CEJ
COMPLIANCE
ELLIANCE
JOURNAL

FROM WHISTLEBLOWING TOOLS TO AI-SUPPORTED DATA ANALYSIS

A compliance practitioner`s view on IT-tools for different aspects of investigations

Markus Endres

AUTHOR

The author of this whitepaper is an attorney-at-law with a strong focus on criminal healthcare compliance. He is an accredited member of the Association of Certified Fraud Examiners (ACFE). After working in the healthcare compliance practice of a Big4 firm, he now serves as Chief Compliance Officer for a global pharmaceutical company in Germany. Markus also supports the CEJ as a member of the Advisory Board.

TABLE OF CONTENTS

I. INTRODUCTION

For some years now, and accelerated again by the Corona pandemic, the workplace has been experiencing a significant change and is becoming increasingly digital. This also leads to new potential white-collar crimes: cybercrime[1] serves as a new buzzword and there are plenty new digital business models with still little-known vulnerabilities for fraud. The compliance and forensics community must respond to these challenges with new and innovative investigation approaches. Forensics will have to become more digital to keep up with this new workplace reality. In this regard, IT-tools can support forensic and compliance functions in various steps of an investigation. This whitepaper aims to provide an overview of the chances provided by digital forensic tools, but also to highlight their potential risks. In addition, it describes the specific challenges for a use of AI-based forensic tools by government agencies.

II. EUROPEAN WHISTLEBLOWER PROTECTION LEGISLATION LEADS TO A BOOM IN WEB-BASED WHISTLEBLOWING TOOLS

Web-based whistleblowing tools are currently very popular in Germany due to the Whistleblower Protection Act[2], which came into force on July 2, 2023. But since the Whistleblower Protection Act finds its legal basis in the EU whistleblower directive[3] that applies to all EU member states and is to be implemented in their local laws[4], the same is true throughout Europe. This European whistleblower protection initiative is based on the lawmaker`s understanding that the implementation of transparent whistleblowing processes is an important tool to foster corporate responsibility[5]. In Germany, it is therefore seen as an element of the German Sustainable Development Strategy[6]. And while the specific goals may be different, the Whistleblower Protection Act continues a legislative trend, with the implementation of internal whistleblowing channels already required in some fields of law[7], while in others their introduction is imminent[8]. But these legislative initiatives will not only foster the importance of the whistleblowing process itself but will also make a transparent follow-up process including internal investigations increasingly important. On the market, this did not only lead to a new impetus for the industry of ombudsmen, but also to an almost countless number of web-based whistleblowing tools. Some of which are already reflecting not only the current legal requirements of

[1] Cyber crime in the narrower sense refers to crime that is directed against the internet, other data networks, IT systems or their data, and in a broader sense to crime that is committed with the support of IT systems (BKA Cybercrime Unit, What is Cybercrime?, (Sep. 18, 2023, 8:07 PM) https://www.bka.de/DE/DasBKA/OrganisationAufbau/Fachabteilungen/Cybercrime/cybercrime_node.html)

[2] Gesetz für einen besseren Schutz hinweisgebender Personen sowie zur Umsetzung der Richtlinie zum Schutz von Personen, die Verstöße gegen das Unionsrecht melden, passed May 31, 2023

[3] DIRECTIVE (EU) 2019/1937 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2019 on the protection of persons reporting on breaches of Union law, passed Oct. 23, 2019

[4] While the deadline for implementation ended Dec. 17, 2021, eight member states including Germany were charged by the European Union due to a delayed implementation in their national law.

[5] Explanatory memorandum of DIRECTIVE (EU) 2019/1937 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 23 October 2019 on the protection of persons reporting on breaches of Union law, par. (47), p. 9

[6] Bundestag-Drucksache 20/3442, Entwurf eines Gesetzes für einen besseren Schutz hinweisgebender Personen sowie zur Umsetzung der Richtlinie zum Schutz von Personen, die Verstöße gegen das Unionsrecht melden, p. 38

[7] For example, “Stellen zur Bekämpfung von Fehlverhalten im Gesundheitswesen”, § 197a Abs. 2 Sozialgesetzbuch V

[8] For example, “Beschwerdestelle”, § 8 Lieferkettensorgfaltspflichtengesetz (LkSG)

the Whistleblower Protection Act but also future requirements that will arise from sustainability-related legislation such as the Supply Chain Due Diligence Act[9].

