Navigating the Risks of the use of Digital Reporting and E-Invoicing in the Age of Advanced Fraud Detection

Professor Madeleine Merkx and Breyli Luna Calzado LL.M.[1]

Abstract:
With the EU set to implement mandatory e-invoicing and digital reporting by 2030, automated fraud detection tools are expected to become standard in VAT enforcement. This article seeks to address the potential risks of automatic fraud detection tools implemented by EU Member States or at the EU level. The article examines these risks and identifies four key concerns: the risk of treating algorithmic predictions as evidence of fraud; limitations in data quality and model accuracy; the potential for bias, and function creep; and the lack of legal protection. These risks could have negative consequences for businesses. For this reason, the article argues that such risks should be addressed through well-founded decisions about the choice of analytical methods, based on the purpose of the analysis, along with measures to prevent, correct, or compensate for irregularities and discriminatory outcomes. Only sound and scientific statistical research should be used, and systems should be regularly validated and audited. Transparency in the operation of fraud detection tools is also crucial. Most importantly, fraud detection systems should support human decision-making, not replace it.

Key words: E-invoicing, digital Reporting, fraud detection system, ViDA, risk assessment, machine Learning, VAT Fraud, VAT Gap, enforcement, legal protection

1. Introduction

VAT fraud, and in particular carousel or missing trader intracommunity (MTIC) fraud is a persistent problem in the EU, demonstrated by the yearly VAT gap reports of the European Commission.[2] It is regarded as one of the most widespread types of tax fraud in EU Member States.[3] VAT fraud has serious social and economic consequences, because it can e.g. lead to inadequate resources for redistribution of income and financing of public services on EU Member State level.[4] What's more, money obtained with VAT fraud is used to finance (organized) crime[5] and even terrorist organizations.[6]

After a failed attempt to remedy VAT fraud with the definitive VAT system[7] the European Commission proposed mandatory e-invoicing and digital reporting in December 2022. With this proposal the European Commission seeks to provide EU Member States with (near) real time data on cross border transactions on a transaction-by-transaction basis.[8] With this (near) real time information it should be easier to detect and investigate VAT fraud cases, subsequently to prosecute those committing VAT fraud and try to retrieve the

[1] Madeleine Merkx is a professor of indirect taxes at the department Law & Tax of Erasmus School of Law Erasmus University Rotterdam and a partner at the tax research center of BDO Accountancy, Tax & Legal B.V. in Tilburg. Breyli Luna Calzado is an Indirect Tax Specialist at the Omron Group.

[2] See the recent 2024 report: European Commission, 'VAT Gap in the EU. Report 2024', p. 26 where it is stated that in 2022 the VAT compliance gap was estimated at EUR 89.3 billion in the whole EU. It should be noted that the compliance gap does not only encompass VAT fraud. It also covers VAT lost due to, for example, insolvencies, bankruptcies, administrative errors, and legal tax optimization (see p. 174 of the report). For a critical review of the VAT gap estimation see: Lisette van der Hel-van Dijk and Menno Griffioen, 'Intra-Community VAT Fraud: Is It Really a 50 Billion Euro Problem?', International VAT Monitor, 2025 (Volume 36), No. 1, https://doi.org/10.59403/2t6m5a6.

[3] Umut Turksen, 'Countering Tax Crime in the European Union', Hart Publishing 2023, p. 63.

[4] Umut Turksen, supra. 3.

[5] EPPO, 'MTIC (Missing Trader Intra Community) fraud', https://www.europol.europa.eu/crime-areas/economic-crime/mtic-missing-trader-intra-community-fraud, last consulted on 25 October 2024.

6 Marius-Cristian Frunza, 'The use of tax frauds - including VAT or carrousselfraud - to finance terrorism - questions, public Hearing European Parliament, p. 1-7.

[7] Proposal for a Council Directive amending Directive 2006/112/EC as regards the introduction of the detailed technical measures for the operation of the definitive VAT system for the taxation of trade between Member States, Brussels 25.5.2018, COM (2018) 329 final.

[8] Proposal for a Council Directive amending Directive 2006/112/EC as regards VAT rules for the digital age, Brussels 8.12.2022, COM (2022) 701 final, p. 4.

money.[9] This will lead to an estimated reduction of VAT fraud by up to EUR 11 billion a year according to the European Commission.[10] The measures can also have a deterrent effect, since it is a well-known fact that those that know they are being observed are more likely to comply[11] or simply move their fraudulent activities elsewhere. Member States reached a political agreement on this proposal on 5 November 2024 and it was formally adopted on 11 March 2025 with the adopted texts published in the official journal on 25 March 2025.[12]

This article does not seek to address the effectiveness of the measure proposed but deals with the potential risk of automatic fraud detection tools implemented by EU Member States or at EU-level which will likely accompany the adoption of measures like digital reporting and e-invoicing, if these tools are not already in place.[13] Although honest businesses may be expected to contribute to the fight against VAT fraud (as part of their societal obligations) and are willing to do so (since fraudulent transactions cause unfair competition in their sector), actions taken as a result of automatic fraud detection should not have disproportionate negative consequences, as those consequences may threaten honest businesses' existence. The proposed research question is therefore: What risks do advanced fraud detection based on e-invoicing and digital reporting entail and how can these risks be mitigated?

This question will be answered by first discussing the obligations for e-invoicing and digital reporting applicable as of 1 July 2030 as well as an analysis of similar obligations currently already in place in EU Member States (section 2). As a second step we will analyse the use of fraud detection systems (section 3) and the risks they entail (section 4). In section 5 the risks of automatic fraud detection in relation to e-invoicing and digital reporting will be addressed and recommendations are proposed. Section 6 contains a conclusion. It should be noted that the extent to which obligations to issue e-invoices, digital reporting and the use of the data obtained through this process interfere with the right to privacy is beyond the scope of this article, as it is a research topic in its own right.

2. E-invoicing and digital reporting

2.1 Introduction

VAT represents a major source of revenue for both the EU and its Member States. For the European Union budget, VAT is a key source of financing, with 0.3% of the VAT collected at national level being transferred to the EU as its own resources, representing 12% of the overall EU budget.[14] Considering the significant role that VAT plays in the budgetary decision-making, the European Commission has announced in 2020, the introduction of its Action Plan, aiming to develop a legislative package that is fairer, more comprehensible, and harmonized. This package will feature updated tax rules capable of adjusting to the new digital economy.[15] The current VAT Directive, created three decades ago, is outdated for today's digital age. It obstructs digital progress by requiring EU Member States to obtain a derogation from the VAT Directive to adopt digital reporting requirements (DRRs) based on obligatory e-invoicing requirements. Consequently,

[9] One of the features of a VAT carrousel fraud is that it happens very quickly. Traders (so called missing traders) disappear within a few months. This makes the detection of fraud and the recovery of the lost VAT income very difficult, because the tax authorities currently rely on monthly or quarterly information, see Questions and Answers: VAT in the Digital Age, Brussels 8 December 2022.

[10] Questions and answers VAT in the digital age, supra. 9, p.

[11] OECD (2017) The Changing Tax Compliance Environment and the Role of Audit, p. 245.

[12] Council Directive (EU) 2025/516 of 11 March 2025 amending Directive 2006/112/EC as regards VAT rules for the digital age OJ L, 2025/516, Council Implementing Regulation (EU) 2025/518 of 11 March 2025 amending Implementing Regulation (EU) No 282/2011 as regards information requirements for certain VAT schemes OJ L, 2025/518 and Council Regulation (EU) 2025/517 of 11 March 2025 amending Regulation (EU) No 904/2010 as regards the VAT administrative cooperation arrangements needed for the digital age OJ L, 2025/517

[13] See e.g. https://taxadmin.ai/ a website about the use of AI by tax administrations in the EU and OECD (2016) Advanced Analytics for Better Tax Administration: Putting Data to Work, OECD Publishing Paris.

[14] Article 2 of the Council Decision (EU, Euratom) 2020/2053 of 14 December 2020 on the system of own resources of the European Union, OJ L 424, 15.12.2020, p. 1–10.

[15] Communication from the Commission to the European Parliament and The Council, an action plan for fair and simple taxation supporting the recovery strategy, Brussels 15.07.2020, COM (2020) 312 final, p.1-3.

the unpreparedness of the system has led to challenges in adapting to the digital economy, facilitating tax fraud.[16]

While digitalization has contributed to the problem, it also represents a valuable tool that can be used to detect and prevent tax fraud. According to an OECD report, these solutions can offer a win-win: better detection of crime, higher revenue recovery, and synergies that can make tax compliance easier for businesses and tax administrations.[17] The European Commission, in its move towards a more efficient VAT System, released on 8 December 2022, a legislative package proposal, colloquially known as ViDA - VAT in the Digital Age.[18] On 5 November 2024 the EU Member States reached a political agreement on this legislative package with a formal adoption of the package on 11 March 2025. The texts on which they reached an agreement introduced new implementation dates and significant amendments to certain areas of the initial proposal were made.[19]

In the upcoming section (2.2), we will examine more closely one of the key proposed changes, which involves VAT reporting through the introduction of digital reporting requirements (DRRs) based on electronic invoicing. The implementation of this change has been delayed, moving from a planned 2028 start to a revised timeline of 1 July 2030. As indicated earlier, DRR has been in place for quite some time. Diverse forms of DRRs have been implemented across multiple Member States. In section 2.3, we will assess the various forms of DRR, highlighting the preferences of Member States among them.

