



ML-based Fraud Detection in Public and Private Sectors: A Legal Analysis of Risks, Benefits and Regulatory Gaps

*Sander Pype**

* Sander Pype holds a Master of Laws from the University of Antwerp and an LL.M. in Intellectual Property Law and Technology Law from The Ohio State University Moritz College of Law. In the academic year 2026–2027, he will pursue an LL.M. in Intellectual Property and ICT Law at KU Leuven. His research interests include intellectual property law, technology law, and the legal challenges arising from technological innovation.

Inhoudstafel

I.	Introduction	1
A.	Evolution of fraud detection systems	1
B.	Research question and objectives	2
II.	Legal framework for ML-based fraud detection	2
A.	General Data Protection Regulation.	3
1.	Applicability	4
2.	Article 5 GDPR	4
(a)	Lawfully	5
§1.	Private entities.....	6
§2.	Public entities	9
(b)	Fairness	10
(c)	Transparency	12
3.	Art. 22 GDPR.....	14
(a)	Applicability	14
(b)	Exceptions to art. 22 GDPR.....	18
B.	AI Act	21
1.	Applicability	22
2.	Risk-level.....	26
(a)	Public ML-based fraud detection.....	28
(b)	Private sector ML-based fraud detection systems	31
3.	Consequences falling under high risk:	33
III.	Case studies	36
A.	Public sector use of ML-based fraud detection systems.....	36
1.	Dutch Child benefit scandal	36

(a) GDPR.....	38
(b) AI Act.....	41
(c) Conclusion.....	43
2. HMRC ‘Connect’.....	44
(a) GDPR.....	46
(b) AI Act	48
(c) Conclusion.....	49
B. Private sector use of ML-based fraud detection.....	50
1. Payment sector.....	51
(a) Paypal and Zelle.....	52
(b) GDPR	55
(c) AI Act	57
(d) Conclusion.....	58
2. Uber fraud detection.....	59
(a) GDPR.....	60
(b) AI Act	62
(c) Conclusion.....	63
IV. Comparison	63
V. Conclusion	65
VI. Bibliography	66

I. Introduction

A. Evolution of fraud detection systems

Over the past decades, fraud detection has undergone a significant transformation. Fraud detection systems have evolved from rule-based systems to systems using machine learning techniques.¹ Rule-based systems work by first manually creating rules (often in an if-then format) which a computer system then applies to individual cases.² When a rule-based system is not updated regularly it can lead to a rigid system.³ In many ways the world of fraud detection is a perpetual arms race⁴ between fraudsters and fraud-system creators. Once fraudsters know and understand the rule that is applied to detect them, they will find ways to avoid detection, thus for a rule-based system to work optimally, constant updates are required.⁵ Another problem is that rule-based systems are based on the subjective opinion of the people creating the rule,⁶ which can lead to rules not being aligned with reality.

Machine learning refers to the capacity of programs to learn in a ‘functional’ sense. Although distinct from human learning, machine learning techniques enable computer programs to adapt their behavior to improve their performance on some task through experience.⁷ In the context of fraud detection, machine learning techniques are employed in order to identify anomalies and fraudulent trends from large data-bases.⁸ Through

¹ Himanshu Gupta, Punit Soni and Rajender Kumar, ‘The Evolutionary Arms Race - Safeguarding Financial Stability through Advanced Fraud Detection Techniques’ in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)* (the Institute of Electrical and Electronics Engineers 2024) 2.

² Crina Grosan and Ajith Abraham, *Intelligent Systems: A Modern Approach* (Springer-Verlag Berlin Heidelberg 2011) 149-150.

³ Gupta (n 1), 2.

⁴ Richard A. Becker, Chris Volinsky and Allan R. Wilks, ‘Fraud Detection in Telecommunications: History and Lessons Learned’ (2010) 52 *Technometrics* 20, 20.

⁵ Brandon Charlotte and Afex Callagher, ‘Artificial Intelligence Approaches to Fraud Detection in Public Procurement’ (unpublished manuscript, 4 September 2025), 2.

⁶ Anthony Simonofski and others, ‘Balancing Fraud Analytics with Legal Requirements: Governance Practices and Trade-Offs in Public Administrations’ (2022) 4 *Data & Policy* e14, e14-3.

⁷ Harry Surden, ‘Machine Learning and Law’ (2014) 89 *Wash L Rev* 87, 89.

⁸ Joy N. Okolo and others, ‘Analyzing the Use of Machine Learning Techniques in Detecting Fraudulent Activities’ (2025) 26 *World Journal of Advanced Research and Reviews* 1198, 1198.

continuous learning , fraud detection systems using machine-learning (ML-based fraud detection systems) are able to adapt to new fraudulent schemes.⁹ The case studies used in this thesis illustrate that ML-based fraud detection systems rarely make a final decision. In most instances, they function as a first filter and are used to calculate a ‘risk score’, that a human reviewer then uses to make a final decision. Still, as will become clear throughout this thesis, this does not stop ML-based fraud detection systems from having given rise to new ethical and legal challenges. While the GDPR and the AI Act offer a degree of protection against the challenges posed by ML-based fraud detection systems, this thesis argues that these frameworks remain not fully adapted to the realities of ML-based fraud detection.

B. Research question and objectives

This thesis will explore how ML-based fraud-detection systems are currently being used in the private and public sector . It will use case studies to make this as concrete as possible. The following case studies are used: The Dutch Child Benefit scandal, the Connect system of the UK tax authorities, the ML-based fraud detection systems used by PayPal and Zelle and the ML-based fraud detection system used by Uber. Throughout these case studies, the thesis aims to answer the following research questions:

- How are ML-based fraud detection systems currently used by public and private entities?
- What are the risks and benefits of using ML-based fraud detection systems in the private and public sector?
- Does the current general legal data protection framework (AI Act, GDPR) provide enough protection against the risks that come with the use of ML-based fraud detection systems?

II. Legal framework for ML-based fraud detection

This thesis does not aim to provide an exhaustive overview of the complete legal framework governing fraud detection. Given the focus of this thesis being ML-based fraud detection, only the laws that are particularly relevant to the use of machine learning in

⁹ Okolo (n 8), 1199.

this context will be explored. This includes the GDPR¹⁰ and the AI Act.¹¹ Furthermore, fundamental rights, like the rights of defense¹² and the presumption of innocence¹³ will also not be discussed.¹⁴ While it is true that many fundamental rights can have an impact on the risks that come with the use of ML-based fraud detection, discussing these rights would make the scope of this thesis too large to remain comprehensive.

A. General Data Protection Regulation.

The GDPR is relevant for this thesis due to one of the objectives of the GDPR specifically being the protection of natural persons in relation to the processing of their personal data.¹⁵ Since ML-based fraud detection systems often engage in some form of personal data collection,¹⁶ it is essential to investigate the impact of the GDPR on ML-based fraud detection.

¹⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016

on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[2016] OJ L 119/1 ('GDPR').

¹¹ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 ('AI Act').

¹² Charter of Fundamental Rights of the European Union [2012] OJ C 326/391 ('Charter'), art 48(2).

¹³ Charter, art 48(1).

¹⁴ See David Hadwick and Shimeng Lan, 'Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany' (2021) 13 World Tax Journal. which provides an analysis of the fundamental-rights implications of the Dutch childcare allowance scandal.

¹⁵ GDPR, art 1.

¹⁶ Kadir Ider and Andreas Schmietendorf, 'Data Privacy for AI Fraud Detection Models: A framework for GDPR compliant AI' in Jaime L. Mauri and others (eds) *ICDS 2020 : The Fourteenth International Conference on Digital Society* (International Academy, Research, and Industry Association (IARIA) 2020) 103.

1. Applicability

Almost all forms of ML-based fraud detection necessitate some form of personal data collection and thus will have to comply with the GDPR.¹⁷ Personal data is defined by the GDPR as any information related to an identified or identifiable natural person.¹⁸ The recitals of the GDPR determine that only true anonymous data that is not able to be related in any reasonable way to an identifiable person will not be considered as personal data. Data that is pseudonymized, but can still be attributed to an identifiable natural person with the use of additional information should still be considered as personal data.¹⁹

While in practice there is still a lot of discussion about what personal data entails,²⁰ it is clear that most data processed by ML-based fraud detection systems will fall under the scope of the GDPR. While certain stages of ML-based fraud detection, such as the training phase, may rely on anonymized or synthetic data,²¹ in practice ML-based fraud detection systems are designed to detect fraud and link this fraud to a certain identifiable natural person. If a fraud detection system is only able to detect fraud, but unable to attribute it to a natural person, the system is functionally useless, as it would not allow entities to stop the fraud. ML-based fraud detection systems will thus in almost every case process personal data. An exception however exists when the ML-based fraud detection system is used to detect fraud committed by legal persons.

2. Article 5 GDPR

Art. 5 of the GDPR lays down important principles governing data processing. It determines that personal data shall be processed:

- Lawfully

¹⁷ Ider n (14), 103.

¹⁸ GDPR, art 4.

¹⁹ GDPR, rec 20.

²⁰ Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data under the GDPR' (2020) 10 International Data Privacy Law 11, 14.

²¹ ML-based fraud detection systems sometimes use anonymized data during their training phase. For an example see: Spyridon D. Lampropoulos, Georgios L. Thanasas and Georgia N. Kontogeorga, 'Fraud Detection in Banking Transactions with the Use of Artificial Intelligence and Anonymized Data' (2025) 24 Journal of European economy 656.

- Fairly
- transparently²²

Each of these three principles will be explored in the context of ML-based fraud detection in the following sections.

(a) Lawfully

Data is processed lawfully when one of the following applies:²³

- The data subject has given consent to the processing of the data for one or more specific purposes²⁴
- The processing is necessary for the performance of a contract to which the data subject is a party²⁵
- The processing is necessary to comply with a legal obligation²⁶
- The processing is necessary in order to protect the vital interests of the data subject or of another natural person²⁷
- The processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller²⁸
- The processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which

²² GDPR, art. 5 (1) (a).

²³ GDPR, art. 6 (1).

²⁴ GDPR, art. 6 (1) (a).

²⁵ GDPR, art. 6 (1) (b).

²⁶ GDPR, art. 6 (1) (c).

²⁷ GDPR, art. 6 (1) (d).

²⁸ GDPR, art. 6 (1) (e).

require protection of personal data, in particular where the data subject is a child²⁹

To see which of these lawful bases apply to ML-based fraud detection systems, it is important to make a distinction between private entities using ML-based fraud detection systems and public entities, as there are differences between the lawful bases that both types of entities tend to rely on.

§1. Private entities

Private entities rely on art. 6 (1) (f) GDPR as a lawful basis. This lawful basis allows the processing of personal data when the processing is necessary for the purposes of a legitimate interest pursued by the controller or a third party. The following conditions must be fulfilled for a private entity to be able to rely on art. 6 (1) (f) GDPR:

- The pursuit of a legitimate interest by the controller or by a third party
- The need to process personal data for the purposes of the legitimate interest pursued
- The interests or fundamental freedoms and rights of the concerned data subject do not take precedence over the legitimate interest of the controller or of a third party³⁰

Recital 47 of the GDPR indicates that fraud prevention can constitute a legitimate interest of a data controller. The EDPB considers that, in principle, fraud detection can be considered as a form of fraud prevention, since individuals who have committed fraud are at risk to reoffend and since identifying instances of fraud is necessary to uncover weaknesses in a fraud prevention system. Still, the EDPB considers that a case-by-case analysis is required to uncover if a fraud detection system can be seen as a fraud prevention system.³¹ In my view drawing a strict distinction between fraud detection and fraud prevention would be artificial. Any system that is designed to detect fraud inevitably

²⁹ GDPR, art. 6 (1) (f).

³⁰ European Data Protection Board, 'Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR', 1/2024 (8 October 2024), para 6.

³¹ European Data Protection Board, 'Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR', para. 102.

has preventive effects. For this reason, this thesis will not differentiate between fraud detection and fraud prevention systems.

As already mentioned, invoking fraud detection as a legitimate reason will not in itself be enough to rely on art. 6 (1) (f) GDPR as a lawful basis for the processing of personal data. The data controller must also demonstrate that the legitimate interest pursued overrides the fundamental rights of the data subject. This is a balancing exercise and will have to be judged on a case-by-case basis. Some aspects that have to be taken into consideration when doing this balancing exercise will be:

- The relationship between the data subject and the controller and the reasonable expectations of the data subject based on his relationship with the data controller³²
- The impact of the processing on data subjects³³
- The nature of the personal data that is processed³⁴
- The context of the processing³⁵
- The existence of additional safeguards which could limit the impact on data subjects (data minimization, privacy enhancing technology...)³⁶

The legitimate interest pursued by the data controller must be clearly articulated. Just a general reference to ‘fraud prevention’ will not suffice, otherwise it would be impossible to do the balancing exercise.³⁷ The controller therefore must be specific about the type of fraud that they are trying to prevent and what data they require for that purpose.

³² GDPR, rec. 47.

³³ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 32.

³⁴ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 32.

³⁵ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 32.

³⁶ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 56-57.

