different sectors (e.g., public vs. private, healthcare vs. education) to analyze in which cases greater gains in terms of utility are observed.

Another potential area of research lies in analyzing the reduction of the workload currently carried out by employees, which in the future could be handled by AI, particularly in the classification of reports, identification of false reports, prioritization based on the severity of the reports, and other tasks.

Finally, it would be important to analyze the usefulness of a single whistleblowing channel in Public Administration, equipped with AI and iAP, to achieve synergies in terms of saving financial and human resources, improving the efficiency of information exchange, and enhancing the speed of processes.

5. CONCLUSION

Through this study, it is possible to conclude that there are several advantages to the use of AI and iAP applied to whistleblowing channels, both from the whistleblower's perspective when reporting and from the perspective of the entities responsible for handling such reports. These advantages translate into process efficiency, foreseeing lower costs in the analysis and handling of cases.

Citizens show an average positive level of knowledge regarding AI and iAP, with no significant differences between these variables. Similarly, citizens consider these technologies reliable and useful for application in whistleblowing channels, which serves as an important starting point for their implementation.

References

ACFE. (2024). The Nations ® Occupational Fraud 2024: 2 Foreword Occupational Fraud 2024: A Report to the Nations.

Almeida Prado Cestari, J. M., Loures, E. de F. R., Santos, E. A. P., & Panetto, H. (2020). A capability model for public administration interoperability. *Enterprise Information Systems*, *14*(8), 1071–1101. https://doi.org/10.1080/17517575.2018.1564154

AMA. (2022). Guia para a Inteligência Artificial - Guia para uma inteligência artificial ética, transparente e responsável na administração pública. *AMA – Agência para a Modernização Administrativa, I.P.*

Bazart, C., Beaud, M., & Dubois, D. (2019). *Whistleblowing vs. Random Audit: An Experimental Test of Relative Efficiency*.

Džino, J., Džino, S., & Injac, D. (2023). Analysis of public administration, effects and impact of digitalization and interoperability in public administration. *JITA, 13*, 48–53. https://doi.org/10.7251/JIT2301048D

Gao, L. (2020). *Technology Advancements for Whistleblowing Reporting Platforms and Employees' Decision to Blow the Whistle.* https://doi.org/10.2308/API-2020-001

Kagias, P., Sariannidis, N., Garefalakis, A., Passas, I., & Kyriakogkonas, P. (2023). Validating the Whistleblowing Maturity Model Using the Delphi Method. *Administrative Sciences*, *13*(5). https://doi.org/10.3390/admsci13050120

Kang, M. M. (2023). Whistleblowing in the Public Sector: A Systematic Literature Review. *Review of Public Personnel Administration*, *43*(2), 381–406. https://doi.org/10.1177/0734371X221078784

Mrowiec, D. (2022). Factors influencing internal whistleblowing. A systematic review of the literature. *Journal of Economics and Management*. https://doi.org/10.22367/JEM.2022.44.07

Noor, N. R. A. M., & Mansor, N. (2019). Exploring the Adaptation of Artificial Intelligence in Whistleblowing Practice of the Internal Auditors in Malaysia. *Procedia Computer Science*. https://doi.org/10.1016/J.PROCS.2019.12.126

Previtali, P., & Cerchiello, P. (2022). Organizational Determinants of Whistleblowing. A Study of Italian Municipalities. *Public Organization Review*, *22*(4), 903–918. https://doi.org/10.1007/s11115-021-00554-0

Putra, I., Sulistiyo, U., Diah, E., Rahayu, S., & Hidayat, S. (2022). The influence of internal audit, risk management, whistleblowing system and big data analytics on the financial crime behavior prevention. *Cogent Economics & Finance*. https://doi.org/10.1080/23322039.2022.2148363

Silva, G. R., & Sousa, R. G. (2017). The influence of the anonymous whistleblowing channel on accounting fraud detection in organizations. *Revista de Contabilidade e Organizações*, *30*, 46-57.

Zouvia, K., Investigator, S., & Ombudsman, G. (2020). *Artificial Intelligence and Whistleblowing: Can A.I. be useful for Whistleblowing processes?* http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=12302