2.2 Digital reporting and e-invoicing under ViDA

2.2.1 Combatting VAT fraud

The rules under ViDA aim to combat VAT fraud and to reduce the EU Member State's VAT gap. Following the establishment of the EU Single Market in 1993, fiscal controls at internal borders were no longer carried out between the European countries in the Schengen area. Primarily to promote the smooth functioning of the internal market and foster economic growth. However, this borderless market inadvertently created an opportunity for fraudulent activities among international traders, such as MTIC fraud.[20] In order to understand MTIC, it is crucial to understand how cross-border supplies of goods are treated in the current VAT system. The EU VAT System is designed to split a cross-border taxable event into two: the supplier's exempted/zero rated (with the right of deduction) intra-Community supply of goods from the Member State of dispatch, and the acquirer's intra-Community acquisition of goods in the Member State of arrival.[21] As a way of example, if a business established in the Netherlands (Company A) sells a goods to a company established in Portugal (Company B), that supply will be zero rated/exempt with a right to deduct. Company B will report the acquisition in the country of arrival against the local VAT rate of 23% (which VAT is deductible in the same VAT return) and applies 23% Portuguese VAT when selling the goods domestically to a company established in Portugal (Company C). Fraud occurs when the VAT on the domestic supply (the supply by B to C) is not passed on to the national tax authorities. A basic illustration of MTIC fraud unfolds when company B in Portugal would disappear without accounting for the VAT in Portugal. The VAT exemption for the Intra-Community supply from country A to country B enables this type of fraud by eliminating any irrecoverable VAT for the fraudulent trader. In this type of fraud goods can be supplied by party C back to party A and the process can take place again and again. It is therefore also called carousel fraud.

---

[16] Proposal for a Council Directive supra. 8, p. 3.
[17] OECD (2017), Technology Tools to Tackle Tax Evasion and Tax Fraud. Paris: OECD Publishing.
[18] Proposal for a Council directive, supra 8, Proposal for a Council Implementing regulation amending Implementing Regulation (EU) No. 282/2011 as regards information requirements for certain VAT schemes, COM (2022) 704 final and Proposal for a Council Regulation amending Regulation (EU) No 904/2010 as regards the VAT administrative cooperation arrangements needed for the digital age, COM (2022) 703 final.
[19] VAT in the Digital Age legislation, supra 12.
[20] European Parliament, Missing Trader Intra-Community Fraud (2021), https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690462/IPOL_BRI(2021)690462_EN.pdf (accessed 28.03.2024).
[21] CJEU 27 October 2022, Case C-641/21 (Climate Corporation Emissions Trading GmbH v Finanzamt Österreich), ECLI:EU:C:2022:842, paras. 46- 47.

To prevent detection, fraudsters make use of honest traders. Multiple fraudulent chains can also take place at the same time, creating a different pattern from the traditional supply chain explained above[22] with a complex chains of companies.[23] It should be noted that fraudsters try to make it look like invoicing and tax declarations are in order to prevent discovery.[24] Fraud can also take place with services. A well-known fraudulent scheme is that with greenhouse gas emissions allowances. Supplies of greenhouse gas emission allowances are supplies of services subject to VAT in the Member State of the recipient of the supply in case of B2B supplies, where the VAT is reverse charged to the customer if the supplier is not established in the recipient's Member State. Again, the fraudster can therefore purchase the services without VAT, pocketing the VAT that he charges to its customer on domestic sales. In case of services cross-border transport is unnecessary. The fact that greenhouse gas emission allowances are exchanged via a stock exchange also makes fraud easy. This type of fraud can be addressed by EU Member States through applying a reverse charge mechanism on domestic supplies under art. 199a (1) (a) VAT Directive.

Intra-community supplies and services on which VAT is reverse charged under the current art. 196 VAT Directive must currently be reported periodically in recapitulative statements or EC Sales listing. Recapitulative statements have, however, revealed their inadequacy in mitigating cross-border fraud.[25] First, in terms of data content and format, the data sent through the recapitulative statement lacks granularity, since the information is aggregated and is not provided on a transaction-by-transaction basis i.e. detailed information on each transaction (including taxable amount, VAT rate, and VAT due). Second, in terms of frequency, the data provided is delayed. In accordance with Article 263 of the VAT Directive, the recapitulative statement should be filed every month. In practice that means that the cross-border transactions are reported to the tax authorities around a month after the transaction has occurred through the recapitulative statements. Such a wide gap creates a window for fraudsters to disappear without having paid the VAT to tax authorities. In addition, data may be available to tax authorities in other Member States too late, not only because of the monthly filing frequency, but also because of the time it takes for local tax authorities to exchange that information. Third, in terms of the data scope, there is a partial scope limitation, which only covers intra-Community supplies. The intra-Community acquisition data is not automatically exchanged.[26]

The introduction of an EU DRR for intra-Community supplies along with the option for Member States to adopt similar DRR systems for domestic transactions, could mark a significant shift toward a more efficient and transparent approach to tax reporting. This initiative offers an automated method for cross-checking transactions on a detailed, almost real-time basis, applicable to both intra-community supplies and, if chosen, domestic transactions.

The efficacy of the DRR is not just theoretical; its practical implementation in certain Member States has proven its ability to improve processes by rectifying existing inefficiencies. This successful application reinforces the DRR's capacity to significantly improve the audit system.

2.2.2 Mandatory e-invoicing and digital reporting requirements

ViDA aims to harmonize the e-invoicing landscape within the EU by adopting near real-time digital reporting requirements (DRR) based on e-invoicing. The rules indicate that all businesses, including non-resident businesses, will be required to issue e-invoices via an EU standard format for B2B intra-Community supplies of goods, services on which the VAT is reverse charged to the customer and transfer of own goods,[27] within

[22] R.A. Wolf, 'Carrouselfraude'(Carousel fraud), Den Haag: Sdu Uitgevers 2010, section 3.2.
[23] EPPO, 'Operation Admiral: EPPO uncovers organized crime groups responsible for VAT fraud estimated at € 2.2 billion', 29 November 2022, https://www.eppo.europa.eu/en/news/operation-admiral-eppo-uncovers-organised-crime-groups-responsible-vat-fraud-estimated-eu22#:~:text=It%20took%20the%20EPPO%20less,VAT%20fraud%20discovered%20so%20far.%22 (last consulted on: 17 December 2024).
[24] EPPO, supra 23.
[25] See European Commission, 'VAT in the Digital Age, final report. Volume 1, Digital reporting requirements, p. 85
[26] Commission Staff Working Document Impact Assessment Report, Brussels 8.12.2022, SWD (2022) 393 final.
[27] Unless a taxable person registers for the new special scheme, under the new Article 369xa.

ten days of the taxable event.[28] E-invoicing will be the standard method for issuing invoices. Next to this certain data on the invoice[29] needs to be reported by both the supplier and the customer. The rules further indicate that data can be transmitted either directly by the taxable person, through a third party, or via a public portal if one is accessible.[30]

With the implementation of the DRR, comes the removal of the recapitulative statement, which is the current reporting system of intra-Community transactions. Currently, the data included in these recapitulative statements are maintained in national VAT databases, which are interconnected through an electronic interface known as the VAT Information Exchange System (VIES).[31] The Commission supports inter-Member State communication via VIES, while individual Member States are tasked with the development of their own national applications. Under ViDA a central electronic system for the obligatory exchange of VAT information will be established, known as 'Central VIES,' starting from 1 July 2030.[32] Once information from the taxable person is collected by the competent authorities, it should be recorded in the central VIES within one day. [33] It would be the responsibility of each Member State to develop, maintain, host and technically manage a national electronic system to automatically transmit different categories of information to the central VIES.[34]

For domestic transactions, Member States will have the option to implement a digital reporting system which should be in line with the EU model. Member States that have already implemented digital reporting requirements as of January 1, 2024, or have been granted authorization by Article 395 of the VAT Directive to do so on 1 January 2024 (or where no such authorization is required have adopted national legislation to that end), must ensure their systems align with that used for cross-border supplies of goods and services by 2035.[35]