All these whistleblowing tools include a variety of functions that are very useful from a forensic point of view such as an automated and legally compliant documentation of all hints received, the monitoring of deadlines or the possibility of chat communication, both with the whistleblower and with relevant functions within the own organization. But there are also significant differences between the tools available on the market, for example when it comes to the anonymity of the whistleblower. While some forensic practitioner might prefer a non-anonymous reporting channel because it is easier to address any additional questions that might arise during the investigation, quite a few tools rely on anonymous reporting as a default. This corresponds with the understanding of the German lawmaker, as initial drafts of the Whistleblower Protection Act also required the implementation of anonymous reporting channels[10]. Even though this requirement was waived for the final version of the law it clearly emphasizes that anonymity is considered the best safeguard to protect whistleblowers and to ensure participation in whistleblowing systems[11].

From my personal forensic experience, the idea of anonymity being one of the success factors of a whistleblowing system seems worthy of discussion. Whenever I asked a whistleblower out of curiosity why he or she chose to report a case, the most important factor was their trust in the transparency of the investigation as well as trust in the integrity of the persons involved in its execution. So even when opting for an anonymous whistleblowing tool it will still be crucial to get the basic requirements like a fair, unbiased, and transparent investigation process right.

Besides the different handling of anonymity matters, also the technical functionality can considerably vary from one tool to another. While some tools only serve as kind of a digitalized mailbox, others also provide helpful support for the planning, execution, and documentation of forensic follow-up activities. Such comprehensive "investigation management tools" make things a lot easier for the forensic team because all investigations steps can be designed, managed, and archived by the use of just one IT-tool acting as a single source of information. But there are also regulatory aspects to be considered especially when it comes to the processing of personal data on a web-based whistleblowing platform. Organizations based in the EU must comply with the rules of the GDPR when operating such a whistleblowing tool. Therefore and among other requirements, companies will have to carefully consider the physical localization of data. For example, after the Schrems II decision of the European Court a data transfer to the US using standard contractual clauses would only be GDPR-compliant if additional safeguards such as BYK-encryption were applied12. Only now after the implementation of the EU-US Data Privacy Framework (DPF), such a transfer can be based on that newly established adequacy decision[13], which will make cooperation with US-tool providers easier in the future (as long

[9] Lieferkettensorgfaltspflichtengesetz (LkSG) to come into effect as of Jan. 1, 2024

[10] BT-Drucks. 20/5992, Entwurf eines Gesetzes für einen besseren Schutz hinweisgebender Personen sowie zur Umsetzung der Richtlinie zum Schutz von Personen, die Verstöße gegen das Unionsrecht melden vom 14.03.2023, § 16 Abs. 1

[11] BT-Drucks. 20/3442, Entwurf eines Gesetzes für einen besseren Schutz hinweisgebender Personen sowie zur Umsetzung der Richtlinie zum Schutz von Personen, die Verstöße gegen das Unionsrecht melden vom 19.09.2023, p. 66

[12] European Data Protection Board, Empfehlungen 01/2020 zu Maßnahmen zur Ergänzung von Übermittlungstools zur Gewährleistung des unionsrechtlichen Schutzniveaus für personenbezogene Daten, Version 2.0, passed June 18, 2023, p. 38

[13] European Commission, Implementing Decision of 10.7.2023 pursuant to Regulation (EU) 2016/678 of the European Parliament and of the Council on the adequate level of protection of personal data under the EU-US Data Privacy Framework, C(2023)4745 (final)

as the data recipient is DPF-certified[14]). And finally, companies must apply adequate technical and organizational measures to ensure that all data is safely stored. This technical aspect of data protection is of particular importance since the information received via a whistleblowing tool can contain all different kinds of data reaching from health data of employees to most sensitive trade secrets of the company.

