Artificial Intelligence and Whistleblowing: Can A.I. be useful for Whistleblowing processes?

Kalliopi Zouvia

Lawyer, DPO and Senior Investigator
Greek Ombudsman
PhD Candidate Panteion University
of Social and Political Sciences

zouvia@synigoros.gr

ABSTRACT

In recent years, there is a constantly growing discussion about the contribution of whistleblowers and the need to protect them. Another rising topic of discussion is the increasing use of Artificial Intelligence (AI), the gains from it, the hardships, the moral dilemmas. This article refers to the development of whistleblower protection in the EU and discusses how AI can be applied in a whistleblowing context.

CCS Concepts

• **Applied Computing → Law, social and behavioral sciences → Law**

• **Computing Methodologies → Artificial Intelligence**

Keywords

Artificial Intelligence; Whistleblowing; Law Technology.

1. INTRODUCTION

In the last decade a lot of important whistleblower cases came to light, such as the E. Snowden case [1], Lux Leaks [2], Panama Papers, the Novartis case. In the last few months the value of whistleblowing became even more obvious in the context of battling the global COVID-19 pandemic. Dr Li Wenliang was one of a group of medics who were reprimanded by local Chinese authorities for sharing information about the early cluster of patients with SARS-like symptoms admitted to Wuhan hospitals in December 2019. A few months later, healthcare professionals, in several countries around the world, came forward with alerts about shortages of personal protective equipment, which were forcing them to work in unsafe conditions [3].

One of the definitions of the term whistleblowing is "the disclosure by organization members (former or current) of illegal, immoral or illegitimate practices under the control of their employers, to persons of organizations that may be able to effect action" [4].

The phenomenon of individuals coming forward to speak about important issues has a long history, but the term 'whistleblowing' came to prominence in the US in the 1960 and early 70's. Some of the fist law intended to protect whistleblowers emerged in the US.

In Europe, the Committee on Legal Affairs and Human Rights of the Parliamentary Assembly of the Council of Europe submitted, in 2009, a report which concluded that although whistleblower laws and regulations existed in the various states, more action was needed at national level in order to implement them effectively [5]. In 2010 the Assembly of the Council of Europe adopted Resolution 1729, calling on all Member States to review their whistleblower protection legislation following specific guidelines [6].

In 2014 the Council of Europe developed a legal instrument on protecting individuals who report or disclose information on acts and omissions in the workplace that represent a serious threat or harm to the public interest. The Recommendation 2014/7 set out a series of principles to guide member States when reviewing their national laws or when introducing legislation and regulations or making amendments that are necessary and appropriate in the context of their legal system [7].

It was after major scandals, such as LuxLeaks, that the European Parliament took steps towards the protection of whistleblowers and in 2017 the plenary Assembly adopted a Resolution on legitimate measures to protect whistleblowers acting in the public interest when disclosing confidential information of companies and public bodies [8].

Finally, in April 2019, the EU passed the Whistleblower Directive or Directive 2019/1937 of the European Parliament and of the Council on the protection of persons who report breaches of Union Law and outlines a basic standard of protection that will apply across all 27 member states of the EU [9]. The Directive sets rules for who can be protected as a whistleblower and the kinds of issues they can make reports on. It defines how the idea of protecting whistleblowers interacts with other legal and ethical obligations. The Directive needs to be transposed into national legislation by the end of December 2021.

The Directive refers to protecting individuals who report breaches of EU Law and its purpose, as set out in Article 1, is to enhance the 'enforcement of Union law and policies in specific areas', such as public procurement, financial services, safety of products, transport safety, protection of the environment.

A lot has already been written in the aftermath of the adoption of the Directive [10]. In this article only a few key points are going to be mentioned, in order then to examine whether artificial intelligence could be used in whistleblower protection.

Regarding the reporting channels, the Directive follows the three –tiered model [11], which means that whistleblowers can report internally [12], within a legal entity, externally to competent authorities [13] or, as a last resort, they can disclose their information publicly, to the media [14].

Confidentiality and anonymity are really important for whistleblower protection. When we talk about confidentiality it means that the identity of a whistleblower is known to specific individuals, whereas anonymity means that a whistleblower is not known at all. According to the Directive 'Member States shall ensure that the identity of the reporting person is not disclosed to anyone beyond the authorized staff members competent to receive or follow up on reports, without the explicit consent of that person. This shall also apply to any other information from which the identity of the reporting person may be directly or indirectly deduced' [15]. As far as anonymity is concerned, the Directive lets Member States decide whether organizations and competent authorities are required to accept and follow up on anonymous reports. This is not in line with current best practice. While anonymous disclosures can make it harder to investigate a concern, this should not prevent a concern being taken seriously [10]. In order to secure the identity of the whistleblower there are hotlines and technology has been a game changer in anonymous whistleblowing, with the advent of the anonymous online drop boxes, which use encryption and other privacy enhancing technologies to obscure the identity of the reporting person.