2.3 EU Member States' experience with digital reporting systems

As outlined in section 2.2, ViDA introduces a harmonized EU-DRR system, utilizing e-invoicing for B2B Intra-Community transactions, while leaving the implementation for domestic transactions at the Member States' discretion. DRRs are already introduced within the EU, as several Member states have implemented these systems at national level. The main objectives of these systems is to combat fraud, under-collection and error.[36] While tax authorities have noted improvements in tax control measures, particularly in terms of more accurate and effective risk analysis like identifying suspicious taxpayers and transaction chains, [37] it's important to recognize that DRR alone cannot tackle all tax compliance risks or prevent fraud, especially not EU-wide fraud.[38] This section will explore the various types of DRR that have

[28] New art. 222 first paragraph VAT Directive
[29] The data to be transmitted differs per transaction and can be found in the new art. 264 VAT Directive
[30] New art. 263 (3) VAT Directive.
[31] Explanatory memorandum of the proposal for a Council Regulation amending Regulation (EU) No 904/2010 as regards the VAT administrative cooperation arrangements needed for the digital age, Brussels 8.12.2022, COM (2022) 703 final, p. 3.
[32] Explanatory memorandum, supra 31, p. 8.
[33] Explanatory memorandum supra 31, p. 7.
[34] Explanatory memorandum supra 31, p. 6.
[35] Art. 5 of the draft council directive dealing with the transposition.
[36] Expert Group, A Next Generation Model for Electronic Tax Reporting and Invoicing', 3 August 2022, p. 5.
[37] See European Commission, supra 25, p. 15 where it is stated that the increase in VAT revenue during the 2014-2019 period is estimated to be between EUR 19 and EUR 28 billion in the Member States which have introduced DRRs in this period, corresponding to an annual increase of VAT revenue of between 2.6% and 3.5%.
[38] This was also emphasized by L. van der Hel- van Dijk and M. Griffioen Digitale real-time rapportage als oplossing voor intracommunautaire btw-fraude? (Digital real time reporting as a solution to VAT fraud), WFR 2023/297 . where they stated that DRR alone will not be sufficient and is not the ultimate solution. See European Commission, supra 25, , p. 76 where it is stated domestic Digital Reporting Requirements (DRRs) are not proving to be an effective measure for combating intra-EU VAT frauds. Given that intra-EU VAT transactions account for approximately 40% of the total VAT, the resulting VAT gap is substantial. This ineffectiveness stems from two main issues: firstly, many domestic DRRs do not include intra-EU transactions in their scope; secondly, even when such transactions are covered, the data is not automatically shared with foreign tax authorities, preventing any automated risk analysis or identification of discrepancies.

been adopted at the national level across the EU, namely the Periodic Transaction Controls (PTCs) and the Continuous Transaction Control (CTCs).[39]

As indicated by their names, both systems can be distinguished based on the time at which information is to be submitted. The PTCs require the data to be reported periodically to the tax authorities, whereas CTCs require the data to be reported electronically to the tax authorities either before, during or immediately after the supplier sends an invoice to the buyer. Within the PTCs the most common models are the VAT listing and the Standard Audit File for Tax (SAF-T). Countries like Bulgaria, Croatia, Czech Republic, Estonia, Latvia and Slovak Republic have implemented the VAT listing. Meanwhile, Lithuania, Poland and Portugal implemented the SAF-T model, which is based on the OECD template.[40] Notably under the new text of art. 273 VAT Directive Member States can keep or implement PTCs in addition to the harmonized Digital reporting at EU-level (see section 2.2.2)

Within the CTCs, the two possibilities are real-time and e-invoicing. In the real-time system, taxpayers are required to submit specific data shortly after completing a transaction, although it is not mandatory for them to use and share e-invoices with the tax administration. Spain and Hungary have both implemented real-time systems, yet each country's system differs significantly from the other. In 2017, Spain became the first EU Member State to introduce a real-time obligation with the SII (*Suministro Inmediato de Información del IVA)*. This system functions as a bookkeeping mechanism directly managed within the electronic platform of the tax authority. To report the invoices in the VAT registers, taxpayers are required to transmit invoicing details to the tax authority for both issued and received invoices within four working days following the date of the invoice issuance.

Unlike the real-time system, where invoices are transmitted directly from supplier to customer, with subsequent reporting of the invoice data to the tax authorities, e-invoicing requires all taxpayers to issue invoices using a structured e-invoice. The term structured indicates that the e-invoice must conform to a machine-readable standard, so that it can be automatically processed. Subsequently, these e-invoices must be transmitted to tax authorities either before or immediately after issuance. One possibility is where the taxpayer sends the e-invoice directly to their customers while also sharing it with the tax authority, referred to as no-clearance e-invoicing. Alternatively, the taxpayer may be required to go through the tax authority first, either to obtain preliminary authorization or by utilizing a centralized IT platform. This platform then delivers the e-invoice to the customer, this is referred to as clearance e-invoicing. Currently, Italy stands as the first Member State within the EU to obtain the derogation ex art. 395 of the VAT Directive for implementing a mandatory e-invoicing with clearance, known as the Sistema di Interscambio.[41]

In this centralized clearance system, the Italian tax authorities centrally process invoices before they are sent to recipients. When a taxpayer creates an invoice, it first goes through the tax authorities for approval. Before approval, several checks and controls are carried out, including verifying the file's uniqueness to avoid duplicates, confirming the authenticity of the signature certificate, and assessing the content of the invoice, including the validity of tax codes and VAT numbers.[42] Once approved, it's forwarded to the customer. Additionally, as part of the e-invoicing mandate, all companies are required to use the Italian e-invoicing standard FatturaPA.

3. Fraud detection systems

The data collected and shared under ViDA will be analysed both by the central VIES and Member States individually. Likely the data will be used in fraud detection systems since one of the main objectives of digital reporting through e-invoicing is the detection of fraud. Fraud detection can be defined as the

[39] See European Commission, supra 25, p. 15
[40] See: SAF-T_Schema_v_2.00 (oecd.org) (accessed 9.05.2024).
[41] eInvoicing in Italy (europa.eu) (accessed 5.04.2025)
[42] The more detailed list can be found at the website: Electronically Invoicing the Public Administration - Exchange System documentation (fatturapa.gov.it) (accessed 5.04.2025).

automated process of identifying high-risk instances.[43] Recognizing the transformative potential of this approach, governments and companies have shifted from traditional auditing to implementing advanced selection tools such as nowadays, artificial intelligence (AI), advanced algorithms, and modern big data analytics to combat fraud. These fraud detection systems offer significant benefits, as they enable the government to exchange data, link files, and analyze information through data analysis, thereby enhancing the effectiveness and efficiency of oversight for the tax authorities.[44] The linking of the data files helps in selecting which cases to audit.[45] This is important because choosing which taxpayers to audit has always been a key issue for tax administrators. Focusing on high-risk cases by selecting a small number of taxpayers with a high probability of tax fraud has several expected benefits for the tax audit process: reduced audit costs, as authorities can focus their efforts on high-risk cases rather than randomly auditing a large number of taxpayers; increased taxpayer awareness of regulatory compliance, since taxpayers will be more diligent in complying with tax regulations knowing that audits are more likely for those suspected of fraud; and a reduction in the heavy burden of audit defense for the majority of taxpayers, as fewer people will be subjected to audits, thus alleviating the burden on those who are compliant.[46] However, the use of these fraud detection systems also come with associated risks, which will be discussed in section 4. This section will explore the different types of fraud detection systems currently in use, assessing their advantages and limitations while underscoring the benefits of combining them. Additionally, it will explore how these techniques have enhanced the detection of VAT fraud or tax evasion by tax administration.

Fraud detection systems have evolved over time, shifting from traditional methods, such as the use of performing random audits, to more modern technology driven fraud prevention system. As time progressedmethodologies were developed that utilized statistical analysis and the construction of financial or tax ratios, eventually leading to the creation of rule-based systems, structured as ¨If X and Y, then Z¨.[47] This rule-based engine can be defined as a system that checks predefined rules and gives alerts when certain conditions are met.[48] Operating on a straightforward 'if-then' principle, it takes action when certain criteria are met. One of the main limitations of the rule-based engine is that, while they are relatively simple to implement and are effective against known types of fraud, they struggle in case of complex or emerging patterns. Traditional systems such as rule-based become inaccurate over time as fraudster behavior changes.[49] To tackle this, these systems require frequent pattern updates, resulting in high maintenance costs.[50]

Another point of attention is that traditional methods also rely on manual auditing and inspection by tax investigators and auditors, which have proven to be time-consuming, resource-intensive, costly and inefficient in detecting fraudulent activity in a timely manner, conserving that the point is to detect fraudulent activity before any amount is refunded. Given the time sensitivity nature of this matter, focus has shifted to modern fraud detection methods like data mining and machine learning, which allow tax authorities to enhance operational efficiency and detect correlations in or across datasets that may be undetectable to human cognition.[51] The use of these techniques has proven to be more efficient in