³⁷ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 17.

Furthermore, data controllers also need to consider if there are reasonable alternative ways to achieve the same legitimate interest. If there are reasonable and just as effective, but less intrusive ways to achieve the same legitimate interest, the processing will not be considered as necessary.³⁸

Art. 6 (1) (f) GDPR is not the only lawful basis that private entities are able to rely on. Private entities can also rely on art. 6 (1) (c) when the processing is necessary for compliance with a legal obligation to which the controller is subject. An example of such an obligation is the duty of obliged entities under the anti-money laundering regulation to monitor transactions.³⁹ The term ‘law’ in art. 6 (1) (c) refers both to European Union law and the law of a member state to which the controller is subject.⁴⁰

When a private entity has been entrusted with a task carried out in the public interest, art. 6 (1) (e) GDPR can also be a possible lawful basis.⁴¹ However, due to this lawful basis primarily being invoked by public entities as they are more often entrusted with a task carried out in the public interest, this lawful basis will be explored more in depth when discussing the lawful bases for public entities.

Other lawful bases appear significantly less suited for fraud detection. Art. 6 (1) (a) allows personal data processing when consent for this processing was given by the data subject. Consent should be given freely.⁴² Consent is presumed not to be given freely when the performance of a contract is dependent upon this consent, despite the consent not being necessary for the performance of the contract.⁴³ The EDPB considers that fraud detection will likely go beyond what is necessary for the performance of a contract with a data

³⁸ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para. 29.

³⁹ Regulation (EU) 2024/1624 of the European Parliament and of the Council of 31 May 2024 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing [2024] OJ L 19.6.2024 (‘AMLR’), art 26; Mariusz Krzysztofek, GDPR: Personal Data Protection in the European Union (Wolters Kluwer 2023) 91.

⁴⁰ GDPR, art 6 (3).

⁴¹ Krzysztofek (n 39), 90-91; ‘Article 6 GDPR’ (*GDPRhub*) <https://gdprhub.eu/index.php/Article_6_GDPR> accessed 31 March 2026.

⁴² GDPR, rec 32.

⁴³ GDPR, rec. 43.

subject.⁴⁴ This is also the reason why art. 6 (1) (b) GDPR is unable to be used as a lawful basis for fraud detection.⁴⁵ Furthermore, it is difficult to imagine someone planning to commit fraud to give consent when the consent is not required for the underlying contractual relationship to be fulfilled. Lastly, art. 6 (1) (d) will almost never be able to be used as a lawful basis for fraud detection. While fraud is a serious matter, being able to detect it is in most cases not a ‘vital’ interest. Data processing will be in the vital interest of the data subject when it is necessary to protect an interest which is essential for the life of the data subject or that of another natural person.⁴⁶ Only in exceptional circumstances will economic interests be considered as vital interests.⁴⁷

§2. Public entities

Public entities will mostly rely on art. 6 (1) (c) GDPR and art. 6 (1) (e) GDPR as lawful bases for the processing of personal data for fraud detection purposes. As explained above, art. 6 (1) (c) allows the processing of personal data when it is required to comply with a legal obligation. An example of a law requiring public entities to process personal data and therefore allowing them to rely on art. 6 (1) (c) GDPR is the Belgian law: ‘Wet van 3 augustus 2012 houdende bepalingen betreffende de verwerking van persoonsgegevens door de Federale Overheidsdienst Financiën in het kader van zijn opdrachten’.⁴⁸ art. 3 of this law determines that the Federal Public Service Finance is allowed to process personal data in order to execute its legal obligations.

Another lawful basis that public entities can use for the processing of personal data for fraud detection purposes is art. 6 (1) (e) GDPR. This article allows the processing of personal data when it is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. The difference between article 6 (1) (e) and art. 6 (1) (c) GDPR is subtle. Just like in art. 6 (1) (c) GDPR,

⁴⁴ European Data Protection Board, ‘Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects’, 2/2019 (8 October 2019), para. 50.

⁴⁵ European Data Protection Board, ‘Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects’, 2/2019 (8 October 2019), para. 50.

⁴⁶ GDPR, rec. 46.

⁴⁷ Krzysztofek (n 39), 88.

⁴⁸ Wet 3 augustus 2012 houdende bepalingen betreffende de verwerking van persoonsgegevens door de Federale Overheidsdienst Financiën in het kader van zijn opdrachten, *BS* 24 August 2012 (ed. 2).

the task carried out in the public interest needs to be laid down by either EU law or member state law.⁴⁹ The difference lies in the amount of specificity required. Under art. 6(1) (c) GDPR, the controller must rely on a clear, concrete legal obligation, while under art. 6 (1) (e) GDPR it suffices that the law entrusts the controller with a task in the public interest or the exercise of official authority.⁵⁰ Similar to art. 6 (1) (f) GDPR the processing of personal data under art. 6 (1) (e) GDPR is required to be necessary for the performance of the task carried out in the public interest. There is no lawful basis when there is a reasonable and less intrusive way to achieve the same result.⁵¹

Art. 6 (1) (a), (b) and (d) GDPR are unable to be used as lawful bases by public entities for the processing of personal data for the same reasons that these bases are unable to be used by private entities, as was discussed in chapter 2.1.2.1.1 of this thesis. Furthermore, the GDPR determines that public entities are unable to use art. 6 (1) (f) GDPR as a lawful basis.⁵²

(b) Fairness

Art. 5 (1) (a) of the GDPR obligates data controllers to process data fairly. While the concept of fairness has an important position in the GDPR, pinning down what it exactly means is difficult as both the European Court of Justice and the GDPR have given little indication of what fairness in the context of the GDPR means.⁵³ Fairness is mentioned in the GDPR in conjunction with lawfulness and transparency.⁵⁴ It is therefore often

⁴⁹ GDPR, art 6 (3).

⁵⁰ Krzysztofek (n 39), 90-91.

⁵¹ Case C-200/23 *Agentsia po vpisvaniyata v OL* EU:C:2024:827, para. 111; ‘Public Task’ (ICO) <<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/public-task/>> accessed 31 March 2026.

⁵² GDPR, art 6 (1).

⁵³ Paul Maynard and Madelaine Harrington, ‘It’s just not fair! The ambiguity of GDPR fairness principle’ [2024] *Privacy laws & business* 20, 20-21.

⁵⁴ Gianclaudio Malgieri, ‘The Concept of Fairness in the GDPR’ in Mireille Hildebrandt and others (eds.), *FAT* ’20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for

Computing Machinery 2020) 154-166; Andrea Palumbo, ‘The Unexplored Potential of the Fairness Principle under the GDPR: Lessons from the Recent TikTok Case’ (*CiTiP blog*, 7 December 2023) <<https://www.law.kuleuven.be/citip/blog/the-unexplored-potential-of-the-fairness-principle-under-the-gdpr-lessons-from-the-recent-tiktok-case/>> accessed 31 March 2026.

assumed that when data processing is carried out in compliance with the procedural rules of the GDPR, that that data processing will be in compliance with the fairness principle of the GDPR.

An example of this kind of reasoning can be found in a case of the Italian data protection authority against Foodinho.⁵⁵ Foodinho is a food delivery labour platform similar to Takeaway and DoorDash. In this case the Italian data protection authority did not use the fairness principle on a stand-alone basis, but instead determined that the infringement of the transparency principle of the GDPR also caused the fairness principle of the GDPR to be infringed due to the obligation to inform employees as part of an employer-employee relationship mirroring the general principle of fairness.⁵⁶

Some authors however propose that the GDPR should include a form of substantive fairness, as purely adhering to the procedural rules found in the GDPR can lead to unfair results.⁵⁷ An example of data processing that, while adhering to the procedural rules in the GDPR is often considered unfair, is personalized pricing. Personalized pricing can be described as “differentiating the online price for identical products or services partly based on information a company has about a potential customer”.⁵⁸ It is perfectly possible for a company to use personalized pricing and respect all procedural provisions of the GDPR,⁵⁹ while many consider such a practice unfair.

⁵⁵ Garante per la protezione dei dati personali, ‘Ordinanza n 234 del 10 giugno 2021’, docweb 9675440 (2021). English abstract can be found at: < <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9677611> > accessed 23 April 2026.

⁵⁶ Garante per la protezione dei dati personali, ‘Ordinanza n 234 del 10 giugno 2021’. English abstract can be found at : < <https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9677611> > accessed 23 April 2026.

⁵⁷ Atoosa Kasirzadeh and Damian Clifford, ‘Fairness and Data Protection Impact Assessments’ in Marion Fourcade and others (eds), *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI* (association for Computing Machinery 2021) 146; Sandra Wachter, Brent Mittelstadt and Chris Russell, ‘Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and ai’ (2021) 41 *Computer Law & Security Review* 1; Andreas Häuselmann and Bart Custers, ‘Substantive Fairness in the GDPR: Fairness Elements for Article 5.1A GDPR’ (2024) 52 *Computer Law & Security Review* (2024) 1.

⁵⁸ Frederik Z. Borgesius and Joost Poort, ‘Online Price Discrimination and EU Data Privacy Law’ (2017) 40 *Journal of Consumer Policy* 347, 348.

⁵⁹ Borgesius (n 58), 364.

The idea of substantive fairness has received some attention from the EDPB, as a reference to substantive fairness was made in the Guidelines 03/2022.⁶⁰ In these guidelines the EDPB defined fairness as “an overarching principle which requires that personal data shall not be processed in a way that is detrimental, discriminatory, unexpected or misleading to the data subject”.⁶¹ Furthermore, the concept of fairness being an independent principle seems to be accepted by the EDPB as in their guidelines on deceptive design patterns, it is determined that deceptive design patterns do not comply with the fairness principle regardless of compliance with other data protection principles.⁶² The concept of deceptive design patterns infringing the fairness principle independently was also accepted in a decision of the EDPB, where the EDPB again underlined that the fairness principle has an independent meaning and that compliance with the transparency principle does not automatically rule out non-compliance with the fairness principle.⁶³ Still, while the concept of substantive fairness has gained the attention of the legal literature, there is currently no agreement on what substantive fairness exactly entails. As will be shown when discussing the child benefit controversy, the concept of substantive fairness can have an impact on ML-based fraud detection systems, most notably when the system is discriminatory.⁶⁴

(c) Transparency

Transparency is an overarching principle of the GDPR that applies to three key areas,

- The provision of information to data subjects regarding fair processing

⁶⁰ European Data Protection Board, ‘Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them’, 03/2022 (14 February 2023).

⁶¹ European Data Protection Board, ‘Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them’, para 9.

⁶² European Data Protection Board, ‘Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them’, para 9.

⁶³ European Data Protection Board, ‘Binding Decision 2/2023 on the dispute submitted by the Irish SA regarding TikTok Technology Limited (Art. 65 GDPR)’, 2/2023 (2 August 2023), para. 100.

⁶⁴ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, Onderzoeksrapport z2018-22445 (2020), 47.

- The manner in which data controllers communicate with data subjects regarding their rights under the GDPR
- How data controllers facilitate the exercise by data subjects of their rights.⁶⁵

The GDPR tries to realize transparency by laying out certain specific requirements in art. 12-15 of the GDPR.⁶⁶ Information provided to data subjects should be concise, transparent, intelligible and plain language must be used.⁶⁷ A central consideration is that the data subject should be able to determine in advance what the scope and consequences of processing entail.⁶⁸ Most relevant for ML-based fraud detection systems is ‘the right to explanation’, recently recognized by the ECJ.⁶⁹ The ‘right to an explanation’ refers to the idea that when an automated system makes a decision, data subjects have the ‘right to explanation’ of that decision.⁷⁰ The explanation provided must describe the procedure and principles actually applied in such a way that the data subject can understand which of his or her personal data have been used in what way in the automated decision-making at issue.⁷¹

The impact of the ‘right to explanation’ is limited for ML-based fraud detection systems due to trade secrets and the nature of fraud detection. Fraud detection systems rely on secrecy to function effectively, as too much transparency could allow fraudsters to game the system.⁷² While the ECJ has addressed the trade secrecy issue by allowing a balance

⁶⁵ Article 29 Working Party, ‘Guidelines on transparency under Regulation 2016/679’, (29 November 2017), para. 1.

⁶⁶ Article 29 Working Party, ‘Guidelines on transparency under Regulation 2016/679’, para. 4.

⁶⁷ GDPR, art 12 (1).

⁶⁸ Article 29 Working Party ‘Guidelines on transparency under Regulation 2016/679’, para. 10.

⁶⁹ Case C-203/22 *CK v Dun & Bradstreet Austria and Magistrat der Stadt Wien* EU:C:2025:117.

⁷⁰ Elanor Taylor, ‘Explanation and the Right to Explanation’ (2023) 10 Journal of the American Philosophical Association 467, 468.

⁷¹ *CK v Dun & Bradstreet Austria and Magistrat der Stadt Wien* (n 69), para. 61.