A timely response after a report is also important. According to the Directive, a reporting person is entitled to have their report acknowledged within seven days and receive feedback within three months [12].

There is not much data about whistleblower reports in Europe. One of the few countries for which there is data is the Netherlands: according to the Dutch Whistleblowers Authority 2018 Annual Report, a total of fifty requests for investigation had been submitted between 2016 and 2018. Of these, 28 were declared inadmissible [16].

In the US, however, there is an increase in whistleblower reports in the last few years. According to the 2019 US Securities and Exchange Commission Annual Report to Congress on the Whistleblower Program, which provides incentives and protection to whistleblowers, the Commission received more than 5.000 reports, which represents a 74% increase since whistleblower data collection started in 2012 [17].

The Occupational Safety and Health Administration, a government agency responsible for protecting workers in a wide range of industries, received in 2018 more than 9.500 reports, creating a significant backlog in light of a reduction in the number of investigators. More than 3.000 of those complaints resulted in a full investigation. With such great workload, it is important to have the means to react quickly and in the most appropriate way. This raises the question whether Artificial Intelligence could be of use.

2. ARTIFICIAL INTELLIGENCE FOR WHISTLEBLOWER PROTECTION

Artificial Intelligence is distinguished in two big general categories. Artificial Narrow Intelligence (ANI), sometimes referred to as Weak AI, that specializes in one area and Artificial General Intelligence (AGI), sometimes referred to as Strong AI, or Human-Level AI. Artificial General Intelligence refers to a computer that is as smart as a human across the board — a machine that can perform any intellectual task that a human being can [18].

AI is a collection of technologies that combines data, algorithms and computing power. The evolution of Artificial Intelligence has been rapid in recent years with the result that it is at the center of

the "Digital Single Market" [19]. AI can be used to bring the benefits of the technology to society and economy, for citizens, for business development and for services of public interest [20].

As far as whistleblowing is concerned, thoughts are already being made about the use of Artificial Intelligence in whistleblowing practice [21]. In the following sections, we elaborate on possibilities of using AI for whistleblowing in three areas: **reporting systems**, **vetting process automation** and **proactive monitoring**.

2.1 AI for whistleblowing reporting systems
Advances in technology are already influencing changes in the way speak up reports are performed. AI can be a technology enabler for whistleblower reporting systems in several ways.

Virtual AI agents (chatbots) can assist or entirely handle the submission of whistleblower reports. Chatbots are AI-enabled software applications that can conduct conversations with human users with voice or text [22]. They use natural language processing (NLP) capabilities to recognize what their human interlocutors say and to formulate responses that try to resemble what a human agent would reply. Chatbots are widely used today in a range of industries, for example for customer care in telecommunications, banking, and healthcare. They are particularly effective as interactive guides of users to structured processes or transactions, such as submitting an issue report to a telecommunications provider, reporting a lost credit card or scheduling a medical appointment. Virtual agents can be helpful for whistleblowers submitting an allegation by interactively providing instructions throughout the process. They can also help reduce the number of incomplete or non-eligible reports by advising users on what are the requirements for a submission and the scope of handled cases. A recent study reports that whistleblowers were more likely to report to an online platform when a virtual agent handled the reports because they believed that it is more efficient and provides greater control while reporting [23].

AI-based live translation can help increase the accessibility of whistleblowing reporting services across ethnic and cultural communities. Automatic real time translation can enable interactive cross-language reporting through hotlines, i.e. over phone to a human agent or through a chatbot (voice or text). Voice-based hotlines are a very common submission channel for whistleblowers. For example, it is by far the most frequent reporting method (37% of reports in 2019) in corporate ethics compliance systems, as reported by the largest provider of ethics reporting software in the US [24]. Real-time machine translation is a relatively mature technology, fueled particularly by advances in neural networks and deep learning [25]. There are commercial offerings of translation-as-a-service by major technology providers which can be used by the developers of whistleblowing applications to allow cross-lingual communication between submitters of whistleblowing reports and handing agents on a 24/7 basis without the huge costs of having stand-by human interpreters for multiple languages [23]. There are already real-life applications of live translation in other domains, such as communication of teachers with parents of different cultures in middle schools in the US [26].

2.2 AI for making the vetting process more efficient
As seen from the above-mentioned scandals and a lot of others that see the light all the more often the last years, there is a lot of misconduct. While whistleblower protection evolves, many organizations will engage in some type of whistleblowing

investigation, whether it is conducted internally or from an outside agency. It is important for an organization to give the tone from the top and cultivate a culture that promotes ethical behavior and a speak-up culture. Being able to understand the motives of whistleblowers is also important, but vetting and investigating the complaints is critical. To this end, one must have the appropriate auditing tools.