[43] Van Vlasselaer et al.: GOTCHA! Network-Based Fraud Detection for Social Security Fraud 3094 Management Science, 2017, vol. 63, no. 9, pp. 3090, © 2016 INFORMS
[44] T.M. Berkhout, G.A. Raven en M. van Engers, 'Effect van wet- en regelgeving omtrent algoritmen op fiscale boekenonderzoeken' (The effect of legislation regarding algorithms on tax audits), MBB 2024/30, p. 6
[45] T.M. Berkhout, G.A. Raven en M. van Engers, supra 44, p. 6
[46] Changro Lee, 'Deep learning-based detection of tax frauds: an application of property acquisition tax', Data Technologies and Applications Vol. 56 No 3, 2022, p. 338.
[47] Pankaj Gupta, Leveraging Machine Learning and Artificial Intelligence for Fraud Prevention SSRG International Journal of Computer Science and Engineering, Volume 10 Issue 5, 47-52, May 2023, ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V10I5P107
[48] Improving a Rule-based Fraud Detection System with Classification Based on **Association** Rule Mining. Available from: https://www.researchgate.net/publication/349244021_Improving_a_Rule-based_Fraud_Detection_System_with_Classification_Based_on_Association_Rule_Mining [accessed 4.06.2024].
[49] Anna Nesvijevskaia, Sophie Ouillade, Pauline Guilmin and Jean-Daniel Zucker, The accuracy versus interpretability trade-off in fraud detection model Data & Policy (2021), 3: e12 doi:10.1017/dap.2021.3, p. 3-12-3-15. Also OECD (2020), Tax Administration 3.0: The Digital Transformation of Tax Administration, OECD, Paris. p.7
[50] Sundong Kim et. al., 'Active Learning for Human-in-the-Loop Customs Inspection' IEEE Transactions on Knowledge and Data Engineering PP(99), doi:10.1109/TKDE.2022.3144299, p. 2.
[51] M. Finck, 'Reasoned A(I)dministration: explanation requirements in EU law and the automation of public administration', European Law Review 47 (3), p. 378

detecting patterns of fraud or evasion more quickly and accurately.

The modern technology systems are driven by machine learning and artificial intelligence. Although machine learning and artificial intelligence are often mentioned together, they are not synonymous. AI can be defined as the research field that aims at performing machine learning to obtain an intelligent machine that can perform tasks on behalf of the user.[52] Machine learning is a subfield of artificial intelligence that studies the ability to improve performance based on experience.[53] In recent years, numerous systematic review papers have been published that explore the application of machine learning in fraud detection. From a broad perspective, current approaches can be categorized into unsupervised, supervised, and active learning techniques, which combine elements of both supervised and unsupervised methods to capitalize on their strengths and minimize their weaknesses.[54] In straightforward terms, supervised techniques depend on a labeled training dataset to create a fraud detection model, whereas unsupervised approaches do not require labeled data.[55] A more detailed discussion of these techniques will be provided in the upcoming paragraphs.

Supervised machine learning relies on labeled data to learn. The category is known as "supervised" because the system receives guidance from a "teacher," who provides the correct expected outputs, enabling the algorithms to learn the relationship between inputs and outputs.[56] These supervised models are developed using data collected from prior audits or similar interventions. These models are trained to identify patterns that most accurately predict the correct outcome: determining whether a case was non-compliant.[57] Common methods in this class include logistic regression, decision trees, support vector machines (SVM), Bayesian networks, and neural networks.[58] Overall, the main advantage of supervised models is their ability to decrease the number of cases mistakenly marked for intervention, thereby saving tax auditors time. However, this also limits them, as they are generally ineffective in identifying new or previously unknown types of risk.[59] Consequently, unsupervised techniques are crucial for identifying novel anomalous patterns.[60] In this approach, the algorithm is provided only with input data and is tasked with extracting meaningful information from it.[61] These models are trained to detect unusual patterns in the data, rather than learning from the outcomes of specific cases.[62] Unsupervised methods are particularly useful in situations where labeling is costly, such as when multiple analysts are required to review a large volume of data points.[63] Common techniques in this class include clustering analysis, association rules, self-organizing maps, fuzzy rules, time series analysis and unsupervised anomaly detection methods. In the context of fraud detection, anomaly detection operates on the assumption that fraudulent behavior is both rare and distinct from the compliant majority behaviors.[64] Anomaly detection aims to identify entities that display a conduct that deviates from the common behavior within the norm.[65]

However, studies have shown that for anomaly detection and classification tasks, supervised methods that rely on manually reviewed transactions (labeled data), often outperform unsupervised learning because it directly leverages known examples.[66] As mentioned earlier, supervised methods struggle to identify new

[52] Hala Z Alenzi and Nojood O Aljehane, Fraud Detection in Credit Cards using Logistic Regression International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 2020 , p. 541
[53] S. Russell, and P. Norvig, Artificial Intelligence: A Modern Approach (4th ed.). Pearson Education 2021, p. 1
[54] D. Labança L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, "Amaretto: An Active Learning Framework for Money Laundering Detection." *IEEE Access*, vol. 10, 2022, p 41722
[55] Jellis Vanhoeyveld, David Martens and Bruno Peeters, Customs fraud detection : assessing the value of behavioral and high-cardinality data under the imbalanced learning issue Pattern analysis and applications - ISSN 1433-7541 - 23(2020), p. 2
[56] Andreas C Müller, and Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly Media, 2016, p. 2
[57] See e.g. supra 13 and OECD (2016) Advanced Analytics for Better Tax Administration: Putting Data to Work, OECD Publishing Paris, p.23
[58] Changro Lee, supra 46,, p. 330.
[59] See supra 13 and OECD supra 57, p.23
[60] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero supra 54, p. 41722
[61] Andreas C. Müller & Sarah Guido, supra 56 p. 131.
[62] See supra 13 and OECD supra 57, p.23
[63] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, supra. 54, p 41722
[64] Jellis Vanhoeyveld, David Martens, Bruno Peeters supra 55, p. 3
[65] Jellis Vanhoeyveld, David Martens, Bruno Peeters, supra 55, p. 3
[66] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, supra 54, p 41723

fraudulent patterns, which can lead to fraud remaining undetected. By contrast, unsupervised learning is better suited for detecting novel anomalous patterns. To address the limitations of supervised learning, active learning plays an important role in bridging the gap between supervised and unsupervised methods, enabling models to learn from both labeled and unlabeled data.[67] An experimental evaluation was conducted using Amaretto, an active learning tool, to detect money laundering in capital markets. Amaretto integrates both supervised and unsupervised approaches. It uses an unsupervised model to detect both known and unknown anomalous patterns and then feeds the data into a supervised learning model to continuously enhance the system's performance.[68] The research demonstrated that this active learning method was able to process over 29 million transactions, extracting aggregated features and highlighting customer behavioral patterns over time to detect unusual correlations.[69]

Building on the importance of adapting fraud detection methods within the VAT context, the role of social network analysis (SNA) should also be highlighted. Within the EU, Spain serves as a prime example of a country using an SNA tool called TESEO,[70] which is employed by AEAT (Agencia Estatal de Administración Tributaria)to visualize and analyze the relationships between taxpayers in a graph.[71] Similarly, countries like New Zealand and Singapore are carrying out social network analysis to help detect VAT carousel fraud and other group-level risks.[72] The use of SNA becomes particularly useful when individual assessments fail to identify risks, especially in cases where fraudulent activities or patterns arise from the interactions between multiple entities rather than from a single entity. It can help identify connections between individuals, such as through company directorships, joint bank accounts, or shared telephone numbers, and organizes these connected individuals into easily visualized networks.[73] The use of SNA is done alongside predictive models for the assessment of VAT risks, as demonstrated by at least two tax administrations.[74] The reason for this is because the relationships within the data are likely to vary depending on the type of risk involved.[75] For example, the patterns used to identify an error on a VAT return are likely to differ from those used to detect VAT carousel fraud. When a single model is tasked with identifying both types of patterns, it often struggles to perform well. As a result, administrations like Norway have chosen to use separate models for each type of risk. This means that the model designed to predict fraud and errors in VAT declarations is not the same one used to detect carousel fraud.[76] It is worth mentioning that, while the multi-model approach takes more time and resources to implement, its value has become evident in the context of fraud detection systems. For instance, it helps caseworkers focus their attention on the most high-risk aspects of a case, leading to more effective case management and greater adoption. Furthermore, if the connections between input variables and risk differ across risk types, this approach will produce more precise predictions.[77]