In summary, web-based whistleblowing platforms are great tools to make reporting processes more convenient, not only for the whistleblower but also for the forensic team who manages and follows-up on each hint received. On the other hand, organizations are well advised to ensure that the chosen tool reflects not only their functionality needs but also all applicable regulatory requirements.

III. (AI-BASED) IT-TOOLS SUPPORT THE AUTOMATED ANALYSIS OF LARGE AMOUNTS OF DATA DURING AUDITS AND INVESTIGATIONS

But even with web-based whistleblowing tools becoming increasingly common, many investigations will still be triggered by other incidents such as internal audit or annual audit findings. During both kinds of audit, the use of IT-tools to support automated data analysis has a long history that dates way back before the invention of artificial intelligence. As part of a risk-oriented audit approach, it is part of an auditor's daily business to use IT-tools to pre-identify potentially critical transactions based on various criteria such as day and time of transaction, amount, posting texts, name similarities and other manually defined red flags. For example, if the records show a transaction over a round amount that was done on the weekend at 3 a.m. in the morning chances are that an audit software will identify this transaction as "potentially irregular" in a matter of split seconds. Of course, this does not necessarily mean that the transaction was fraudulent, but it gives enough reason to be examined by the audit team in more detail.

But even without such a manual definition of warning indicators, the analysis of large amounts of data has been a standard audit approach for decades. The scientific background for this kind of analysis was initially provided by Simon Newcomb in an article in 1881[15] and re-invented by Frank Benford in his publication in 1938[16]. Both papers dealt with the distribution of leading digits in empirical data and concluded that the higher the leading digit, the rarer it will occur. Honoring Newcomb and Benford, this correlation is now known as Newcomb-Benford's law and assumes the following distribution of leading digits in empirical data sets[17]:

[14] Data Privacy Framework Program, Data Privacy Framework List, (Aug. 9, 2023, 9:02 a.m.) www.dataprivacyframework.gov/s/participant-search

[15] Simon Newcomb, Note on the Frequency of the use of different Digits. In: Natural Numbers, in American Journal of mathematics, 1881, p. 39.

[16] Frank Benford, The Law of Anomolous Numbers. In: Proceedings of the American Philosophical Society 1938, p. 551 et seqq.

[17] Mark J. Nigrini, I´ve got your number. In: Journal of Accountancy, 1999, p. 80

Leading Digit	Likelihood of occurrence in %
1	30,103
2	17,609
3	12,494
4	9,691
5	7,918
6	6,695
7	5,799
8	5,115
9	4,576

There are countless examples for Newcomb-Benford`s law and it can be observed with the size of files stored on a hard drive, the length of rivers[18] or the development of market prices[19] as well as many more.

But there are also more forensic-focused use cases of Newcomb-Benford's law as it also applies to numerous financial figures such as accounts payable data, estimations in the general ledger, customer refunds and others[20]. Important academic work in this regard was done by Mark J. Nigrini who showed in his dissertation at the University of Cincinnati that Newcomb-Benford's law can also be used to detect white collar crime or more specifically tax evasion[21]. With his work, Negrini laid the foundation of a forensic use of Newcomb-Benford's law, which quickly became an effective and proven way for detecting compromised data sets as a starting point for more specific investigation measures.

So, how does it work? M. Dworschak offers the following explanation: "*The 1 is no further away from the 2 on the scale of numbers than the 5 from the 6. However, for the real things that are counted, measured or weighed, the path from 1 to 2 can be very long: in order to cover it, they have to grow twice as much. A 5 on the other hand, is only a fifth short of becoming a 6.*"[22] Let me illustrate this

with the following example: Company A has a turnover of EUR 1,000,000, so the leading digit is 1. To move from leading digit 1 to a leading digit 2, the company must double its current turnover to EUR 2,000,000. Company B has a turnover of EUR 5,000,000, so the leading digit is 5. To get a leading digit 6, company B must also increase its turnover by EUR 1,000,000 in absolute terms, but only by 20 percent in relative terms. In other words, the leading digit 1 is much more resilient to change than higher leading digits. And the more resilient a leading digit is to change, the more often it will occur.