⁷² Joshua A. Kroll, ‘Accountable Algorithms’ (PhD thesis, Princeton University 2015), 64-65; Christian Sandvig and others, ‘Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms’ [2014] Data and Discrimination: Converting Critical Concerns into Productive Inquiry 1, 9.

to be struck between data rights and trade secrets, offering as a solution the disclosure of trade secrets to the competent court or supervisory authority that then determines what and how much information should be shared with the data subject,⁷³ in practice trade secrets will still remain a barrier to a full right to explanation. Furthermore, it is still unclear how the ECJ will deal with the argument that transparency can help fraudsters game the system. Still, it is unlikely that this argument will allow entities to completely avoid the transparency requirements laid out in the GDPR, as a broad refusal to share any information about the processing of personal data would undermine the very essence of the data subjects 'right to explanation'. Moreover, the 'right to explanation' puts a heavy burden on data subjects as even ordinary requests demand an enormous amount of time and persistence.⁷⁴

3. Art. 22 GDPR

(a) Applicability

Art. 22 of the GDPR gives data subjects the right not to be subject to a decision based solely on automated processing, when this decision produces legal effects concerning the data subject or when this decision similarly significantly affects the data subject.

Essential for a decision to fall under the scope of art. 22 is that it is solely based on automated processing. Only systems where there is no human involvement fall under the scope of art. 22.⁷⁵ However, data controllers cannot circumvent art. 22 by fabricating human involvement. To qualify as human involvement, the controller must ensure that the human oversight over the decision is actually meaningful and that it is carried out by someone who has the competence and authority to change the decision.⁷⁶ While the requirement that a decision be made solely based on automated processing causes many ML-based fraud detection systems to be excluded from the scope of art. 22, as many of these systems are solely used to calculate risk scores which humans then use to make a

⁷³ *CK v Dun & Bradstreet Austria and Magistrat der Stadt Wien* (n 69), para. 67-76.

⁷⁴ Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security & Privacy 46, 51.

⁷⁵ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', (11 April 2018), 20.

⁷⁶ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', 21.

final decision, the SCHUFA case of the European court of justice demonstrates that in some cases such risk scoring can still fall under the scope of art. 22 GDPR.⁷⁷

In the SCHUFA-judgement, the central question asked was if the automated calculation of a credit score, which was subsequently used by a human decision-maker in assessing whether to grant a loan, could fall under art. 22 GDPR. Although the final decision to give a loan or not was still determined by a human, the German Federal Association of Consumer Organizations argued that the person making the decision was so highly influenced by the automatically calculated credit score that the decision was akin to being taken solely by the automated calculation of the credit score. The European Court of Justice accepted this functional approach and determined that art. 22 GDPR still applies if the credit score plays a decisive role in the decision-making process even though formally a human still intervened.⁷⁸

It is clear that the SCHUFA-judgement has an impact on the way that ML-based fraud detection systems will be treated under art. 22 GDPR. Even if an ML-based fraud detection system implements a human-in-the-loop and only uses the system as a risk-calculation tool, the system can still fall under the scope of art. 22 when the risk-score plays a decisive role in the decision-making process.

The SCHUFA case demonstrates that the application of Article 22 becomes significantly more complex once a system no longer fits a simple binary model of either human or automated decision-making. As has been pointed out by Reuben Binns and Michael Veale⁷⁹ systems making use of multi-layered or selective automation are a square peg in the round hole of art. 22 GDPR. When only one level in the decision-making process is automatized or only a specific aspect of the decision-making process, it becomes hard to determine if it is solely decided by automated processing or not. Multi-stage decision-making processes can be divided into decision-supporting systems and triaging systems. Supporting systems are used to provide information to a human decision-maker, while triaging systems are

⁷⁷ Joined Cases C-26/22 and C-64/22 *UF AB v Land Hessen* EU:C:2023:958.

⁷⁸ *UF AB v Land Hessen* (n 77), para. 73.

⁷⁹ Reuben Binns and Michael Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR' (2021) 11 International Data Privacy Law 319.

used to determine which cases get to a human decision maker and which cases get passed into another automated system.⁸⁰

ML-based fraud detection systems often make use of both triaging and decision-supporting systems. An example of an ML-based fraud detection system making use of a decision-supporting system is Connect of the HMRC.⁸¹ Connect is the ML-based fraud detection system used by the UK tax authorities. It relies on large databases to calculate fraud risk scores. With the support of those risk-scores a human makes the final decision.⁸² The Connect system will be discussed further later in the thesis when the case studies are discussed in chapter 3. ML-based fraud detection systems used by credit card companies often make use of triaging techniques. Only when the system flags a certain transaction as suspicious, a human will review the transaction.⁸³

The combination of triaging techniques and decision-supporting techniques in an ML-based fraud detection system can make it difficult to determine whether a decision is taken solely by a machine or not. A key problem in a multi-stage decision process is locating when the decision happened.⁸⁴ Where a triaging technique is used to determine whether a data subject will be subjected to further review, it is unclear whether this constitutes a decision in itself, or whether only the final fraud determination by the human reviewer should be considered as a decision. Davis and Schwemer argue that relevant to answer this question is the legal or similarly significant-effect criterion of art. 22 of the GDPR.⁸⁵

⁸⁰ Binns (n 79), 322.

⁸¹ Richard Dyson, 'What does the taxman know about you, your finances and your lifestyle?' *The Telegraph* (25 June 2015) <<https://www.telegraph.co.uk/finance/personalfinance/tax/11697816/What-does-the-taxman-know-about-you-your-finances-and-your-lifestyle.html>> accessed 1 April 2026.

⁸² Letter from Sir Robert Devereux and Jon Thompson to Meg Hillier (10 June 2016) <<https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf>> accessed 21 April 2026.

⁸³ Binns (n 79), 322; Imane Sadgali, Nawal Sael and Faouzia Benabbou, 'Human Behavior Scoring in Credit Card Fraud Detection' (2021) 10 *IAES International Journal of Artificial Intelligence (IJ-AI)* 698, 698.

⁸⁴ Binns (n 79), 325.

⁸⁵ Peter A. E. Davis and Sebastian F. Schwemer, 'Rethinking Decisions under Article 22 of the GDPR: Implications for Semi-Automated Legal Decision-Making' [2023] *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace* 1, 5.

They argue that when there is a human in the loop somewhere in the decision-making process the relevant question becomes whether the effects of previous automated decisions were significant enough to be considered under art. 22 GDPR.⁸⁶

The legal or significant effect criterion is a crucial element to determine if a decision falls under the scope of art. 22 and determines that a decision based on automated processing will only fall under the scope of art. 22 when that decision produces legal or similarly significant effects for the data subject. The term legal effects refers to every decision or action that affects someone's rights or legal status, or even their rights under a contract. Examples include but are not limited to: impacts on the right to vote, the right to receive a monthly pension allocation or the ability for someone to legally enter the country.⁸⁷

With the similarly significant effect criterion, art. 22 opens a broad scope of application, making it hard for data controllers to determine if they fall under the scope of this article. Similarly significant effects include those consequences that, although they do not necessarily have a legal impact on the lives of data subjects, still have a significant weight in their lives. The effect will be considered impactful enough if the use of automated processing leads to an individual finding themselves in need of protection due to the use of automated processing influencing their behavior, choices or circumstances.⁸⁸ Examples given by the recitals of the GDPR of decisions with similarly significant effects include the automatic refusal of an online credit card application and e-recruiting practices without human intervention.⁸⁹

Davis and Schwemer's argument to consider preliminary automated decisions as decisions under art. 22 of the GDPR and to focus on the significant criterion effect to truly determine in those cases if a decision falls under art. 22⁹⁰ is supported by Advocate General Pikamäe.⁹¹ Pikamäe argues that the absence of a formal definition of 'decision' in art. 22

⁸⁶ Davis (n 85), 7.

⁸⁷ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', 21; Sandra Barbosa and Sara Félix 'Algorithms and the GDPR: An analysis of article 22' [2021] *Anuário da Proteção de Dados*, 75.

⁸⁸ Barbosa (n 87), 76.

⁸⁹ GDPR, rec. 71.

⁹⁰ Davis (n 85), 7.

⁹¹ Joined Cases C-26/22 and C-64/22 *UF AB v Land Hessen* EU:C:2023:958, Opinion of AG

indicates that this concept should be constructed broadly and include a number of acts capable of affecting the data subject.⁹² The crucial element should not be if a decision was made or not, but instead the effect that the decision has on the data subject.⁹³

However, triaging techniques likewise complicate the assessment of whether a given outcome amounts to a similarly significant effect. Consider the example of an ML-based fraud detection system that uses a triaging mechanism to determine which accounts should be examined more closely by human reviewers. Even if we assume that the human reviewer does not merely rubber-stamp the decision of the triaging system as was the case in SCHUFA,⁹⁴ the flagged individual could argue that the outcome would not have occurred had the automated system not flagged their account, as no review would then have taken place. On this view, the automated triage would be seen as producing a significant effect.⁹⁵

However, it is my view that this would misconstrue the language of art. 22 GDPR, which only grants the right not to be subject to decisions based *solely* on automated processing. The final fraud determination is based both on the ML-based fraud detection system and the human reviewer, making the decision not *solely* based on automated processing. In my view the person being accused of fraud in my scenario would only be able to invoke art. 22 to attack the decision that determined that he would be checked by a human reviewer. Only when the triaging decision itself had enough of a significant effect would art. 22 apply. This would be the case, for example, if the triaging step did not merely trigger human review but also resulted in immediate consequences such as the freezing of an account.

(b) Exceptions to art. 22 GDPR

There are three exceptions to the right of data subjects not to be subject to a decision based solely on automated processing. A decision solely based on automated processing is allowed when:

Pikamäe.

⁹² *UF AB v Land Hessen*, Opinion of AG Pikamäe (n 91), point 38.

⁹³ *UF AB v Land Hessen*, Opinion of AG Pikamäe (n 91), point 43.

⁹⁴ *UF AB v Land Hessen* (n 77).

⁹⁵ Binns (n 79), 328.

- the automated processing is necessary to enter into, or perform the contract between the data subject and a data controller⁹⁶
- the automated processing is allowed by EU or member state law and if the law lays down suitable measures to safeguard the data subject's right, freedoms and legitimate interests⁹⁷
- the automated processing is based on the explicit consent of the data subject⁹⁸

At first sight it might seem that entities using ML-based fraud detection systems will rely on the exception that allows the use of automated processing when the processing is necessary to enter into a contract.⁹⁹ The Article 29 working party however links this exception to art. 6 (1) (b) GDPR.¹⁰⁰ Thus, just as for the purposes of art. 6 (1) (b) GDPR fraud detection is likely to go beyond what is objectively necessary for the performance of a contract with a data subject,¹⁰¹ fraud detection will likely not be considered as necessary in the context of art. 22 (2) (a) GDPR.

Furthermore, consent is also not a feasible exception for ML-based fraud detection to rely on. As is explained in chapter 2.1.2.1 of this thesis, consent has to be given freely¹⁰². Consent is not given freely when the performance of a contract is dependent upon the consent, despite the consent not being necessary for the performance of the contract.¹⁰³

ML-based fraud detection systems will however in many cases be able to rely on EU or member state laws to make decisions based on automated processing. Although most

⁹⁶ GDPR, art. 22 (2) (a).

⁹⁷ GDPR, art. 22 (2) (b).

⁹⁸ GDPR, art. 22 (2) (c).

⁹⁹ GDPR, art. 22 (2) (a) GDPR.

¹⁰⁰ Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', 23.

¹⁰¹ European Data Protection Board, 'Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects', para. 50.

¹⁰² GDPR, rec. 32.

¹⁰³ GDPR, rec. 43.

member states have not made use of this exception,¹⁰⁴ the member states that have made use of this exception have taken differing approaches. Some member states have adopted a sector-specific approach by authorizing automated data processing in particular domains,¹⁰⁵ while other member states have opted for a more general, procedural model that permits automated processing provided certain safeguards are met.¹⁰⁶ An example of a more sector-specific approach can be found in Germany, where the insurance sector can use automated decision-making when it is used in the context of providing services.¹⁰⁷ An example of a general approach can be found in the Netherlands,¹⁰⁸ where art. 22 of the GDPR can be circumvented if the processing is necessary for the fulfillment of a task of general interest.¹⁰⁹

EU law that allows for the automated processing of data includes the AMLR.¹¹⁰ This regulation allows entities who fall under the scope of the AMLR to adopt decisions resulting from automated processing when this is strictly necessary for the purposes of preventing money laundering and terrorist financing, provided that:

- The data is limited to data obtained pursuant to chapter III of the AMLR¹¹¹
- any decision to enter or refuse to enter into or maintain a business relationship with a customer or to carry out or refuse to carry out an occasional transaction for a customer, or to increase or decrease the extent of the customer due diligence measures applied pursuant to Article 20 of the AMLR, is subject to meaningful

¹⁰⁴ Gianclaudio Malgieri, 'Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations' (2019) 35 Computer Law & Security Review 1, 7-9.

¹⁰⁵ Malgieri (n 104), 7-9.

¹⁰⁶ Malgieri (n 104), 7-9.

¹⁰⁷ Federal Data Protection Act of 30 June 2017 (Federal Law Gazette I p. 2097); Malgieri (n 104), 7.