Expanding on the methods used in VAT to address fraud, another study illustrates how potential users of false invoices in a given year can be identified based on their tax payment information, historical performance, and characteristics, by utilizing various types of fraud detection systems.[78] The issue of false invoices in the context of VAT arises from the way tax liabilities are calculated. When a company accepts a false invoice, it creates a fictitious purchase, which artificially increases its tax credit and lowers its VAT payment. To characterize and identify patterns, the research started with unsupervised techniques like Neural Gas and Self-Organizing Maps (SOM), to identify relationships between tax payments related to false invoices and the behavioral variables linked to their usage. Subsequently, classification techniques were applied in cases where the conduct of fraud and no fraud was already known. By relying on supervised methods such as decision trees, neural networks, and Bayesian networks,

[67] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, supra 54,p 41723
[68] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, supra 54, p 41737
[69] D. Labança, L. Primerano, M. Markland-Montgomery, M. Polino, M. Carminati, and S. Zanero, supra 54, p 41737
[70] While TESEO seems to be an abbreviation its exact meaning or full form isn't explicitly detailed in the available sources.
[71] I. González García, en A. Mateos Use of Social Network Analysis for Tax Control in Spain. Universidad Politécnica de Madrid, p. 165
[72] See e.g. supra 13 EU and OECD supra 57, p.21
[73] See e.g. supra 13 and OECD, supra 57, p.21
[74] See e.g. supra 13 and OECD supra 57, p.21
[75] See e.g. supra 13 and OECD supra 57, p.21
[76] See e.g. supra 13 and OECD supra 57, p.22
[77] See e.g. supra 13 and OECD supra 57, p.22
[78] P.C. González, and J.D. Velásquez, J. D. (2014). Characterization and detection of taxpayers with false invoices using data mining techniques, Expert Systems with Applications, Volume 40, Issue 5, April 2013, p 1427

it was possible to identify variables linked to fraudulent and non-fraudulent behavior and detect associated behavioral patterns.[79] The research pinpointed that particularly the neural gas method found that it was possible to identify some relevant variables to differentiate between good or bad behavior, not necessarily associated with the use and sale of false invoices. [80] The decision tree method, applied to cases where the outcomes of fraud and non-fraud were known, proved effective in identifying variables that differentiate between fraudulent and non-fraudulent behavior. These variables varied depending on the size of the company. For small businesses, the key distinguishing factors were primarily the percentage of tax credits derived from invoices relative to total credit and the history of previous audits with negative results.[81] For medium and large companies, the most notable variables were the surplus credit carried over from previous periods, the proportion of credit tied to invoices, the cost-to-asset ratio, the degree of informality in accounting practices, and the company's age. Other crucial factors included the frequency of irregularities in past invoices, the number of payment orders, and the company's track record of not responding to notifications.[82]

The research concluded that, given the practical limitation of monitoring only a small group of companies each year, the most effective approach would be to combine the results from neural networks, decision trees, and Bayesian networks. This strategy would allow for the prioritization of companies for audit, focusing on those identified as fraudulent by the neural network and those showing the highest probability of fraud based on the decision tree and Bayesian network.[83]

4. Risks of fraud detection systems

The use of fraud detection systems, though potentially effective, also comes with certain risks which considering the vast amount of data analysed can have disastrous effects if these risks are not taken into account seriously. When it comes to the use of fraud detection systems it is first of all important to note that these systems provide a prediction. A prediction by definition holds a certain amount of (un)certainty.[84] Self-learning algorithms can find correlations in data without causation or reasoning power. There is a high probability of a relationship, but it is uncertain whether there is indeed a relationship and what it entails.[85] In other words correlation doesn't mean causation.[86] It is therefore important to remember that, although machine learning is a powerful tool for detecting fraud, it is crucial to recognize that it is not flawless and will inevitably make mistakes. An example of a common mistake occurs when legitimate transactions are incorrectly flagged as fraudulent. This incorrect positive assessment is known as a false positive.[87] The other possible mistake occurs when fraudulent transactions are mistakenly flagged as legitimate transactions, a situation known as a false negative. In statistics, a false positive is referred to as a type I error, while a false negative is known as a type II error. For clarity and ease of understanding, we will use the terms "false positive" and "false negative," as they are more explicit and easier to remember.[88] False positives and false negatives can result in erroneous decisions.[89] In case of false negatives there is a risk to overlook someone or a situation that should be suspicious.[90] In case of a false positive someone or a situation is erroneously regarded as a high risk and there may be a

[79] P.C. González and J.D.Velásquez, supra 78, p. 1427.
[80] P.C. González and J.D.Velásquez, J. D. supra 78, p. 1435.
[81] P.C. González and J.D.Velásquez, supra 78, p. 1435.
[82] P.C. González and J.D.Velásquez, supra 78, P. 1435.
[83] P.C González, P. C and J.D. Velásquez, supra 78, p. 1435.
[84] T.M. Berkhout, G.A. Raven en M. van Engers, supra 44, p. 24 and S. van Schendel, 'The Challenges of Risk Profiling Used by Law Enforcement: Examining the Cases of COMPAS and Syrl' in L. Reins, 'Regulating New Technologies in Uncertain Times', Springer 2019, p. 228.
[85] T.M. Berkhout, G.A. Raven en M. van Engers, supra 44, p. 27 and M. Finck, 'Reasoned A(I)dministration: explanation requirements in EU law and the automation of public administration', European Law Review 47 (3), p. 385.
[86] M.B.A. van Hout, 'Rechtsbescherming in het tijdperk van big data' (Legal protection in the era of big data), WFR 2017/165, para. 6.
[87] R. Wetzels et al. 'De daad bij het woord voegen: Fraudedetectiesystemen verankeren in de strategie van een organisatie' (Putting words to action: Embedding fraud detection systems in an organization's strategy), TaxTech 2023/6.
[88] R. Wetzelfs et al. , supra 87. See also: S. van Schendel, supra 84, p. 236 and R. Wetzels et al. supra 87.
[89] S. van Schendel, 'Data used in governmental automated decision making and profiling: Towards more practical protection' in B. van der Sloot and S. van Schendel, 'The Boundaries of Data', Amsterdam University Press 2024, p. 138.
[90] S. van Schendel, supra 84, p. 234.

risk of hyperenforcement.[91] The Dutch Wetenschappelijke Raad voor het Regeringsbeleid (WRR, Scientific Council for government policy) has pointed out that e.g. the use of profiling can lead de facto to a reversal of the burden of proof where an individual needs to prove that the risk linked to a certain group or situation does not apply to him, instead of the government providing the proof that the profile applies in his individual case.[92] This is one of the consequences of the fact that the outcome of profiling can cause a biased attitude among the authorities.[93] This is also referred to as automation bias. This is a phenomenon where humans ascribe a certain authority to outcomes suggested by automated processes that lead them to neglect other available information or counter indications.[94] Wetzels et al. describe that both the risk of false negatives and false positives should be taken into account. In the Netherlands after the so called Bulgar fraud the focus has shifted to false negatives instead of false positives which subsequently had consequences for how alleged fraud was addressed regarding child benefits, and which has played a role in the child benefit scandal.[95] Since the outcome of a fraud detection system is a likelihood, chance or probability it is important that the outcome is only used for signaling risks. Such risks should subsequently be investigated by humans.[96]

Secondly, the success (and failure) of a fraud detection system depends on its accuracy. Training data can be incomplete, incorrect, out of date or not representative,[97] which enlarges the risks of erroneous outcomes such as false positives or negatives. The more comprehensive and accurate the training data, the better the system can understand the relationship between inputs and outputs. Van Schendel points out that data in the field of criminal law for example is limited. Crime data is not real time data of actual crime. They reflect crime that was caught or reported and recorded. The same is true in the field of VAT. Available data is data of fraud (patterns) that have been discovered. For all data it must be considered that they are gathered in a certain way and that they are always a representation of reality.[98] In its report Advanced Analytics for better Tax Administration the OECD concludes that data available to train predictive models comes from highly biased samples.[99] Fraud detection modes also require a different approach to data. Data points that are not relevant for operational purposes may be highly relevant for analysis.[100] Last, but certainly not least, it is important that the models are retrained using new data. Without retraining the performance of the model declines rapidly.[101] In the context of customs operations, for example the list of countries procuring a particular product will change over time and products that are foreign to the system (e.g. because of new technology) may be declared.[102] Kim et al. in this respect propose a strategy to address these changes. It combines the approaches of exploitation and exploration. The exploitation approach selects the most likely fraudulent and highly profitable items to secure short term revenue for customs administration. The exploration on the other hand selects uncertain items at the risk of temporary revenue regret, but with the possibility to detect novel fraud patterns.[103] When a risk based model is used, e.g. for profiling, it is important that the risk indicators used are valid and reliable. Valid means that there is a statistic correlation between the indicator and a certain behaviour