As fascinating the automated search for manually defined warning indicators and general data analysis by applying Newcomb-Benford's law (as well as numerous other automated fraud pattern

[18] Hans Humenberger, Warum 1 so oft vorne steht – das eigenartige Gesetz von Newcomb und Benford, 2011, p. 9

[19] Tarek el Sehity, Erik Hoelzl, Erich Kirchler, Price developments after a nominal shock, Benford's Law and psychological pricing after the Euro introduction, in: International Journal of Research in Marketing, 22, 2005, p. 471 et seqq.

[20] Mark J. Nigrini, I´ve got your number. In: Journal of Accountancy, 1999, p.81

[21] Mark Nigrini, The detection of Income Tax Evasion Through an Analysis of Digital Frequencies, 1992

[22] Manfred Dworschak, Weiter Weg zur Zwei. In: Der Spiegel 47/1998, p. 228

checks[23]) might be for the non-forensic community, it is a common standard approach during forensic investigations and audits. So, what are the additional benefits the forensic community can expect from the use of artificial intelligence during investigations? To answer this, it is necessary to define the term "artificial intelligence" that has become so omnipresent in our lives. Artificial intelligence mimics human cognitive abilities by recognizing and sorting information from input data and can be based on programmed processes or generated by machine learning[24]. In machine learning, an algorithm learns to perform a task independently through repetition. The machine is guided by a predefined quality criterion and the information content of the data. Unlike conventional algorithms, no solution is provided. The computer learns to recognize the structure of the data on its own[25]. This makes AI-supported tools very useful during forensic investigations and allows the analysis of the most extensive and diverse unstructured data while independently identifying relevant analysis parameters. In addition, AI-based open source intelligence (OSINT) tools can support the collection and visualization of relevant data by scraping publicly accessible sources such as websites, social media or public registers. This helps the auditor to identify and better understand the relationship between relevant individuals and/or entities. As an example, let`s assume a kick-back scheme in which the internal perpetrator conspires with an external service provider, he went to university with years ago. While it is quite likely nowadays, that there is some publicly available proof of that relationship (e.g. a connection on social media or a year book on a website), it might be very difficult, if not impossible, to find if one doesn`t know where to look. An OSINT tool may help to find the missing link such as the fact that both individuals founded an offshore entity years ago and served as its legal representatives, both facts well documented in the commercial register of the company's country of residence. And finally, AI-supported forensic tools can also help to avoid compliance cases in the first place by proactively identifying business model-specific fraud risks which can then be addressed by corresponding internal compliance processes and controls.

IV. GOVERNMENTAL USE OF AI IN LAW ENFORCEMENT HELPS TO INCREASE THE EFFECTIVENESS OF POLICING BUT CAN ALSO LEAD TO SIGNIFICANT RISKS

But the use of AI-supported tools is not limited to private organizations. Law enforcement agencies are also increasingly relying on AI-supported tools. For example, as part of a predictive policing approach agencies may use AI to map crime hot spots or predict the risk of a person becoming involved in a violent or serious crime etc.[26]. Another well known (and well debated) IT-use in the context of law enforcement is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), a tool that supports courts in assessing the risk that an offender will commit another crime after being released from custody by taking numerous factors into consideration. But the use of AI-based decision aids for law enforcement does not come without risk and some studies suggest a bias (e.g. regarding

[23] Some IT forensic service providers include more than 170 different pattern checks in their data analysis, for example see IT Compliance Systeme GmbH, Forensic Data Mining/Red Flag Analysis/Journal Entry Testing, (Sep. 29, 2023, 9:47 p.m.); www.compliance-systeme.de/en/service/it-forensics/forensic-data-mining-red-flag-analysis

[24] Fraunhofer-Institut für Kognitive Systeme, *Künstliche Intelligenz (KI) und maschinelles Lernen*, (Sep. 17, 2023, 10:14 p.m.), https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html