¹⁰⁸ Uitvoeringswet Algemene verordening gegevensbescherming.

¹⁰⁹ Uitvoeringswet Algemene verordening gegevensbescherming, art. 40.

¹¹⁰ AMLR, art. 76 (5).

¹¹¹ AMLR, art. 76 (5) (a).

human intervention to ensure the accuracy and appropriateness of such a decision¹¹²

- the customer may obtain an explanation of the decision reached by the obliged entity, and may challenge that decision, except in relation to a report as referred to in Article 69 of this Regulation¹¹³

The impact of the exception to art. 22 of the GDPR provided by the AMLR is especially relevant for ML-based fraud detection systems used by the payments sector. This will be explored further in chapter 3.2.1.2 of this thesis.

B. AI Act

The AI Act is widely considered as the world's first comprehensive AI legislation.¹¹⁴ The purpose of the AI Act is to lay down a uniform legal framework about the development, placing on the market and use of AI systems that ensures a high level of protection of health, safety and other fundamental rights, while also supporting innovation.¹¹⁵ The AI Act came in force on the 1st of august 2024. It was planned that 36 months after the entry into force, on the 2 of august 2026, most of the provisions of the AI Act would be enforced.¹¹⁶ However, backlash against many provisions of the AI Act has caused the European Commission to reconsider this enforcement date via the Digital Omnibus Proposal. The Digital Omnibus proposal¹¹⁷ wants to improve the competitiveness of Europe by

¹¹² AMLR, art. 76 (5) (b).

¹¹³ AMLR, art. 76 (5) (c).

¹¹⁴ Council of the EU 'Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI' (European Council Council of the European Union 21 May 2024) <<https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-AI-Act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/>> accessed 21 April 2026.

¹¹⁵ AI Act, art. 1; AI Act, rec. 1; Martin Ebers 'Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act' (2025) 16 European Journal of Risk Regulation 684, 685.

¹¹⁶ AI Act, art. 113.

¹¹⁷ European Commission, 'Proposal for a regulation of the European Parliament and of the Council amending Regulations (EU) 2016/679, (EU) 2018/1724, (EU) 2018/1725, (EU) 2023/2854 and Directives 2002/58/EC, (EU) 2022/2555 and (EU) 2022/2557 as regards the simplification of the digital legislative framework, and repealing Regulations (EU) 2018/1807, (EU) 2019/1150, (EU) 2022/868, and Directive (EU) 2019/1024 (Digital Omnibus)' COM(2025) 837 final ('Digital Omnibus proposal').

simplifying, solidifying and clarifying the effectiveness of the rules applicable to the ICT sector and their enforcement.¹¹⁸ The hope is that this would lead to compliance with the rules coming at a lower cost, while still maintaining high standards and the agreed objectives.¹¹⁹ Recently, the European Parliament has accepted the postponing of certain rules governing high-risk AI systems.¹²⁰ As the AI Act is, at the time of writing, not yet applied in practice, there is no relevant case law, making its exact impact difficult to assess.

As will be discussed in the following chapter, ML-based fraud detection systems will often be considered as AI systems under the AI Act.¹²¹ An analysis of the impact of the AI Act on ML-based fraud detection systems is thus highly relevant for the purposes of this thesis.

1. Applicability

The scope of the AI Act is broad, as it applies not only to the providers of AI systems, but also to deployers of AI systems, importers, distributors, product manufacturers and authorized representatives of AI systems.¹²² Almost every actor in the value chain of AI-development will therefore be subject to the AI Act.

An AI system is defined by art. 3 of the AI Act as: ‘a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments’.¹²³ This very broad definition encompasses any software, system or device that autonomously processes data to produce outputs.¹²⁴

¹¹⁸ Digital Omnibus proposal, 3.

¹¹⁹ Digital Omnibus proposal, 2.

¹²⁰ European Parliament ‘Amendments adopted by the European Parliament on 26 March 2026 on the proposal for a regulation of the European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI)’, A10-0073/2026.

¹²¹ AI Act, art. 3 (1).

¹²² AI Act, art. 2 (1).

¹²³ AI Act, art. 3 (1).

¹²⁴ Slobodan Tomić and Vid Štimac ‘Pinning down an octopus: towards an operational definition of AI systems in the EU AI Act’, [2025] *Journal of European Public Policy* 1, 13.

The guidelines on the definition of artificial intelligence in the AI Act¹²⁵ divide the definition of AI systems¹²⁶ into the following seven elements: (1) a machine-based system; (2) designed to operate with varying levels of autonomy; (3) that may exhibit adaptiveness after deployment; (4) and that, for explicit or implicit objectives; (5) infers, from the input it receives, how to generate outputs (6) such as predictions, content, recommendations, or decisions (7) that can influence physical or virtual environments.¹²⁷

Almost all fraud detection systems will fulfill elements (1), (2), (4), (6) and (7). The term machine-based refers to the fact that AI systems are developed with and run on machines.¹²⁸ The term machine includes both the hardware and software components that enable an AI system to function.¹²⁹ Evidently, all modern fraud detection systems are developed and work on machines.

Most fraud detection systems will fulfill element (2). The AI Act does not require AI systems to be fully autonomous to fall under the scope of the AI Act. However, at least some degree of independence of actions from human involvement and some capabilities to operate without human intervention is required.¹³⁰ The exact threshold of autonomy required remains unclear. It is evident that systems designed to operate solely through full manual human control will be excluded from the scope of the AI Act.¹³¹ The guidelines describe the required level of autonomy as a reasonable degree of independence from human actions.¹³² While fraud detection systems rarely make final determinations without human intervention, they do independently analyze data and generate risk scores or flag

¹²⁵ European Commission, ‘Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’ COM(2025) 5053 final.

¹²⁶ AI Act, art. 3 (1).

¹²⁷ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 9.

¹²⁸ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 11.

¹²⁹ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 11.

¹³⁰ AI Act, rec. 12.

¹³¹ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 17.

¹³² European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 20.

potentially fraudulent activity without real-time human involvement. The detection process therefore operates largely autonomously.

Furthermore, fraud detection systems always have a clear objective, namely the detection of fraud, regardless of whether the developer has made this objective explicit or not.¹³³ Fraud detection systems also always generate predictions. A prediction is an estimate of an unknown value (the output) from known values supplied to the system (the input).¹³⁴ This is exactly what fraud detection systems do as they make predictions about the risk that a certain transaction or entity is fraudulent based on the input that they have received. It is also clear that fraud detection systems influence the environment in which they are deployed.¹³⁵ Being flagged by a fraud detection system will, at a minimum, trigger additional investigation.

However, not all fraud detection systems will fulfill elements (3) and (5). Purely rule-based fraud systems do not exhibit adaptiveness after deployment. Adaptiveness refers to self-learning capabilities, which allows AI systems to change while being deployed.¹³⁶ While a fraud detection system not exhibiting adaptiveness is not automatically excluded from the AI-definition in the AI Act,¹³⁷ due to the use of the word *may* in the definition, the presence of adaptiveness is a factor in favor of considering a system as AI.

The decisive distinction for excluding many rule-based fraud detection systems from the scope of the AI Act lies in element (5). The guidelines help us determine what is meant with *infer*. *Infer* refers both to inference during the building phase of AI and inference during the use phase of AI.¹³⁸ This means that an AI system may still qualify as an AI system under the AI Act even if it does not perform inference on input data during its use, provided that inference was used during its development.

¹³³ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 24.

¹³⁴ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 54.

¹³⁵ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 60.

¹³⁶ AI Act, rec. 12.

¹³⁷ AI Act, art. 3 (1).

¹³⁸ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 28.

What exactly constitutes ‘inferring’ remains difficult to determine. The guidelines suggest that systems using ‘basic’ rules will fall outside the scope of the AI Act, whereas systems based on complex rules may fall within the scope of the AI Act, even if the rules are solely defined by humans.¹³⁹ The AI Act does not define this term explicitly, leaving an important role for the European and national courts to determine the exact scope of this element.¹⁴⁰

Recital 12 of the AI Act gives two examples of AI-techniques that enable interference.¹⁴¹ The first category of AI-techniques are machine learning approaches that learn from data how to achieve certain objectives.¹⁴² These machine learning approaches include techniques, such as supervised learning, unsupervised learning, self-supervised learning, reinforcement learning and deep learning.¹⁴³ This makes it clear that ML-based fraud detection systems will thus fulfill element (5), as ML-based fraud detection systems clearly make use of machine learning approaches.

The other category of AI-techniques mentioned by recital 12 of the AI Act are logic-and knowledge-based approaches. The difference between machine learning approaches and logic-and knowledge-based approaches lies in their source of learning. While machine-based approaches learn from data, AI systems being developed via logic-and knowledge-based approaches learn from knowledge including rules, facts and relationships encoded by human experts. Based on this knowledge, these systems reason via deductive or inductive engines or use operations including sorting, searching, matching and chaining. By using logical interference to draw conclusions, these systems apply formal logic,

¹³⁹ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 26-51.

Rosie Nance, Marcus Evans and Peter McBurney ‘The Commission’s guidelines on AI systems – what can we infer?’ (Norton Rose Fulbright 13 February 2025) <<https://www.dataprotectionreport.com/2025/02/the-commissions-guidelines-on-AI-systems-what-can-we-infer/>> accessed 21 April 2026.

¹⁴⁰ Kristof Cox, Dries Rooses and Thierry Wuyts ‘Een juridische verkenning van de AI Act’ (2024) 88 *Rechtskundig Weekblad* 283, 285.

¹⁴¹ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 30-39.

¹⁴² European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 32.

¹⁴³ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 32-38.

predefined rules or ontologies to new relations.¹⁴⁴ Some rule-based fraud detection systems may thus be considered as performing inference where they do not merely execute predefined condition-action rules, but instead derive new conclusions through the structured application of rules, facts, and relationships. However, as this thesis focuses on ML-based fraud detection systems this will not be further explored.

2. Risk-level.

The AI Act divides AI systems in 4 risk categories based on the risk-level that they represent. A higher risk-level will require an AI system to fulfill more stringent requirements. The first category includes systems that represent an unacceptable risk-level.¹⁴⁵ Systems that represent an unacceptable risk-level are never allowed to be used. Art. 5 of the AI Act lays out which systems are considered to fall under this risk-level. The AI systems falling under this risk-level are forbidden due to being particularly harmful, abusive and these practices contradicting EU values of respect for human dignity, freedom, equality, democracy, and the rule of law, as well as fundamental rights enshrined in the Charter of Fundamental Rights of the European Union.¹⁴⁶ Most commonly used ML-based fraud detection systems will not be considered to fall under the scope of this risk-level.

The second risk-category includes systems that represent a high-risk.¹⁴⁷ There are two ways for an AI system to fall under this high-risk category. Firstly, an AI system will be considered as high-risk when they are products or safety components of products covered by EU harmonization legislation.¹⁴⁸ Annex I of the AI Act provides an exhaustive list of the relevant EU harmonization legislation for this purpose. Secondly, an AI system will be considered as high risk when it is used in one of the areas mentioned in annex III of the AI Act.¹⁴⁹ As the legislation mentioned in Annex I of the AI Act will rarely be relevant

¹⁴⁴ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 39.

¹⁴⁵ AI Act, art. 5.

¹⁴⁶ AI Act, rec. 28.

¹⁴⁷ AI Act, chapter III.

¹⁴⁸ AI Act, art. 6 (1).

¹⁴⁹ AI Act, art. 6 (2); Hannah Ruschemeier and Jascha Bareis, ‘Searching for Harmonised Rules: Understanding the Paradigms, Provisions, and Pressing Issues in the Final EU AI

for ML-based fraud detection systems,¹⁵⁰ the classification of such systems as high-risk will primarily depend on whether they fall within one of the use cases listed in Annex III.

Art. 6 (3) determines that an AI system referred to in annex III of the AI Act will not be considered high-risk, when it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision-making.

However, most ML-based fraud detection systems will not be able to rely on the exception provided by art. 6 (3) of the AI Act, as an AI system referred to in annex III will always be considered a high-risk AI system if it engages in the profiling of natural persons.¹⁵¹ Profiling in the context of the AI Act has to be interpreted consistently with the GDPR.¹⁵² The GDPR defines profiling as: “any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements”.¹⁵³ Most ML-based fraud detection systems will engage in profiling as defined by GDPR, as they involve automated processing of personal data to analyze and predict a natural person’s likelihood of engaging in fraudulent behavior.¹⁵⁴

The third risk-category includes systems that represent a transparency risk.¹⁵⁵ ML-based fraud detection systems will rarely fall under this category, as these systems rarely

Act’ in Rita Gsenger and Marie-Therese Sekwenz (eds), *Digital Decade: How the EU shapes digitalization* (Nomos 2025) 55.

¹⁵⁰ An ML-based fraud detection system may nevertheless be considered as a high-risk AI system via annex I of the AI Act, if the ML-based fraud detection system can be considered as a safety component of a product covered by the EU harmonization legislation see: art. 3 (14) and 6 (1) (a) AI Act. However, the sectors covered by EU harmonization legislation listed in annex I generally do not make significant use of ML-based fraud detection systems, limiting the practical relevance of this pathway.