Currently, the impact of AI is mostly discussed in financial audits and is especially pronounced in the area of data acquisition (data extraction, comparison, and validation). This means that AI-enabled technology can locate relevant information, extract it from documents, and make it usable for the human auditor, who can devote more time to areas requiring higher-level judgment. For example, AI enables full automation of time-consuming tasks such as payment transaction testing, including extraction of any supporting data for further substantive testing [27][28].

The procedural aspect of a whistleblower investigation is more or less the same as any other investigation that corporations are subjected to. What is critical in an internal whistleblower investigation is to determine if the allegation is true [29]. Whistleblower reports are usually followed by a great amount of data. Sometimes though only a small part of this data may actually be relevant to the action or omission that represents a serious threat or harm. Advanced analytic tools and AI can be used in order to locate and extract critical information from whistleblower reports. Artificial Intelligence in the form of machine learning can group similar documents for faster review. Data sampling is useful to provide insight into larger data.

One could distinguish between the use of AI for structured data analytics and its use for analytics over unstructured data [30]. Structured data is comprised of clearly defined data types whose pattern makes them easily searchable; while unstructured data – "everything else" – is comprised of data that is usually not as easily searchable, including formats like audio, video, and social media postings

In structured data analytics, AI and analytics tools can be set up to look for example for transactions that exceed norms, transactions with vague or missing detail, rush requests, unusual cash disbursements, or transactions that circumvent typical approval processes. All these could be characterized as red flags and could direct the investigator towards the person or persons responsible for those transactions. On the other hand, in unstructured data (documents, email, and other messaging) keyword language filters could be used in order to expose use or change in language that may indicate unethical or noncompliant behavior.

2.3 AI for proactive monitoring

Artificial Intelligence can also be used for proactive monitoring, in order to identify potential areas of risk. If for example a certain type of bad behavior is becoming more common in an organization's industry sector, that particular subject matter could be targeted for routine monitoring using AI and linguistic and analytics tools to flag worrisome language or sentiment in data stores. Analytical tools and predictive software will enable companies to combine whistleblowing data with information from across the business and identify where problems are most likely to occur in the future [28].

3. CONCLUSIONS – FUTURE OUTLOOK

More organized steps were taken just recently in order to better regulate whistleblower protection in the EU and already there are thoughts about the use of AI in whistleblower protection. Even the idea that robots could replace whistleblowers is being explored as a potential future development [31]. If robots gradually take the place of workers (let's say in the automotive industry), will there be a time that robots conceive and report wrongdoings? One could argue that the 'penalisation' of the whistleblower will no longer exist [32].

Nevertheless, the potential replacement of whistleblowers by AI raises a lot of questions, some of them being whether a robot will know what to blow the whistle on and if it will be able to follow the three tiered model. Moreover, depending on the country and legislation, whistleblowers can report illegalities, irregularities, wrongdoings or immoral actions. Can a robot know immoral actions? [31] Scandals as LuxLeaks did not entail illegal acts but obscure legal practices that were considered immoral by society. Will a machine ever be able to report such acts even internally? Artificial Intelligence is a useful tool in detecting wrongdoing. Nevertheless, a fair balance between human actions and the use of AI will always be a crucial aspect.

4. REFERENCES

[1] Hackett, K. (2013). Edward Snowden: The New Brand of Whistle-Blower? *Quill Magazine.* 101, 5 (Sept/Oct 2013), 26-31.

[2] Tansini, C. (2018). The Luxleaks Whistleblowing Case: a Philosophical Assessment.

[3] Abazi, V. (2020). Truth distancing? Whistleblowing as remedy to censorship during covid-19. *European Journal of Risk Regulation*, *11*(2), 375–381.

[4] Near, J.P., Miceli, M.P. (1985) Organizational dissidence: The case of whistle-blowing. *Journal of Business Ethics* 4, 1–16.

[5] Protection of "whistle-blowers", Doc. 12006 (14 September 2009), Report of the Committee on Legal Affairs and Human Rights, [http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=12302]

[6] Vandekerckhove, W. and Lewis, D. 2011. The Content of Whistleblowing Procedures: A Critical Review of Recent Official Guidelines. *Journal of Business Ethics.* 108, 2 (Nov. 2011), 253–264.