[91] Sofia Ranchordás, 'Administrative Blindness: All the Citizens the State cannot see', Inaugural Lecture Tilburg University 2024, p. 35. Such a hyperenforcement took place in the Dutch child benefit scandal. In this scandal innocent citizens were regarded as fraudsters based on the used methods of risk based supervision. In case of mistakes or inaccuracies or if part of the costs could not be justified, the full benefit was reclaimed from the citizens. See: Parlementaire enquêtecommissie Fraudebeleid en Dienstverlening (Dutch Parliamentary Inquiry Committee on Fraud Policy and Service Delivery), 'Blind voor mens en recht' (Blind for people and justice), 26 February 2024, p. 163 and 181.
[92] WRR, 'Big Data in een vrije en veilige samenleving' (Big Data in a free and safe society', report nr. 95, 2016, p. 142.
[93] R.N.J. Kamerling en A.K.H. Klein Sprokkelhorst, 'Renseignering in het 'big data' tijdperk', (Information gathering in the big data era), WFR 2020/199, para. 4.
[94] M. Finck, supra 85 p. 387.
[95] R. Wetzels et al. supra 87.
[96] T.M. Berkhout, G.A. Raven en M. van Engers, supra 44, p. 24 and Parlementaire enquêtecommissie supra 91, p. 271.
[97] M. Finck, supra 85, p. 379 and Parlementaire enquêtecommissie supra 91, p. 59.
[98] S. van Schendel, supra 89, p. 148 and Parlementaire enquêtecommissie supra 91, p. 269 and 270.
[99] OECD supra 57, p. 51.
[100] OECD supra 57, p. 52.
[101] A. Nesvijevskaia et al. 'The Accuracy versus Interpretability Trade-off in Fraud Detection Model', Data & Policy (2021), 3: e12 doi:10.1017/dap.2021.3, para. 2.3.4.
[102] Sundong Kim et. al., 'Active Learning for Human-in-the-Loop Customs Inspection' IEEE Transactions on Knowledge and Data Engineering PP(99), doi:10.1109/TKDE.2022.3144299, p. 1.
[103] Sundong Kim et. al., supra 102, p. 1.

that the indicator seeks to predict. Reliable means that the indicator maintains the same correlation in different contexts.[104]

Another commonly reported and important risk are biases which could result in prohibited discrimination and stigmatization.[105] In case of a bias disadvantageous effects can hit certain groups without a good objective reason.[106] The more datadriven systems become the more difficult it is to disentangle the biased data from decision making, making the bias more hidden and exacerbating the inequalities.[107] Authorities using the system may thus be unaware of discrimination happening in their dataset or algorithms.[108] Raven describes a situation where the trainingset contains biased data, because a certain group is deliberately or unintentionally under or overrepresented. This imbalance of the dataset can substantially affect the accuracy of the systems trained on them.[109] If based on this dataset predictions are made certain groups will be overrepresented in the selection of the algorithm. By checking selected cases and including the results in the next training cycle overrepresentation will increase, creating a self-fulfilling prophecy with disadvantageous effects for certain groups without a good and objective justification.[110] Another bias is the confirmation bias. This relates to the fact that people are inclined to seek and interpret information in such a way that it substantiates their hypotheses and that they will ignore information that proves the contrary. This can also create a self-fulfilling prophecy in the sense that fraud is more frequently investigated within a certain group, also causing fraud to be found more frequently within that group. This confirms the hypothesis, which will result in a more frequent selection of people within this group and a further reinforcement of the hypothesis.[111] What's more, we should also be aware of function creep. Function creep is the use of data for a different purpose than for which they are collected. It is as easy as connecting one system to another that has a different function.[112] The extent to which function creep poses a risk depends on the type of data and the consequences of the analysis. Function creep poses a risk when data, organizations or functions are combined that were previously separated for good (legal) reasons.[113]

Finally legal protection is an important issue to consider. This has to do with the so-called black box problem of AI. Even though the input and output are clear, the process and the reason a system reached its conclusions can be opaque to a more or less extent.[114] From a democratic standpoint, the balance of power between the state and citizens, as well as between companies and citizens, is shifting. A transparency paradox emerges in this context, since on the one hand, citizens are becoming increasingly transparent to both governments and businesses; while on the other hand, the methods used for analysis remain opaque to the public, rather profiles and algorithms are seldom transparent or traceable.[115]The complexity of the algorithm makes the decision making process opaque, with the result that a good, known and transparent explanation of a decision is lacking.[116] Providing such an explanation would require disclosing to a human how different factors were weighed by the system to reach a decision and what input was determinative.[117] In this respect research has been done about explainable AI or XAI or white box AI. The aim is to show how the AI system's input affected its output by revealing the link between the

[104] Parlementaire enquêtecommissie supra 91, p. 269 and 322. For example in the Dutch child benefit scandal nationality was used as a risk indicator. The risk indicator nationality is irrelevant, as it is not necessary to have the Dutch nationality to obtain the child benefit. The same applies for the assumption that guest parent agencies that disregard the rules about quality are also financially inaccurate as this has never been investigated or established. Therefore this risk indicator is neither valid nor reliable. Parlementaire enquêtecommissie supra 91, p. 316 and 321.
[105] S. van Schendel, supra 89, p. 138 and R.N.J. Kamerling en F. Muis-Visser, 'Nieuw feit in een tijdperk van big data en data-analyse: rechtsbescherming op de tocht?' (New fact in an era of big data and data analysis: legal protection at risk?), WFR 2021/62, para. 4.
[106] T.M. Berkhout, G.A. Raven en M. van Engers, 'supra 44, p. 17 and Parlementaire enquêtecommissie supra 91, p. 236.
[107] S. van Schendel, supra 89, p. 148 and T.M. Berkhout, G.A. Raven en M. van Engers supra 44, p. 24.
[108] S. van Schendel, supra 84, p. 237.
[109] A. Nesvijevskaia et al., supra 101, p 12-6.
[110] G.A. Raven, 'Het juridisch kader voor algoritme-gedreven technologieën in boekenonderzoeken' (The Legal Framework for Algorithm-Driven Technologies in Tax Audits), MBB 2024/23, para. 3.1.
[111] Parlementaire enquêtecommissie supra 91, p. 270.
[112] M.B.A .van Hout, supra 86, para. 6.
[113] WRR, supra 92, p. 91.
[114] M. Finck, supra 85, p. 377.
[115] WRR, supra 92, p.93
[116] G.A. Raven, supra 110, para. 3.2 and R.N.J. Kamerling en A.K.H. Klein Sprokkelhorst, supra 93, para. 4.
[117] M. Finck, supra 85, p. 384.

data ingested by the system and the decision it makes.[118] When it comes to explainable AI or transparency a balance should be struck between the interest of the subject of the decision made by AI on the one hand to know how the system reached its decision and the interest of the government authority using the AI on the other, to avoid disclosing so much about the system that those subject to it know how to avoid detection. Van Hout points out that providing openness about the decision process does not mean that the tax administration should disclose all information about enforcement strategies used.[119]

5. Risks related to the use of digital reporting and e-invoicing

5.1 Introduction

As outlined in section 2, the EU's proposal for Digital Reporting Requirements (DRR) aims to introduce a real-time digital reporting for VAT purposes based on e-invoicing. Once implemented, Member States will have more information they need to step up the fight against VAT fraud, especially carousel fraud or MTIC fraud.[120] This initiative represents astep forward, aligning with the global shift toward real-time reporting and the adoption of technology for detecting and preventing fraud by tax authorities. Beyond enhancing transparency and improving the efficiency of tax compliance processes, the data collected through DRR will provide tax authorities with significant opportunities to conduct advanced analysis using fraud detection systems. For instance, the data collected can be utilized to train fraud detection models, enabling the prediction of fraudulent activities or the identification of anomalies, as discussed in section 3. However, as highlighted in section 4, there are inherent risks associated with the use of such systems. In this section, we will focus on evaluating the risks associated with the use of the DRR and e-invoicing, concluding with a set of recommendations.