¹⁵¹ AI Act, art. 6 (3).

¹⁵² AI Act, art. 3 (52).

¹⁵³ GDPR, art. 4 (4).

¹⁵⁴ Klaus Wiedemann ‘Profiling and (automated) decision-making under the GDPR: A two-step approach’ (2022) 45 Computer Law & Security review 1, 3.

¹⁵⁵ AI Act, Chapter IV.

interact directly with natural persons. Only when an ML-based fraud detection system engages directly with natural persons, will it fall under this category.¹⁵⁶

When an AI system does not fall under the scope of any of the other risk-categories, it will be considered as falling under the scope of the fourth risk-category. The AI Act does not include any obligations for these systems.¹⁵⁷

To determine the impact of the AI Act on ML-based fraud detection systems, it is thus necessary to focus primarily on the high-risk category, as this constitutes the most legally consequential classification. For the purposes of this thesis, it is useful in this context to analyze public and private ML-based fraud detection systems separately.

(a) Public ML-based fraud detection

AI systems are classified under the AI Act as high-risk based on their specific functions within specific sectors.¹⁵⁸ While the purpose of an ML-based fraud detection system will not differ, as it is always used to detect fraud, the sector where the system is utilized does have an impact on the risk-level that will apply to the AI system. The best illustrated examples of ML-based fraud detection systems used by public entities can be found within the social welfare sector and the tax sector.¹⁵⁹

Public ML-based fraud detection systems used to detect fraud within the social welfare sector will be considered as high-risk AI systems when they fall under art. 5 (a) annex III of the AI Act. Art. 5 (a) annex III determines that AI systems used to evaluate the eligibility of natural persons for essential public assistance benefits and services, including healthcare services, as well as to grant, reduce, revoke, or reclaim such benefits and services are to be considered as high-risk.

¹⁵⁶ AI Act, art. 50 (1).

¹⁵⁷ European Commission, 'Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)', COM(2025) 5052 final, para. 2; Seliha Buelens, Wout Platteau and Wouter Torfs, 'AI-Verordening. Inzoomen op de eerste verplichtingen' [2026] *Nieuw Juridisch Weekblad* 2, 3.

¹⁵⁸ David Hadwick, 'slipping Through the Cracks, the Carve-outs for AI Tax Enforcement Systems in the EU AI Act' (2024) 9 *European Papers* 936, 944.

¹⁵⁹ OECD, *Governing with Artificial Intelligence The State of Play and Way Forward in Core Government Functions* (OECD Publishing 2025) 30.

As will be shown by the child benefit scandal case study an ML-based fraud detection system flagging someone as fraudulent can be considered as a basis to revoke certain benefits or services. Recital 58 of the AI Act gives as examples of essential public assistance benefits and services healthcare services, social security benefits and social services that provide protection in cases like maternity, illness, retirement, loss of employment, housing and industrial accidents. An ML-based fraud detection system that causes the loss of a public assistance benefit and/or service will thus be considered as an high-risk AI system.

For tax authorities using ML-based fraud detection systems art. 5 (a) annex III of the AI Act will typically not cause their system to be considered as high-risk due to these systems not having an impact on the decision to revoke any public benefits as this is often the responsibility of social security institutions or welfare agencies.¹⁶⁰

Instead, what is relevant to determine if ML-based fraud detection systems used by tax authorities are to be considered as high-risk AI systems is art. 6 Annex III of the AI Act. Specifically, art. 6 (e) of annex III is relevant in the context of ML-based fraud detection used by tax authorities. This article determines that AI systems intended to be used by or on behalf of law enforcement authorities or by EU institutions, bodies, offices or agencies in support of law enforcement authorities for the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of the detection, investigation or prosecution of criminal offences are considered as high-risk AI systems.

Art. 3 (4) of the Law Enforcement Directive¹⁶¹ determines that profiling means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location or movements. This is the

¹⁶⁰ In some jurisdictions the tax sector is tasked with disbursing social benefits. Still the division between uses within the social welfare sector and the tax sector makes sense as even when both tasks are executed by the same administrative entity, both actions involve different policy objectives and are treated differently under the AI Act.

¹⁶¹ Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L 119/89 (' Law Enforcement Directive').

same definition as the definition of profiling under the GDPR.¹⁶² As was discussed in chapter 2.2.2 of this thesis, most ML-based fraud detection systems will engage in profiling.

The recitals of the AI Act provide that administrative proceedings by tax and customs authorities as well as by financial intelligence units carrying out administrative tasks analyzing information pursuant to EU anti-money laundering law should not be classified as high-risk AI systems used by law enforcement authorities for the purpose of prevention, detection, investigation and prosecution of criminal offences.¹⁶³

While the idea behind a distinction between administrative and criminal proceedings is understandable as criminal sanctions are typically graver than administrative sanctions, tax law does not necessarily follow this logic, with certain administrative fines reaching several million euros.¹⁶⁴ Furthermore, recital 38 of the AI Act seems to assume that a different ML-based fraud detection system is used for inconsistencies that lead to a penal sanction and for inconsistencies that lead to administrative sanctions. However ML-based fraud detection systems are used to detect non-compliance. A later investigation then determines if the non-compliance will be punished via an administrative sanction or a penal one.¹⁶⁵ The same ML-based fraud detection system is thus used both to detect fraud that leads to administrative sanctions and fraud that leads to penal sanctions. The *SS SIA*¹⁶⁶ case of the ECJ leads me to believe that when an ML-based fraud detection system is used both in the context of administrative proceedings and in the context of penal proceedings, the system will not be labelled as high-risk.

In *SS SIA* the court had to determine whether the GDPR or the Law Enforcement Directive should apply to the processing of personal data by tax authorities if it is not yet clear if this processing happens under the administrative tax supervision capacity or for criminal law enforcement purposes.¹⁶⁷ The GDPR is not applicable to the processing by competent authorities of personal data for the purposes of the prevention, investigation,

¹⁶² GDPR, art. 4 (4)

¹⁶³ AI Act, rec. 59.

¹⁶⁴ Hadwick (n 158), 948-949.

¹⁶⁵ Hadwick (n 158), 952.

¹⁶⁶ Case C-175/20 *SS' SIA v Valsts ieņēmumu dienests* EU:C:2022:124.

¹⁶⁷ *SS' SIA v Valsts ieņēmumu dienests* (n 166) , para. 30-47.

detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.¹⁶⁸ The reason for this exception lies in the fact that the Law Enforcement Directive was specifically created to protect personal data used in this context.¹⁶⁹ Thus when personal data is processed by public authorities for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security, the Law Enforcement Directive applies, while if personal data is processed for other purposes the GDPR applies.¹⁷⁰

The court determined that only when data is collected for the specific purpose of prevention, investigation, detection or prosecution of criminal offences, the Law Enforcement Directive is applicable, whereas if the data is collected for another purpose or for mixed purposes the GDPR will apply.¹⁷¹ The Court thus adopted a strict reading of law enforcement purposes, requiring a clear and specific criminal law orientation at the moment of processing.

As in *SS SIA* it was decided by the court that when it is unclear if personal data will be used by public authorities for administrative or criminal proceedings, the GDPR, rather than the Law Enforcement Directive should apply, it makes sense to expand this interpretation to art. 6 (e) of annex III. If mixed-purpose processing defaults to the GDPR rather than the Law Enforcement Directive, then mixed-purpose AI systems used by tax authorities should likewise default to the general AI Act regime rather than the high-risk category reserved for law enforcement tools. Consequentially, , ML-based fraud detection systems used by tax authorities thus will rarely be considered as high-risk systems.

(b) Private sector ML-based fraud detection systems

As the best illustrated examples of private sector ML-based fraud detection systems are found within the payments sector and within labour platforms, these are the sectors that will be explored in this chapter.

¹⁶⁸ GDPR, art. 2 (2) (d).

¹⁶⁹ GDPR, rec. 19; *SS' SIA v Valsts ieņēmumu dienests* (n 166), para. 42.

¹⁷⁰ Law enforcement directive, rec. 11.

¹⁷¹ *SS' SIA v Valsts ieņēmumu dienests* (n 166), para. 45

ML-based fraud detection systems used in the payments sector will rarely be classified as high-risk AI systems. Recital 58 of the AI Act provides that AI systems used for the purpose of detecting fraud in the provision of financial services should not be considered high-risk where their use is provided for under EU law. Important regulations that are relevant in this context are the payment service directive 2 (PSD2)¹⁷² and the AMLR.

Contrary to ML-based fraud detection systems used in the payments sector, ML-based fraud detection systems used by labour platforms to check if their workers are committing fraud will often be considered as high-risk AI systems. Directive (EU) 2024/2831¹⁷³ determines that a certain service is a labour platform when:

- The service is provided, at least in part at distance via electronic means, such as by means of a website or mobile application¹⁷⁴
- The service is provided at the request of the recipient of the service¹⁷⁵
- The service involves as a necessary and essential component the organization of work performed by individuals in return for payment, irrespective of whether that work is performed online or in a certain location¹⁷⁶
- The service involves the use of automated monitoring systems or automated decision-making systems.¹⁷⁷

¹⁷² Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on payment services in the internal market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC [2015] OJ L 337/35 ('PSD2').

¹⁷³ Directive (EU) 2024/2831 of the European Parliament and of the Council of 23 October 2024

on improving working conditions in platform work [2024] OJ L 11.11.2024 ('Platform-work directive').

¹⁷⁴ Platform-work directive, art. 2 (1) (a) (i).

¹⁷⁵ Platform-work directive, art. 2 (1) (a) (ii).

¹⁷⁶ Platform-work directive, art. 2 (1) (a) (iii).

¹⁷⁷ Platform-work directive, art. 2 (1) (a) (iv).

ML-based fraud detection systems used by labour platforms to detect fraud committed by their workers fall under art. 4 (b) of annex III of the AI Act. Art. 4 (b) of annex III of the AI Act determines that “AI systems intended to be used to make decisions affecting terms of work-related relationships, the promotion or termination of work-related contractual relationships, to allocate tasks based on individual behavior or personal traits or characteristics or to monitor and evaluate the performance and behavior of persons in such relationships” are considered as high-risk AI systems. Recital 57 of the AI Act confirms that platform workers will be considered as having a work-related contractual relationship with the labour platform.

As will be illustrated in the case study on Uber in chapter 3.2.2 of this thesis, ML-based fraud detection systems targeting platform workers can influence the allocation of tasks and may even result in exclusion from the platform. Such a system will thus clearly fall under the scope of art. 4 (b) annex III of the AI Act.

3. Consequences falling under high risk:

In this section a quick overview will be given of the obligations applicable to high-risk AI systems. The exact impact that these obligations have on ML-based fraud detection systems will be explored when the case studies are discussed.

The use of high-risk AI systems is not prohibited by the AI Act. However, the admissibility of these systems depends on the fulfilment of the requirements for high-risk AI systems found in the AI Act.¹⁷⁸ The most important of these requirements are:

- the obligation to create a risk-management system¹⁷⁹ which includes the obligation to test these systems¹⁸⁰
- The obligation to ensure that the data used by the AI system meets certain quality standards¹⁸¹

¹⁷⁸ Dariusz Szostek and Rafał T. Prabucki, ‘High-Risk AI systems’ in Vera L. Raposo (ed), *The European Artificial Intelligence Act Promises and Perils?* (Springer 2025) 166.

¹⁷⁹ AI Act, art. 9 (1).

¹⁸⁰ AI Act, art. 9 (6).

¹⁸¹ AI Act, art. 10.

- the documentation obligations¹⁸²,
- the obligation to have a possibility of effective human oversight¹⁸³
- the obligation to ensure the accuracy, robustness and cybersecurity of these systems.¹⁸⁴

The risk management system has to be understood as a continuous iterative process planned and run throughout the entire life cycle of an AI system. It requires regular systematic review and updating.¹⁸⁵ Only those risks that can be reasonably mitigated or eliminated through the development or design of the high-risk AI system, or via the provision of adequate technological information need to be considered.¹⁸⁶ The development of the risk management system is the responsibility of the provider of the system (i.e. the entity who developed the AI system).¹⁸⁷ The development of the risk management system goes hand in hand with the testing of the AI system as testing is required to identify the most appropriate and targeted risk management measures.¹⁸⁸

The obligation to ensure that the data used by the AI system meets certain quality standards¹⁸⁹ requires activities such as training, validation and testing to be conducted on data sets that meet the quality criteria laid out in paragraph 2 to 5 of that article. The most relevant of these quality criteria for ML-based fraud detection systems are the

¹⁸² AI Act, art. 11-12.

¹⁸³ AI Act, art. 14 AI Act.

¹⁸⁴ AI Act, art. 15; Szostek (n 178), 177.

¹⁸⁵ AI Act, art. 9 (2).

¹⁸⁶ AI Act, art. 9 (3).

¹⁸⁷ AI Act, art. 3 (3); AI Act, art. 16 (a).