[7] Protection of whistleblowers, Recommendation CM/Rec (2014) 7 adopted by the Committee of Ministers of the Council of Europe on 30 April 2014 and explanatory memorandum [http://rm.coe.int/doc/09000016807096c7]

[8] European Parliament resolution of 24 October 2017 on legitimate measures to protect whistle-blowers acting in the public interest when disclosing the confidential information of companies and public bodies (2016/2224(INI)) https://www.europarl.europa.eu/doceo/document/TA-8-2017-0402_EN.html

[9] Directive 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report breaches of Union law [2019] OJ L 305/17 https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32019L1937

[10] Terracol, M. 2019. Building on the EU Directive for Whistleblower Protection, Analysis and Recommendation. Transparency International Position Paper #1 / 2019. ISBN: 978-3-96076-127-3.

[11] Vandekerckhove, W. 2010. European Whistleblower Protection: Tiers or Tears? in *D. Lewis (ed.), A Global Approach to Public Interest Disclosure* (Cheltenham: Edward Elgar, 2010).

[12] Chapter II Internal Reporting and Follow Up, Art. 7-9 in Directive 2019/1937 of the European Parliament.

[13] Chapter III External Reporting and Follow Up, Art. 10-14 in Directive 2019/1937 of the European Parliament.

[14] Chapter IV, Public Disclosures, Art. 15 in Directive 2019/1937 of the European Parliament.

[15] Chapter V, Art. 16 Duty of confidentiality in Directive 2019/1937 of the European Parliament.

[16] Dutch Whistleblowers Authority. March 2019. Annual Report 2018. https://www.huisvoorklokkenluiders.nl/Publicaties/jaarverslagen/2019/03/14/annual-report-2018---dutch-whistleblowers-authority

[17] US Securities and Exchange Commission. 2019. Whistleblower Program 2019 Annual Report to the US Congress. https://www.sec.gov/files/sec-2019-annual-report-whistleblower-program.pdf

[18] Tassis S. 2018. The era of Artificial Intelligence, *Media and Communication Law* (ΔιMEE) 4/2018, 484-494.

[19] Voras A., Mitrou L. 2018. Artificial Intelligence and Data Protection: A view under the light of the General Data Protection Regulation. *Media and Communication Law* (ΔιMEE) 4/2018, 460-466.

[20] European Commission. European Data Strategy. COM(2020) 66 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066

[21] Mohd Noor, N.R.A. and Mansor, N. 2019. Exploring the Adaptation of Artificial Intelligence in Whistleblowing Practice of the Internal Auditors in Malaysia. *Procedia Computer Science*. 163, (2019), 434–439.

[22] Dale, R. 2016. The return of the chatbots. *Natural Language Engineering* 22(5) 811-817.

[23] Gao, L. 2020. Technology Advancements for Whistleblowing Reporting Platforms and Employees' Decision to Blow the Whistle. *Accounting and the Public Interest*. (Oct. 2020). DOI:https://doi.org/10.2308/api-2020-001

[24] Penman, C. (2020). 2020 Risk & Compliance Hotline Benchmark Report. *Navex Global Benchmark Reports*. Retrieved from https://www.navexglobal.com/en-gb/resources/benchmarking-reports/2020-risk-compliance-hotline-benchmark-report

[25] Koehn, P. (2020). Neural machine translation. *Cambridge University Press*. ISBN: 9781108601764.

[26] Le Santo, G. Real-time language translation boosted by AI. *Orange Business Services Blogs*. Online article: https://www.orange-business.com/en/blogs/language-translation-boosted-by-ai

[27] Kokina, J. and Davenport, T.H. 2017. The Emergence of Artificial Intelligence: How Automation is Changing Auditing. *Journal of Emerging Technologies in Accounting*. 14, 1 (Mar. 2017), 115–122.

[28] Ghanoum, S., & Alaba, F. M. 2020. Integration of Artificial Intelligence in Auditing: The Effect on Auditing Process. 1–75. *Master's Thesis. Kristianstad University*. Spring 2020.

[29] Grobart, J. 2020. Preparing for Whistleblower Complaints. *Strategic Finance*. September 2020. https://sfmagazine.com/post-entry/september-2020-preparing-for-whistleblower-complaints/

[30] O'Leary, D.E. 2013. Artificial Intelligence and Big Data. *IEEE Intelligent Systems*. 28, 2 (Mar. 2013), 96–99. DOI:https://doi.org/10.1109/mis.2013.39.

[31] Kafteranis, D. 2019. Can artificial intelligence replace whistle-blowers in the business sector. *International Journal of Technology Policy and Law*. 3, 2 (2019), 160. DOI:https://doi.org/10.1504/ijtpl.2019.104948.

[32] Waytz, A. 2014. Why Robots Could Be Awesome Whistleblowers. *The Atlantic*. October 2014. https://www.theatlantic.com/business/archive/2014/10/why-robots-could-be-awesome-whistleblowers/381216/