5.2 Evaluation of risks of fraud detection under e-invoicing and digital reporting

The first risk we described in section 4 is that fraud detection systems merely provide a prediction that has a certain amount of (un)certainty. One of the primary concerns related to handling the data collected through DRR is therefore that tax authorities may automatically treat a signaled risk of fraud as fraud case ignoring that it could be a false positive and wrongly classifying human mistakes as fraud. This issue was highlighted by the Dutch Parliamentary Inquiry Committee on Fraud Policy and Service Delivery, where it was pointed out that the model used by the Benefits Department failed to distinguish between applications with a high risk of fraud and those with a high risk of error. As a result, applicants that had made errors were mistakenly treated as fraudsters.[121] The same issue can arise when assessing data reported for VAT purposes. If the system used by tax authorities cannot distinguish between transactions with a high risk of fraud and those with a high risk of error, companies may be wrongly classified as fraudsters. This may trigger the response to fraud instead of the response to errors, i.e. hyperenforcement. It should be noted that this can significantly harm companies, not only financially, but also as regards their reputation. Being accused of fraud, which is universally regarded as a criminal offense, could result in even more severe consequences, such as the freezing of assets or the blocking of bank accounts, as stipulated by the criminal codes of various jurisdictions. For instance, in Poland, there is an IT system known as STIR (Clearing House Information and Communication System), which is used to monitor financial account activity in entrepreneurs' bank accounts to prevent tax evasion and fraud. Data is reported by banks and credit unions to the National Clearing House (KIR) which establishes a risk score for each person based on an algorithm.[122] If there is suspicion that an entrepreneur's account is being used for tax fraud, the Head of the National Tax Administration (KAS) has the authority to block the account for

[118] Lukasz Gorski et al., Exploring explainable AI in the tax domain', Artificial Intelligence and Law' 7 May 2024, section 2.
[119] M.B.A .van Hout, supra 86, para. 7.
[120] M. A. Papis-Almansa, VAT in the Digital Age. Real-time digital reporting based on e-invoicing for businesses (comments by Marta Papis-Almansa). Highlights & Insights on European Taxation, 2023(2). p.1. However in itself DRR cannot tackle EU-wide fraud, see section 2.3.
[121] Parlementaire enquêtecommissie supra 91, p. 52
[122] M. Papis-Almansa, The Polish Clearing House System: A 'Stir'ring Example of the Use of New Technologies in Ensuring VAT Compliance in Poland and Selected Legal Challenges, 28 EC Tax Review 1 (1 Feb. 2019), p. 45.

72 hours.[123] This period can be extended up to three months if there is reasonable suspicion that the entrepreneur will fail to meet a tax liability of at least €10,000. Additionally, the National Tax Administration has also the authority to refuse or cancel an entrepreneur's VAT registration based on a risk indicator provided by the KIR[124]. In the latter case information about these companies is made public, including – in case of legal persons – the name of persons authorized the entity and its shareholders. The objective is to protect compliant taxable persons and trustworthy entrepreneurs.[125] We also discussed that not taking into account sufficiently that the system just provides a prediction could lead to automation bias, where other available information or counter indications are neglected. As Jafari, Lamensch and Papis-Almansa explain, the mere lack of compatibility between output and input tax can inform tax authorities on potential irregularities but cannot by itself constitute the basis for the denial of the right to deduct because of fraud. The lack of compatibility could have a different explanation, such as errors.[126] The risk of automation bias, in our opinion, is amplified by the fact that evidence suggests that tax enforcement activity has gravitated towards tackling the 'low-hanging fruit', meaning tax authorities tend to prioritize cases that boost revenue with low administrative costs, thus improving performance statistics.[127] It is crucial to recognize that this approach might pose a risk of bias and stigmatization, potentially leading to certain groups of entrepreneurs, especially small and medium-sized enterprises (SMEs), to be targeted as low-hanging fruit. Literature demonstrates that although most SMEs seek to comply, factors such as outdated systems and a lack of understanding of the tax system have historically contributed to non-compliance,[128] making them more susceptible to being singled out. Simply because they make more mistakes. If more cases of SMEs come up and are investigated after which the data is fed back into the system their cases might be overrepresented. This again can lead to more of these cases being selected and investigated resulting in a self-fulfilling prophecy as discussed in section 4.

It is also inevitable not to think that this situation creates a potential risk, where mismatches in the data collected may arise not only from fraudulent activity but also from simple errors. The combination of tax authorities' focus on meeting revenue goals could, in turn, create a dangerous mix, leading to incorrect decisions, including the misclassification of legitimate errors such as fraud. As regards mandatory e-invoicing and digital reporting obligations Van der Hel-Van Dijk and Griffioen point out that these measures will lead to more standardization and digitization of record keeping, which will reduce errors made by taxpayers who are willing to adhere. This in turn will lead to better compliance and more VAT revenue for the state budget, but not necessarily to the prevention of MTIC fraud.[129]

As discussed in section 4, another key risk associated with fraud detection systems is their reliance on accurate data. In the area of VAT, one should also be aware of the fact that the example used to train a fraud detection system is always limited. The data reflects fraud that was caught and recorded. We therefore recommend an approach of combined supervised and unsupervised learning as discussed in section 3 and exploitation and exploration as discussed in section 4 to make sure that new types of fraud are discovered and can be used to keep the fraud detection system up to date. Even if that approach is taken it should however be kept in mind that there will also be some limitation to data.

A third risk that we discussed are biases. Though we don't have any concrete examples yet about how such a risk would emerge in the field of VAT it is most definitely important to take note of this potential risk. We see function creep within the field of VAT already, as the digital reporting data is e.g. being linked to Transaction Network Analysis (TNA) and the Central Electronic System of Payment (CESOP). The TNA helps tax authorities identify suspicious transaction patterns that may indicate VAT fraud. It analyses

[123] Please see division III B of the Tax Ordinance (Articles 119zg – 119zu). Retrieved from: Counteracting the use of the financial sector for tax fraud - Section 3B - Tax Ordinance. - Journal of Laws 2023.2383 i.e. - OpenLEX accessed 18.12.2024

[124] B. Kuźniacki et al., Towards eXplainable Artificial Intelligence (XAI) in Tax Law: The Need for a Minimum Legal Standard, 14 World Tax J. 4 (2022), Journal Articles & Opinion Pieces IBFD (accessed 2 September 2022). p 9

[125] M. Papis-Almansa, supra 122, p. 47.

[126] Sacha Jafari, Marie Lamensch and Marta Papis-Almansa, 'Proposal for a Secure Digital Reporting Standard for Intra-Community Transactions', International VAT Monitor, 2022 (Volume 33), No. 6, https://doi.org/10.59403/2w3rrvj, p. 242.

[127] Rita. de la Feria, Tax Fraud and Selective Law Enforcement. *Journal of Law and Society*, vol. 00, no. 0, 2020, p.27.

[128] OECD (2014), Tax Compliance by Design: Achieving Improved SME Tax Compliance by Adopting a System Perspective, OECD Publishing, p. 11, Deloitte (2016) Lot 3, p. 72

[129] Lisette van der Hel-van Dijk and Menno Griffioen supra 2, section 5.1.

transaction data across multiple countries, enabling authorities to detect issues such as missing traders or carousel fraud. The CESOP is a system designed to help monitor cross-border payments and ensure proper VAT payment. It collects data on payments between payers and payees in different EU countries or a country outside the EU and an EU country, enabling tax authorities to track these transactions and identify potential fraud. In essence, CESOP ensures VAT compliance across borders. The fourth recital of the preamble of Regulation 2020/283,[130] however demonstrates that CESOP intended to target B2C-transactions. It is now, however, also used to target potential fraud in B2B-cross border transactions. Whether this indeed creates issues is a question beyond the scope of this article as it is a research topic in its own right. In any case the data is still used in the field of VAT and still to target potential fraud. Although the proposal for e-invoicing and digital reporting is accompanied by safeguards which include that the data that Member States share with each other cannot be used for any purpose whatsoever,[131] we do not see such safeguards concerning the data that Member States themselves collect under these rules (which they will subsequently have to share with other member states). In this way, carefully balanced interests that form fundamental rights, such as the right to privacy and the freedom to conduct a business, in our opinion, risk being harmed by function creep at the national level in a manner that is not legally justified.

Finally, we discussed the risk of lacking legal protection. In this respect it is important to note that significant concerns have already been raised about the STIR (as mentioned above), particularly because the AI algorithms used to assess these risks are not disclosed to taxpayers or courts, even during legal disputes with the National Revenue Administration (NRA). This lack of transparency raises serious questions about the compatibility of the STIR system with fundamental taxpayer rights.[132] Moreover, technology-facilitated measures, such as blocking bank accounts or flagging and removing taxpayers from the VAT register, can interfere with the freedom to conduct business. While intended to prevent fraud, these actions may disproportionately impact legitimate businesses, compounding concerns about their implementation and oversight.[133] The Polish rules are in particular intrusive because the VAT registration will be restored only if the taxable person proves that his activities are carried out without fraudulent intent. This burden of proof is particularly heavy because the taxable person does not have complete information regarding the grounds for the decision.[134]

5.3 Recommendations

Without the proper measures and safeguards in place, the risks outlined in section 4 and 5.2 could materialize, potentially leading to unjust outcomes. Generally speaking the risk discussed above can be addressed by authorities by first making a well-founded decision for the method or analysis to be used based on the purpose of the analysis.[135] They should also make sure the data used within the fraud detection system is up to date and the system contains no bias.[136] This can be done by applying technical measures to prevent, correct or compensate for irregularities and discrimination bias.[137] Algorithms and methods should be sound and meet the requirements of scientific good statistic research.[138] Authorities should also validate whether the model or algorithm correctly performs its intended functionality, periodically audit the accuracy of the model or algorithm and document which models, algorithms and

[130] Council Regulation (EU) 2020/283 of 18 February 2020 amending Regulation (EU) No 904/2010 as regards measures to strengthen administrative cooperation in order to combat VAT fraud, OJ L 62, 2.3.2020, p. 1–6.