¹⁸⁸ AI Act, art. 9 (6).

¹⁸⁹ AI Act, art. 10.

quality criteria about data collection processes,¹⁹⁰ formulations of assumptions¹⁹¹ and bias.¹⁹²

The documentation obligations require technical documentation to be drawn up,¹⁹³ logs to be kept of events over the lifetime of the AI system¹⁹⁴ and for providers to ensure that their system is sufficiently transparent to enable deployers to interpret a systems' output and use it appropriately.¹⁹⁵ The idea behind these documentation requirements is to address concerns related to the opacity of AI systems and allow deployers to understand the functionality, strength and limits of AI systems.¹⁹⁶ While these documents are not publicly available, many of the documents created to comply with the documentation obligations have to be at the disposal of the national competent authorities¹⁹⁷, thus enabling transparency, while still taking into account trade secret considerations and the risk of fraudsters 'gaming' the system.

The AI Act also requires high-risk AI systems to be developed in such a way that they can effectively be overseen by humans during the period in which they are in use.¹⁹⁸ Furthermore, Recital 73 of the AI Act explains that human oversight should be assigned to natural persons who have the necessary competence, training and authority to carry out that role. The AI system also needs to include mechanisms to guide and inform a natural person to whom human oversight has been assigned so as to enable them to make informed decisions.¹⁹⁹

Lastly, the AI Act requires AI systems to be developed so as to achieve an appropriate level of accuracy, robustness and cybersecurity.²⁰⁰ Art. 15 (2) of the AI Act determines that

¹⁹⁰ AI Act, art. 10 (2) (e).

¹⁹¹ AI Act, art. 10 (2) (d).

¹⁹² AI Act, art. 10 (2) (f) (g).

¹⁹³ AI Act, art. 11.

¹⁹⁴ AI Act, art. 12.

¹⁹⁵ AI Act, art. 13.

¹⁹⁶ AI Act, rec. 72.

¹⁹⁷ AI Act, art. 18 (1).

¹⁹⁸ AI Act, art. 14 (1).

¹⁹⁹ AI Act, rec. 73.

²⁰⁰ AI Act, art. 15 (1).

the Commission shall encourage the development of benchmarks and measurement methodologies as to address the appropriate levels of accuracy and robustness required. Many of these standards are still under development, making it hard for providers to understand what these standards entail.²⁰¹

III. Case studies

A. Public sector use of ML-based fraud detection systems

In this section, the public sector use of ML-based fraud detection systems will be explored. Two case studies will be discussed: The Dutch ‘Toeslagenaffaire’ (in this thesis called the child benefit scandal) and the Connect system developed and used by the UK tax authorities. These two case studies fall together with the sectors in which the public sector use of ML-based fraud detection is best illustrated, namely the social welfare sector and the tax sector.

1. Dutch Child benefit scandal

The child benefit scandal was a major controversy in the Netherlands in which many parents were falsely accused of fraud by the Dutch tax authorities. These false accusations had severe consequences for the affected parents. This was not only because false accusations led to the loss of child benefits, but also because they often resulted in parents being required to repay benefits they had previously received.²⁰² The impact of this scandal was and continues to be substantial. At this moment 43.702 parents have been recognized as having been wrongfully accused of fraud and this number is still rising every week.²⁰³ Furthermore, the Dutch child benefit scandal led to the highest fine ever imposed by the Dutch data protection authority amounting to 2,750,000 euro.²⁰⁴

²⁰¹ Roberto Tamponi and others, ‘Lost in Vagueness: Towards Context-Sensitive Standards for Robustness Assessment under the EU AI Act’ (unpublished, 19 November 2025), 2-3.

²⁰² ‘Feiten en cijfers’ (Dienst Toeslagen Ministerie van Financiën 3 April 2026) <<https://herstel.toeslagen.nl/dashboard-kinderopvangtoeslag/>> accessed 21 April 2026.

²⁰³ ‘Feiten en cijfers’ (Dienst Toeslagen Ministerie van Financiën 3 April 2026) <<https://herstel.toeslagen.nl/dashboard-kinderopvangtoeslag/>> accessed 21 April 2026.

²⁰⁴ Autoriteit persoonsgegevens, ‘boete belastingdienst kinderopvangtoeslag’ (Autoriteit persoonsgegevens 7 December 2021) <<https://www.autoriteitpersoonsgegevens.nl/documenten/boete-belastingdienst-kinderopvangtoeslag>> accessed 31 March 2026; Friederike van der Jagt-Vink, ‘Toeslagenaffaire: Autoriteit Persoonsgegevens beboet Belastingdienst’ (Mr. Online 29

While the scandal was caused by an interplay of different improper government actions²⁰⁵ including a disproportionate approach to administrative errors,²⁰⁶ non-transparency towards parents being accused of fraud²⁰⁷ and discrimination,²⁰⁸ the wrongful use of an ML-based fraud detection system partly caused and exacerbated many of these improper government actions. The ML-based fraud detection system used by the Dutch tax authorities for the detection of child benefit fraud was a self-learning ‘risk-classification model’.²⁰⁹ The risk-classification model was trained on examples of correct and incorrect child benefit applications. The model calculated a risk score between 0-1 where 0 was the lowest risk and 1 the highest.²¹⁰ The model was supposed to function as a first filter, as the 100 applications with the highest risk-score were checked by humans on correctness.²¹¹ However, many officials hastily marked files labelled as high-risk as fraudulent without exercising the necessary scrutiny needed for a high-impact decision like this.²¹² Furthermore investigations by the Dutch data protection authority have unveiled that the model considered not being a Dutch citizen as a factor that contributed towards being labelled as high-risk, which caused immigrant and minority families to be disproportionately targeted by the risk-classification model.²¹³

In the following sections, the application of the GDPR on this risk-classification model will be explored. Thereafter, the impact that the AI Act would have had on the risk-

December 2021) < <https://www.mr-online.nl/toeslagenaffaire-krijgt-ook-privacyrechtelijk-staartje-autoriteit-persoonsgegevens-beboet-belastingdienst/> > accessed 21 April 2026.

²⁰⁵ Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (Amnesty International 2021) 5.

²⁰⁶ Hadwick (n 14), 612-614.

²⁰⁷ Hadwick (n 14), 617

²⁰⁸ Hadwick (n 14), 619-625.

²⁰⁹ Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (Amnesty International 2021) 5-7.

²¹⁰ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 14-15.

²¹¹ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 15.

²¹² Sofia Ranchordás and Luisa Scarcella, ‘Automated Government for Vulnerable Citizens: Intermediating Rights’ (2021) 30 William & Mary Bill of Rights Journal 373, 398.

²¹³ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 16-18.

classification model if at the time of the child benefit scandal it would have been in force will be explored.

(a) GDPR

The use of the risk-classification model by the Dutch tax authorities raises several important questions under the GDPR. In particular, about the existence of a valid lawful basis, the fairness of the data processing, the applicability of art. 22 of the GDPR and transparency. The decision of the Dutch data protection authority provides valuable insight into how some of these GDPR principles are applied in the context of ML-based fraud detection systems in the public sector.

Firstly, The Dutch data protection authority held that the processing of personal data by the Dutch tax authorities could not be justified by relying on art. 6 (1) (e) of the GDPR.²¹⁴ While the Dutch tax authority were entrusted by law with the implementation of the granting, payment and recovery of the childcare allowance²¹⁵ and thus the Dutch tax authority was entrusted with a public task, the processing of nationality was not necessary to combat fraud.²¹⁶ While some processing of nationality-like data was held to be necessary, as nationality has an impact on the granting of child benefits²¹⁷, simply allowing non-Dutch citizenship to be seen as a risk-factor was too excessive.²¹⁸ As was described in chapter 2.1.2.1.2 of this thesis, entities can only rely on art. 6.1 (e) GDPR as a lawful basis when there exists no reasonable less intrusive way to achieve the same result.²¹⁹ The Dutch data protection authority held that a more accurate result could be reached by processing data that responds to the question if someone possesses the Dutch nationality, or EU nationality with registration in a Dutch municipality, or non-EU nationality with a

²¹⁴Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, page 44.

²¹⁵ Regeling Wet Kinderopvang, art. 1.3.

²¹⁶ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 42-44.

²¹⁷ Vreemdelingenwet 2000, art. 10.1: People without the Dutch nationality can only be granted childcare benefits if they have a residence permit

²¹⁸ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 43.

²¹⁹ *Agentsia po upisvaniyata v OL* (n 51); ‘Public Task’ (ICO) <<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/public-task/>> accessed 31 March 2026.

valid residence permit. This would decrease the chance of nationality having too large of an impact on the acting of the Dutch tax authorities and would be a less far-reaching form of personal-data processing.²²⁰

Furthermore, the Dutch data protection authority held that the processing of nationality for the purposes of the risk-classification model was unfair due to being discriminatory and thus infringed art. 5.1 (a) of the GDPR.²²¹ While the Dutch data protection authority did not develop a comprehensive theory of fairness under the GDPR, the Dutch data protection authority did hold that the fairness principle of the GDPR entails in any case that the processing should not be contrary to fundamental rights, like the right not to be discriminated against.²²² The Dutch data protection authority determined that the processing of nationality was discriminatory here, as there was no objective justification for using ‘non-Dutch nationality’ as a risk-factor.²²³

Interesting here is that the Dutch data protection authority thus seemed to utilize a form of substantive fairness, as the infringement of fairness is assessed independently from the infringement of lawfulness of the data processing. Still, the Dutch data protection authority avoids truly defining substantive fairness, only determining that at the very least substantive fairness means that data processing cannot be contrary to fundamental rights.²²⁴

Lastly, the Dutch data protection authority determined that art. 22 of the GDPR was not applicable, as it held that the decisions in this case were not solely taken via automated

²²⁰ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 43.

²²¹ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 50.

²²² Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 47-48. The non-discrimination principle was held to be a fundamental right, due to it being recognized in art. 26 of the international covenant of civil and political rights, art. 1 of protocol 12 of the European convention of human rights, art. 1 of the Dutch constitution and art. 1 of the International convention on the elimination of all forms of racial discrimination.

²²³ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 49-50.

²²⁴ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 47.

processing.²²⁵ It is true that the risk-classification model used by the Dutch tax authorities did not make any final decisions, as humans reviewed applications flagged by the risk-classification model.²²⁶ However, it is my view that art. 22 of the GDPR could potentially have applied and that, at the very least, a deeper analysis of the applicability of art. 22 of the GDPR by the Dutch data protection authority was warranted here. Especially as it has been suggested that many tax officials did not use the necessary scrutiny needed when reviewing child benefit applications flagged by the risk-classification model and thus almost mechanically approved the results of the risk-classification model.²²⁷

While it is true that art. 22 of the GDPR is only applicable to decisions made solely by automated processing, not just any form of human involvement suffices to make art. 22 of the GDPR inapplicable. To qualify as human involvement, human involvement needs to be meaningful.²²⁸ The Dutch data protection authority should thus in my opinion have investigated the effectiveness of the human involvement and should not have determined that art. 22 was inapplicable by simply determining that somewhere in the decision-making process a human was involved. As was mentioned in chapter 2.1.3.1 of this thesis, the SCHUFA²²⁹ case of the ECJ supports my argument, as in this case the ECJ determined that an automatically calculated risk-score used by humans to make a decision regarding a data subject can still fall under the scope of art. 22 GDPR, if the calculated risk-score was strongly drawn upon.²³⁰

Furthermore, the flagging by the risk-calculation score was used to filter out which child benefit applications should be investigated and thus can be considered as a triaging technique. As was pointed out in chapter 2.1.3.1 of this thesis, the application of art. 22 becomes complicated when triaging techniques are being used. I defended the idea of the triaging decision having to be considered separately from the final decision. In my view,

²²⁵ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 42.

²²⁶ Autoriteit Persoonsgegevens, ‘Belastingdienst/toeslagen De verwerking van de nationaliteit van aanvragers van kinderopvangtoeslag’, 42.

²²⁷ Ranchordás (n 212), 398.

²²⁸ Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’, 21.

²²⁹ *UF AB v Land Hessen* (n 77).

²³⁰ *UF AB v Land Hessen* (n 77), para. 73.

the Dutch data protection authority should thus have analyzed if the triaging decision itself caused enough of a significant effect to fall under art. 22 GDPR.

Strangely, the Dutch data protection authority did not discuss transparency in the context of the risk-classification model. However the black-box nature of the risk-classification model was one of the reasons that the child benefit controversy was able to endure for so long.²³¹ While it remains unclear what the exact impact of the transparency principle of the GDPR is on ML-based fraud detection systems, as it remains unclear how much data controllers can rely on the argument that transparency could facilitate fraudsters, at least a minimum form of transparency should apply to ML-based fraud detection systems. In a case closely resembling the child benefit scandal in which the Dutch government also used an ML-based fraud detection system to identify welfare fraud, a Dutch court of first instance determined that the government should provide some insight in the risk indicators and underlying norms used. The court, however, did not base themselves on the GDPR, but on the right on private and family life found in art. 8 of the European convention of human rights.²³² It is thus my view that the Dutch data protection should have analyzed the transparency principle in the context of the risk-classification model used by the Dutch tax authorities.