[25] Fraunhofer-Institut für Kognitive Systeme, *Künstliche Intelligenz (KI) und maschinelles Lernen*, (Sep. 17, 2023, 10:14 p.m.), https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html

[26] Douglas Yeung, Inez Khan, Nidhi Kalra, Osonde Osob, *Identifying Systematic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Agencies*, 2021, p. 2

the race of the person concerned) in their results[27]. But while the question whether such tools in general or a specific tool in question actually show bias or not goes way beyond the scope of this whitepaper, it is important to understand that the mere possibility of bias in this kind of tools imposes significant risks: A risk for the law enforcement agencies as they rely on the results to plan and execute their tasks or to allocate their resources. A risk to the individuals concerned who have a constitutional right to be treated fair and transparently by public authorities. And finally, a risk to policing itself: Procedural justice and its ingredients such as perceptions of neutrality, the treatment of people with dignity and respect, and the trustworthiness of law enforcement's motives are crucial to the legitimacy of law enforcement[28]. And if the people perceive law enforcement as having legitimate authority, they are more likely to follow the law and cooperate with the police[29].

But criticism is not only engendered by government-operated AI- tools. In January 2023 for example, Meta, the operator of social media platforms such as Facebook and Instagram, sued a private UK-based forensic tool provider for data scraping from 600,000 Meta users with 38,000 fake profiles[30]. One of their clients that allegedly spent millions of dollars for these services was the Los Angeles Police Department (LAPD)[31]. That indicates that law enforcement does not only use tools that are specifically designed for agency use but also intensively cooperates with private forensic service providers. It is unclear how agencies ensure that these publicly available private tools also comply with the procedural justice requirements for law enforcement use as described above.

V. CONCLUSION

The trend towards digitalization and the use of artificial intelligence will undeniably provide new impetus and opportunities not only for the private sector but also for law enforcement. However, there are potential (legal) risks especially in the field of data protection and data security. Where a use is legally permissible, private organizations must carefully weigh up the chances and risks for the individuals affected, not only from a legal but also from an ethical point of view. Special caution is required for the use of AI-based tools by law enforcement. In this regard, it will be crucial to critically assess planned tools, their use cases, and their results for potential bias. The prevention of biased results is a key requirement for the public acceptance of AI-based decision aids in policing as well in other fields of law enforcement.

[27] For a general overview including further reference see: Douglas Yeung, Inez Khan, Nidhi Kalra, Osonde Osob, *Identifying Systematic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Agencies*, 2021

[28] Douglas Yeung, Inez Khan, Nidhi Kalra, Osonde Osob, *Identifying Systematic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Agencies*, 2021, p. 2; Lorraine Mazerolle, Sarah Bennett, Jacqueline Davis, Elise Sargeant, Matthew Manning, *Procedural Justice and Police Legitimacy: A Systematic Review of the Research Evidence*. In: Journal of Experimental Criminology, Vol. 9, No. 3, 2013, p. 246

[29] Douglas Yeung, Inez Khan, Nidhi Kalra, Osonde Osob, *Identifying Systematic Bias in the Acquisition of Machine Learning Decision Aids for Law Enforcement Agencies*, 2021, p. 2; Lorraine Mazerolle, Elise Sargeant, Adrian Cherney, Sarah Bennett, Christina Murphy, Emma Antropus, Peter Martin, *Procedural Justice and Legitimacy in Policing*, 2014, p. 10

[30] Der Spiegel, *Meta verklagt Überwachungsfirma wegen Fake Accounts*, Jan 13, 2023, 2:47 p.m. https://www.spiegel.de/netzwelt/netzpolitik/meta-verklagt-ueberwachungsfirma-wegen-fake-accounts-a-b51bfedc-fe05-43d1-8a76-a1d604a47dcf

[31] Johana Bhuiyan, *NYPD spent millions to contract with firm banned by Meta for fake profiles*, The Guardian, Sep. 8, 2023, https://www.theguardian.com/us-news/2023/sep/08/new-york-police-tracking-voyager-labs-meta-contract