[131] See the 4th and 12th recital of the preamble to the Draft Council Regulation amending Regulation 904/2010 where it is stated that the information in the central VIES is not to be used for other purposes than to monitor the correct application of VAT and combat VAT fraud. Art. 24k (1) describes which officials will be granted access to central VIES and art. 24g (3) makes the storage of information from central VIES in a national system subject to the same access permissions. No such limitations however apply when it comes to data collected by the Member State itself (which it subsequently transfers to central VIES).

[132] M. Papis-Almansa, supra 122, A. Bal, Ruled by Algorithms: The Use of 'Black Box' Models in Taxation, 95 Tax Notes International 12 (2019); M. Rojszczak, Compliance of Automatic Tax Fraud Detection Systems With the Right to Privacy Standards Based on the Polish Experience of the STIR System, 49 Intertax 1, pp. 46-49 (2021)

[133] M.A. Papis-Almansa, The Use of New Technologies in VAT and Taxpayers' Rights. In G. Kofler, M. Lang, P. Pistone, A. Rust, J. Schuch, K. Spies, C. Staringer, & I. Kuniga (Eds.), CJEU: Recent Developments in Value Added Tax 2021, p18.

[134] M. Papis-Almansa, supra 122, p. 47 and 54.

[135] G.A. Raven, supra 110, para. 4.2.

[136] WRR, supra 92, p.139.

[137] G.A. Raven, supra 110, para. 4.2.

[138] WRR, supra 95, p.136 and G.A. Raven, supra 110, para. 4.2.

datasets have been used, to make it possible to validate an audit as mentioned above.[139] They should also evaluate the fraud detection system periodically. According to the Dutch Scientific Council for Government Policy this should include testing whether there is still a need to apply the fraud detection system, whether the system is effective and whether the benefits outweigh the costs. When evaluating the latter the breach of fundamental rights of the persons involved should be taken into account as well.[140] In this respect it is important to note that pursuant to the new art. 271c VAT Directive the European Commission will assess only the effects of e-invoicing and digital reporting on the effectiveness of the VAT collection, the reduction of the VAT Gap, the number of controls carried out by the tax administration and the reduction of the administrative burden and costs savings for taxable persons and the measures put in place by Member States to alleviate the taxpayers' administrative burden. The evaluation does not include fraud detection systems making use of the data collected under e-invoicing and digital reporting.

According to the Dutch Scientifical Council of Government Policy a greater level of transparency is possible, and organizations should draft a policy plan which mentions which methods they use including costs and estimated results.[141] This is in our view crucial both for those working with these models and algorithms and for individuals affected by decisions made using them. In particular, the decisions of tax authorities must comply with the principle of formal motivation. To comply with this principle, all factual and legal grounds on which the decision is based should be mentioned and explained by the tax authorities.[142] The rationale for such decisions must be clear, precise, and genuinely reflect the underlying motives.[143] It should also be noted that even if information is provided on how data is processed and incorporated in the decision making process, technical skills to understand the operation of AI systems are required to fully understand the decision making process. In this respect an explanation about how such a system operates should be provided as well.[144] We believe that by incorporating explainability as a core requirement, particularly in decisions affecting individuals, organizations can foster trust and comply with legal principles.

When these requirements are not met by self-regulation within governmental authorities one can consider taking it a step further by legally regulating e.g. the general requirements for quality of data and methods used or the acceptable margin of error.[145] One can even consider introducing an independent authority supervising that these requirements are met.[146] This is particularly relevant where individuals (including legal entities) are unable to verify the legality of the algorithm as it increases the risk of abuse of power. Certification or control by an independent body, such as a dedicated judicial authority, is necessary where sufficient information about the functioning of a fraud detection system cannot be disclosed.[147]

Given the inherent risks tied to fraud detection systems, we especially believe tax authorities should exercise caution to avoid over-reliance on these tools. While these systems have become indispensable in this modern era due to the immense volume of data that can be collected through digital reporting, it is crucial to use them cautiously. We believe that tax authorities should use fraud detection systems as tools for assistance rather than as definitive solutions. It is important that, while using these systems, they refrain from adopting a black-and-white, all-or-nothing approach that oversimplifies the analysis. Instead of focusing solely on the total number of incorrect transactions when assessing fraud cases, authorities should prioritize identifying genuine instances of fraud.[148] Tax authorities should bear in mind that there are also other reasons why mismatches can occur, and it does not necessarily point to fraud. A mismatch between a declared intra-Community supply but no declared intra-Community acquisition under the VAT identification number of the purchaser can, for example, arise due to a number acquisition/safety net

[139] G.A. Raven, supra 110, para. 4.2.
[140] WRR, supra 92, p. 141.
[141] WRR, supra 92, p. 13 and 18.
[142] B. Kuźniacki et al., supra 124, p. 7
[143] Art. 41 Charter of Fundamental Rights of the European Union (the EU Charter), OJ C 326/391, 26 Oct. 2012
[144] B. Kuźniacki et al., supra 124 Section 3.1.
[145] WRR, supra 92, p.10.
[146] WRR, supra 92, p. 10.
[147] M. Rojszczak, supra 133, p. 48.
[148] Parlementaire enquêtecommissie, supra 91, p. 181

acquisition under art. 41 VAT Directive where the actual acquisition has already been declared in the EU member state of arrival of the goods,[149] or in the situation where there is a concurrence of exemptions/zero rates, such as an intra-Community supply from a customs warehouse, where one Member State takes the position that an intra-Community transaction must be declared, but the other Member State considers that the exemption/zero rate due to supply in a customs warehouse takes precedence. A mismatch in the reported values of a cross-border sale can also occur when both parties use different exchange rates to convert the transaction amount.

6. Conclusion

In this article we addressed the following research question: What risks do advanced fraud detection based on e-invoicing and digital reporting entail and how can these risks be mitigated? We defined four risks that could materialize within the context of e-invoicing and digital reporting currently or in the future. The first and in our view most prominent risk is ignoring that fraud detection systems merely provide a prediction and treating a prediction as a given case of fraud. If this takes place e.g. errors can mistakenly be regarded as fraud with the equivalent response, having a substantial negative impact on honest businesses. Combined with the tax authorities focusing on meeting revenue goals this could create a dangerous mix, leading to incorrect decisions. The second risk that we discussed is the limitation of data and the accuracy of the fraud detection model. A combined approach of supervised and unsupervised learning as discussed in section 3 and exploitation and exploration as discussed in section 4 allows for making sure that new types of fraud are discovered and keeping the fraud detection system up to date, while also securing tax revenue by detecting known types of fraud. The third risk we defined are biases which can result in prohibited discrimination and stigmatization, and function creep, where data is used for a different purpose than for which it is collected. Though we cannot yet pinpoint examples of biases in the field of e-invoicing and digital reporting yet, it is important to take this risk into account. Function creep already materializes itself in the field of VAT by simply linking the central VIES to TNA and CESOP. We also observed that there are no safeguards dealing with for what purpose data reported to Member States under e-invoicing and digital reporting are used, creating the risk that Member States will use this data for different purposes than combatting fraudulent transactions and ensuring VAT compliance. The fourth and last risk we defined is the risk of lack of legal protection as the decision-making process of fraud detection models are opaque with the result that a good explanation of a decision cannot be provided. We already see such a risk materializing in Poland under STIR. The risk is especially prominent as the burden of proof is put on the taxable person to demonstrate that his transactions are not carried out with fraudulent intent, while he lacks information about the ground of the decision.

Generally speaking, the risks associated with fraud detection models should be addressed by well-founded decisions about which method of analysis to be used given the purpose of the analysis, applying measures to prevent, correct or compensate for irregularities and discrimination biases, using only good sound and scientific statistic research and by validating and auditing the system periodically. When evaluating the breach of fundamental rights of the individuals involved should also be considered. A greater level of transparency should be provided, including information about how the system operates. Where this is impossible as that would entail disclosing too much information, an independent (judicial) body could certify or supervise the use of the fraud detection system. Where self-regulation by governmental authorities is insufficient one can consider legally regulating e.g. the general requirements for quality of data and methods used or he acceptable margin of error. In the field of e-invoicing and digital reporting it is in our view important that tax authorities use fraud detection systems as tools of assistance rather than definitive solutions. While AI and machine learning are powerful tools for detecting fraud, they are not infallible and can make mistakes.

[149] In our view it can be inferred from the B-case (CJEU 7 July 2022, C-696/20, ECLI:EU:C:2022:528) that where the intra-Community acquisition has already been declared in the Member State of arrival of the goods it is not necessary to report an intra-Community acquisition in the Member State whose VAT identification number has been used in the transaction. See also M.M.W.D. Merkx case law note on CJEU 7 July 2022, C-696/20, ECLI:EU:C:2022:528, FED 2022/93.