(b) AI Act

At the time of the child benefit scandal, the AI Act was not yet in force. It is nonetheless interesting to examine whether the AI Act would have been capable of preventing the risks and consequences associated with the risk-classification system used by the Dutch tax authorities, as this allows us to assess whether the AI Act is sufficient to address the risks posed by ML-based fraud detection in the social welfare sector.

First off, it seems likely that the risk-classification tool would fall under the scope of the AI Act. As was determined in chapter 2.2.1 of this thesis, 7 elements are relevant to determine if a system falls under the scope of the AI Act. To fall under the scope of the AI Act, the system needs to be (1) a machine-based system; (2) that is designed to operate with varying levels of autonomy; (3) that may exhibit adaptiveness after deployment; (4) and that, for explicit or implicit objectives; (5) infers, from the input it receives, how to generate outputs (6) such as predictions, content, recommendations, or decisions (7) that

²³¹ Hadwick (n 14), 627.

²³² Rb. Den Haag 5 februari 2020, ECLI:NL:RBDHA:2020:1878.

can influence physical or virtual environments.²³³ For fraud detection systems elements (2), (3) and (5) are especially relevant, as the other elements can be assumed to almost always be fulfilled when using a fraud detection system.

The risk-classification model used by the Dutch tax authorities fulfills elements 2, 3 and 5. Firstly, the risk-classification model calculated the risk score completely autonomously. Furthermore, the system was able to adapt throughout its lifespan by using the output data it generated as new input data to learn from.²³⁴ Lastly, as the risk-classification model used a form of machine learning, it will be considered as fulfilling element 5, as machine learning approaches are explicitly mentioned in recital 12 of the AI Act as enabling inference.

Furthermore, as the risk-classification model was used to revoke and reclaim essential public benefits, it would be considered as a high-risk AI system under art. 5 (a) of annex III of the AI Act. The risk-classification model would thus have to comply with the obligations laid out for high-risk AI systems by the AI Act. While it is of course impossible to perfectly predict what the impact of these obligations would have been on the child benefit controversy, it is likely that the obligations around documentation²³⁵, data²³⁶ and risk-management²³⁷ potentially would have been able to prevent this controversy. These obligations would probably have led to greater transparency from the start and potentially could have significantly shortened the investigation by the Dutch data protection authority, limiting the impact of the controversy.²³⁸ While the GDPR already prohibited the use of nationality in the risk-classification model, the AI Act would have allowed this use to be discovered earlier, limiting the impact of this discriminatory system.

²³³ European Commission, ‘Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)’, para. 9.

²³⁴ Hadwick (n 14), 620.

²³⁵ AI Act, art. 11-12.

²³⁶ AI Act, art. 10.

²³⁷ AI Act, art. 9 (1).

²³⁸ Hadwick (n 158), 947-948.

(c) Conclusion

The Dutch child benefit scandal serves as a good case study for the risks inherent in the use of ML-based fraud detection systems within the welfare sector.

Firstly, the risk-classification model demonstrated the impact that false positive detection can have. False fraud detection in the child benefit scandal not only led to parents losing their child benefits, but also in certain cases to parents having to pay some of the benefits received in the past back to the government.²³⁹

Secondly, the risk-classification model exemplified the risk of discrimination. An ML-based fraud detection system risks either intentionally or unintentionally to make use of discriminatory criteria. While the use of nationality in the child benefit scandal was intentional,²⁴⁰ bias in ML-based fraud detection systems can also arise unintentionally, due to self-learning systems being trained on historical data.²⁴¹

The use of discriminatory criteria can be exacerbated by the third risk exemplified by the system used in the child benefit scandal, namely automation bias. Automation bias can be defined as “the human inclination to give undue weight to recommendations or decisions generated by automated systems, even in situations where the information provided by those systems is incorrect or where human judgement would be more accurate”.²⁴² In the child benefit scandal this automation bias can be considered as an explanation as to why it seems that officials did not use the necessary scrutiny on the results of the risk-management system. While the risk-classification model was supposed to function as a filter, some officials used it as a de facto decision maker. Automation bias can thus lead to a system that merely incorporates discriminatory criteria as a filtering tool to produce discriminatory outcomes in practice, as it is effectively treated as the final decision-maker.

²³⁹ ‘Feiten en cijfers’ (Dienst Toeslagen Ministerie van Financiën 3 April 2026) <<https://herstel.toeslagen.nl/dashboard-kinderopvangtoeslag/>> accessed 21 April 2026.

²⁴⁰ Hadwick (n 14), 621.

²⁴¹ Moses Alabi, ‘Bias and Fairness in AI-Based Fraud Detection: An Ethical Assessment in Financial Applications’ (unpublished manuscript 9 September 2023), 3.

²⁴² Hannah Ruschemeier and Lukas J. Hondrich, ‘Automation bias in public administration – an interdisciplinary perspective from law and psychology’ (2024) 41 Government Information Quarterly 1, 1.

Finally, this case study shows that non-transparency can allow for violations of the rights of data subjects to persist for a long time, as a non-transparent system makes it impossible for data subjects to determine when their rights are infringed.²⁴³

2. HMRC ‘Connect’

The Connect system is an ML-based fraud detection system developed by the UK tax authorities in 2015.²⁴⁴ It is a data-matching risk-management tool using machine learning.²⁴⁵ This means that the Connect system works by matching data from many different sources in order to detect any inconsistencies or suspicious patterns. The data used by Connect includes tax data from the UK tax authorities, data from UK government databases, data from banks and financial institutions, data gathered from international tax exchange mechanisms and open-source social and e-commerce data like for example public information found on E-bay, Etsy ²⁴⁶ It is currently estimated that the Connect-system has access to 55 billion items of data.²⁴⁷

Furthermore, the Connect-system uses machine learning. By using machine learning, Connect is able to uncover new patterns that help it predict the likelihood that a person or business might be committing fraud or is non-compliant. Similar to the ML-based fraud detection system used by the Dutch tax authorities, the Connect-system itself does not

²⁴³ Samanta Schweblin, Mariana Enríquez and Valeria Luiselli, ‘Explainable AI for Transparency in Automated Fraud Detection Systems’ (unpublished manuscript 6 December 2025), 2.

²⁴⁴ Richard Dyson ‘What does the taxman know about you, your finances and your lifestyle?’ The Telegraph (25 June 2015) <<https://www.telegraph.co.uk/finance/personalfinance/tax/11697816/What-does-the-taxman-know-about-you-your-finances-and-your-lifestyle.html>> accessed 21 April 2026.

²⁴⁵ Stephen Daly, ‘Artificial intelligence, the Rule of Law and public administration: the case of taxation’ (2024) 83 The Cambridge Law Journal 437, 450.

²⁴⁶ Helen Wood, ‘What is HMRC’s Connect AI System?’ (TaxAssist Accountants 4 November 2025) <<https://www.taxassist.co.uk/resources/articles/what-is-hmrc-s-Connect-AI-system>> accessed 21 April 2026.

²⁴⁷ Kumal Natwhani, Artificial intelligence in automated decision-making in tax administration: the case for legal, justiciable and enforceable safeguards (The institute for Fiscal Studies 2024) 8.

make any final decisions. It merely attaches a risk-level to persons or businesses. Based on this risk level, human tax experts will investigate and make a final decision.²⁴⁸

Data seems to suggest that leads from the Connect-system have helped to generate an additional 4.6 billion pounds in tax revenue.²⁴⁹ Furthermore, some believe that the Connect-system has led to a decrease in the tax gap.²⁵⁰ The tax gap is the difference between the tax amount that theoretically should be paid to the government and the tax amount that the tax authorities have actually collected.²⁵¹ In the 2023-2024 tax-year the tax gap was estimated to be 5.3 percent.²⁵² In the tax year 2005-2006 which is before the use of the Connect system, the tax gap was estimated to be around 7.4 percent.²⁵³ Currently the Connect-system prompts 90 percent of the investigations done by the UK tax authorities.²⁵⁴

The Connect-system has, however, not been without criticism. Concerns have been raised about the impact of this system on privacy.²⁵⁵ Especially the use of public internet data, like social media posts has been controversial. While UK tax investigators already used public internet data before the use of the Connect-system in their investigations, the use of AI to collect social media data in criminal investigations constitutes a significantly

²⁴⁸ Letter from Sir Robert Devereux and Jon Thompson to Meg Hillier (10 June 2016) <<https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf>> accessed 21 April 2026.

²⁴⁹ Jay Sanghrajka, 'HMRC's Connect computer and investigations' [2020] Taxation <<https://www.taxation.co.uk/articles/hmrc-s-Connect-computer-and-investigations>> accessed 21 April 2026.

²⁵⁰ Daly (n 245), 450; 'Got Secrets? HMRC "Connect-System" Targets Tax Evaders Now!' (eClear 10 May 2023) <<https://eclear.com/article/got-secrets-hmrc-Connect-system-targets-tax-evaders-now/>> accessed 21 April 2026.

²⁵¹ Sanghrajka (n 249)

²⁵² HM Revenue & Customs, 'Official Statistics 1. Tax gaps: Summary' (GOV.UK 19 June 2025) <<https://www.gov.uk/government/statistics/measuring-tax-gaps/1-tax-gaps-summary>> accessed 21 April 2026.

²⁵³ HM Revenue & Customs, 'Official Statistics 1. Tax gaps: Summary' (GOV.UK 19 June 2025) <<https://www.gov.uk/government/statistics/measuring-tax-gaps/1-tax-gaps-summary>> accessed 21 April 2026.

²⁵⁴ Sanghrajka (n 249).

²⁵⁵ Daly (n 245), 450.

greater intrusion of privacy than traditional case by case data collection done by human investigators.²⁵⁶

The limited amount of information about the inner workings of the Connect-system has also been subject to numerous criticisms, with some calling the Connect-system a black-box system.²⁵⁷ While the Connect system has not been at the forefront of any major scandal like the risk-classification tool used by the Dutch tax authorities in the child benefit scandal, the non-transparency of the UK tax authorities about the Connect-system might lead to biases or other problematic practices being hidden from the public.

(a) GDPR

While the GDPR as it exists in Europe is not applicable to the United Kingdom, the GDPR was enacted before the withdrawal of the UK out of Europe and is therefore considered as a valid UK law.²⁵⁸ While the UK is under no obligation to follow rules of the GDPR after Brexit, the UK has continued to comply with many of the provisions of the GDPR post-Brexit, via the UK GDPR.²⁵⁹ Still, there have been no major cases against the UK tax authorities for data protection infringements concerning the Connect-system. This does however not mean that the Connect system is not infringing the UK GDPR as the opacity of the system could lead to infringements being hidden.²⁶⁰ This section will analyze the Connect-system under the EU GDPR and not under the UK GDPR to allow for comparison between the different case studies. Still, an analysis under the UK GDPR would show

²⁵⁶ Jack Simpson, 'Eight ways that the taxman is spying on you' *The Times* (6 September 2025) < <https://www.thetimes.com/money/family-finances/article/eight-ways-that-the-taxman-is-spying-on-you-7ds87gs3q>> accessed 21 April 2026.

Tom Gerken, 'HMRC admits using AI to scan social media for tax cheats' *BBC* (12 August 2025) < <https://www.bbc.com/news/articles/cqjyedz202ko>> accessed 21 April 2026.

²⁵⁷ Daly (n 245), 450.

²⁵⁸ Karen McCullagh, 'Post-Brexit data protection in the UK - leaving the EU but not EU data protection law behind' in Gloria González, Rosamunde Van Brakel, and Paul De Hert (eds) *Research handbook on privacy and data protection law* (Edward Elgar 2022) 37-38.

²⁵⁹ Regulation (EU) 2016/679 as retained in UK law (UK GDPR) pursuant to the European Union Withdrawal Act 2018, s 3; Karen McCullagh (n 258), 41.

²⁶⁰ Sylvia Lu, 'Data Privacy, Human Rights, and Algorithmic Opacity' (2022) 110 *California Law Review* 2087, 2103.

little difference as until now there has been no substantial divergence between the UK GDPR and the EU GDPR.²⁶¹

As was mentioned in chapter 2.1.2.1.1 the public sector will either rely on art. 6 (1) c or e of the GDPR as a lawful basis to process personal data. In a policy document, the UK tax authorities have laid out four potential lawful bases that they can rely on when they engage in the processing of personal data, namely art. 6 (1) (a), (b), (c) and (e).²⁶² However, these refer to the lawful bases for all types of processing of personal data done by the UK tax authorities and not solely to the processing of personal data by the Connect-system. As was laid out in chapter 2.1.2.1.1 of this thesis, art. 6 (1) (a) and (b) are unable to be used as a lawful basis for data processing for the purpose of fraud detection. The only lawful bases that thus remain are art. 6(1) (c) and (e) of the GDPR.

For the Connect-system, the most reliable lawful basis is art. 6 (1) e of the GDPR, as there is no specific legal obligation for the UK tax authority to utilize a fraud detection system. A combination of provisions from the taxes management act²⁶³ and various finance acts²⁶⁴ make it clear that the UK tax authority is given by law the task to detect tax fraud, thus the UK tax authorities are able to use art. 6 (1) e GDPR as a lawful basis for the processing of personal data, as long as they keep the data processed limited to what is necessary.

The UK tax authority does not have to comply with the transparency obligations found in art. 13-15 of the GDPR, due to the data protection act exempting risk-assessment systems used for taxation purposes.²⁶⁵ Both the UK GDPR and the GDPR allow for legislators to restrict the scope of certain obligations of the GDPR when this is necessary and proportionate to safeguard certain public interests.²⁶⁶ This essentially allows Connect to function as a black-box without infringing the transparency principle of the GDPR.

²⁶¹ Karen McCullagh, 'post-Brexit UK Data Protection: Staying the Course or Charting a New Direction?' in Edoardo Celeste and others (eds) *Data Protection and Digital Sovereignty Post-Brexit* (Bloomsbury Publishing 2023), 24-25.

²⁶² HM Revenue & Customs 'Guidance HMRC appropriate policy document' (GOV.UK 8 May 2019) < <https://www.gov.uk/government/publications/hmrc-appropriate-policy-document>> accessed 21 April 2026.

²⁶³ Taxes Management Act 1970, ss 8, 9 and 18; Finance Act 2008, sch 36; Finance Act 2011, sch 23; Finance Act 2021, sch 36 (UK).

²⁶⁴ Finance Act 2008, sch 36; Finance Act 2011, sch. 23; Finance Act 2021, sch. 36.

²⁶⁵ Data Protection Act 2018, Sch 2, Pt 1, para 2.

²⁶⁶ Art. 23 GDPR and art. 23 UK GDPR.

Moreover, although Connect does not appear to make final decisions, its function as a risk-scoring and prioritization tool raises important questions under Article 22 GDPR. Even if no formal “decision” is taken solely by automated means, the practical effect of risk scoring may significantly influence whether individuals are selected for investigation, audit, or further scrutiny. This raises the question whether such systems should be considered as producing “legal or similarly significant effects” within the meaning of Article 22 GDPR.

(b) AI Act

If we make abstraction of the fact that the AI Act does not apply to the UK, it is very likely that the Connect-system will be considered as an AI system under the AI Act for many of the same reasons that the risk-classification tool used by the Dutch tax authorities would likely be considered as an AI system under the AI Act. The Connect-system calculates risk-scores autonomously, is adaptable and infers, due to engaging in machine-learning techniques.

As was described above in chapter 2.2.2.1 ML-based fraud detection systems used by tax authorities will likely not be considered as high-risk AI systems under the AI Act, due to AI-tools used by tax authorities for administrative proceedings not being classified as high-risk AI systems used by law enforcement authorities for the purpose of prevention, detection, investigation and prosecution of criminal offences.²⁶⁷ The Connect-system will thus not be considered as a high-risk AI system.

Furthermore, the Connect-system will also not be considered under the third risk-category of the AI Act as it does not directly engage with natural persons.²⁶⁸ The Connect-system will thus most likely fall under the fourth risk-category of the AI Act. As AI systems falling under this risk-category are not subject to specific obligations under the AI Act.²⁶⁹

²⁶⁷ AI Act, rec. 59.

²⁶⁸ AI Act, art. 50.

²⁶⁹ European Commission, ‘Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)’, para. 2.

(c) Conclusion

The Connect-system demonstrates that the use of ML-based fraud detection systems by public entities can be highly effective, as the available data seems to indicate that the Connect-system is currently playing an important role in the fight against tax fraud.²⁷⁰

Connect is not the only ML-based fraud detection system used by a tax authority to detect fraud that has generally been considered a success. The United States Internal Revenue Service (IRS) has been able to achieve a 40 percent increase in fraudulent return detection rates between 2020-2023 thanks to the integration of machine learning algorithms.²⁷¹ ML-based fraud detection systems thus seem to be an effective tool in the fight against tax fraud.

The use of ML-based fraud detection systems by tax authorities to detect tax fraud is however not without risk. The Connect-system case study exemplified the transparency risks that such a system can have. However, the risks are not limited to just transparency. The risks discussed in the child benefit controversy, namely discrimination²⁷² and automation bias²⁷³ are also relevant when using ML-based fraud detection systems to detect tax fraud. These risks exist partly because the safeguards provided by the GDPR may be restricted through national legislation,²⁷⁴ and partly because ML-based fraud detection systems used by tax authorities are likely to fall within the fourth risk category of the AI Act. As already mentioned, the AI Act does not contain any regulations for the fourth risk category.

²⁷⁰Sanghrajka (n 249); Daly (n 245), 450; HM Revenue & Customs, 'Official Statistics 1. Tax gaps: Summary' (GOV.UK 19 June 2025) <<https://www.gov.uk/government/statistics/measuring-tax-gaps/1-tax-gaps-summary>> accessed 21 April 2026; 'Got Secrets? HMRC "Connect-System" Targets Tax Evaders Now!' (eClear 10 May 2023) <<https://eclear.com/article/got-secrets-hmrc-Connect-system-targets-tax-evaders-now/>> accessed 21 April 2026.

²⁷¹ Kehinde O. Ariyibi and others, 'Leveraging Artificial Intelligence for enhanced tax fraud detection in modern fiscal systems' [2024] GSC Advanced research and reviews 129, 131.

²⁷² Ariyibi (n 271) 132.

²⁷³ Ruschemeier (n 242), 1.

²⁷⁴ GDPR, art. 23.

B. Private sector use of ML-based fraud detection

This part will discuss the use of ML-based fraud detection systems by private entities. To ground the discussions of the use of ML-based fraud detection systems by private entities in reality, the following real-life uses will be discussed: PayPal fraud detection advanced²⁷⁵, Zelle²⁷⁶ and the fraud detection system used by Uber.²⁷⁷ I chose these case studies, as these are the best illustrated examples of private ML-based fraud detection systems. Furthermore, they fall together with two very important sectors in which ML-based fraud detection systems are used, namely the payments industry and the labour platform sector.

It is however difficult to uncover the exact workings of ML-based fraud detection systems used by the private sector, due to these algorithms often being protected via trade secrets.²⁷⁸ Furthermore, publicly available information about ML-based fraud detection systems is often limited to what the developing companies choose to disclose, which tends to lack the critical scrutiny that would be necessary to understand the workings of ML-based fraud detection systems. The narrative about these systems is therefore controlled by the creators of these algorithmic systems.²⁷⁹ However, as this thesis is not trying to reconstruct the precise technical architecture of these systems the opacity of the exact workings of these systems does not prevent a legal analysis of the risks and benefits of these systems. This opacity does, however, explain why the legal literature surrounding

²⁷⁵ Paypal Editorial Staff, 'Harnessing machine learning fraud detection technologies' (Paypal.com 5 November 2024) <<https://www.paypal.com/us/brc/article/payment-fraud-detection-machine-learning>> accessed 21 April 2026.

²⁷⁶ David Krause, 'The Future of Consumer Financial Protection: Lessons from the Zelle Lawsuit Dismissal' (unpublished manuscript, 6 March 2026).

²⁷⁷ Edwin Ng and others, 'Project RADAR: Intelligent Early Fraud Detection System with Humans in the Loop' (Uber Blog 31 January 2022) <<https://www.uber.com/tw/en/blog/project-radar-intelligent-early-fraud-detection/>> accessed 26 April 2026.

²⁷⁸ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2016) 193; Sylvia Lu, 'Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence' (2020) 23 Vanderbilt Journal of Entertainment and Technology Law 99, 116; Marietje Schaake, 'Trade secrets shouldn't shield tech companies' algorithms from oversight' (Brookings Institution 2020) <<https://www.brookings.edu/articles/trade-secrets-shouldnt-shield-tech-companies-algorithms-from-oversight/>> accessed 21 April 2026.

²⁷⁹ Pasquale (n 278), 193.

fraud detection tends to focus on public sector use as more information about the workings of public-based ML-based fraud detection systems tends to be available.

1. Payment sector

ML-based fraud detection systems are a very important component of companies providing payment services as it is often their responsibility to implement sufficient safeguards to prevent fraud.²⁸⁰ Losses of payment card issuers and merchants encumbered by fraudulent transactions have been rising constantly, with the total losses from fraudulent bank card transactions worldwide more than doubling between 2012-2017 from USD 11.27 billion²⁸¹ to USD 22.8 billion.²⁸² Fraud prevention and fraud detection has therefore truly become a priority for private entities offering payment services.²⁸³

ML-based fraud detection systems used by the payment industry should not only be able to prevent credit card fraud, but also need to work as efficiently as possible to ensure customer satisfaction.²⁸⁴ Furthermore, it is crucial to avoid false positive fraud detection, as a false fraud detection can have significant financial ramifications.²⁸⁵ Some experts even estimate that the cost of a false fraud accusation can be more costly than the cost of the fraud itself.²⁸⁶ A study researching the impact of false positive fraud detection in the context of credit card fraud detection in the United States, uncovered that one in six of all

²⁸⁰ Art. 95 (1) PSD2.

²⁸¹ 'Nilson Report 1023' (Nilson Report August 2013) <[1023 - Nilson Report](#)> accessed 21 April 2026.

August 2013

²⁸² 'Nilson Report 1118' (Nilson Report October 2017) <[1118 - Nilson Report](#)> accessed 21 April 2026.

²⁸³ Caroline Hroncich, 'Why Fraud Resolution is Critical to Building Customer Trust' (The Financial Brand 27 March 2025) <<https://thefinancialbrand.com/news/customer-experience-banking/why-fraud-resolution-is-critical-to-building-customer-trust-187612>> accessed 21 April 2026.

²⁸⁴ Anna Nesvijevskaia, 'The accuracy versus interpretability trade-off in fraud detection model' (2021) 3 Data & Policy e12-1, e12-2.

²⁸⁵ Roy Wedge and others 'Solving the "false positives" problem in fraud prediction Automated Data Science at an Industrial Scale' in Ulf Brefeld and others (eds), *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part III* (Springer 2019) 372.

²⁸⁶ Florian Wallny 'False Positives in Credit Card Fraud Detection: Measurement and Mitigation' in Tung X. Bui (ed) *Proceedings of the 55th Annual Hawaii International Conference on System Sciences* (HICSS Conference Office 2022) 1572.

legitimate cardholders have experienced at least one decline due to suspected fraud in 2014 resulting in a total of 118 billion dollar declined.²⁸⁷ 26 percent of these customers whose credit card is declined reduce the amount that they spend on the website of the merchant that declined their credit card. 32 percent stop shopping on the merchant's website entirely.²⁸⁸ The pressure on payment service providers to develop fraud detection systems that are capable of accurately identifying fraudulent transactions while minimizing false positives is therefore particularly high, as these systems must strike a delicate balance between effective fraud prevention and the preservation of a smooth and reliable user experience.

(a) Paypal and Zelle

PayPal uses ML-based fraud detection systems to enable fraud teams to conduct in-depth risk analysis and investigations to help identify and mitigate fraud.²⁸⁹ Thanks to data gathered over two decades from 434 million active accounts and 1.68 trillion in annual payment volume PayPal has access to a very strong foundation to create sophisticated AI-models.²⁹⁰

Although PayPal has access to extensive data, challenges remain in effectively addressing issues related to false positive fraud detections. False positive fraud detection has been the cause of multiple lawsuits in the US against PayPal, as false positive fraud detection

²⁸⁷ Kyle Marchini and Aleia Van Dyke, 'Overcoming False Positives: Saving the Sale and the Customer Relationship' (Javelin 21 September 2015) <<https://javelinstrategy.com/research/overcoming-false-positives-saving-sale-and-customer-relationship>> accessed 21 April 2026.

²⁸⁸ Kyle Marchini and Aleia Van Dyke, 'Overcoming False Positives: Saving the Sale and the Customer Relationship' (Javelin 21 September 2015) <<https://javelinstrategy.com/research/overcoming-false-positives-saving-sale-and-customer-relationship>> accessed 21 April 2026.

²⁸⁹ 'Fraud Protection Advanced' (Paypal Developer 28 August 2024) <<https://developer.paypal.com/docs/checkout/advanced/customize/fraud-protection/fraud-protection-advanced/>> accessed 21 April 2026.

²⁹⁰ Dany Kitishian 'Paypal's AI Strategy: Analysis of Dominance in FinTech AI' (Klover.AI 4 August 2025) <<https://www.klover.ai/paypal-ai-strategy-analysis-of-dominance-in-fintech-ai/>> accessed 21 April 2026.

often leads to the freezing of accounts.²⁹¹ Furthermore, in the US PayPal has also been confronted with law suits for not detecting fraud, which led to consumers losing money.²⁹² While both the lawsuits against PayPal for false positive fraud detection and non-detection of fraud were unsuccessful, this is more caused by mandatory arbitration clauses and class-action waiver clauses found in the US terms of use of PayPal which prevents true precedent-setting litigation,²⁹³ than there being no issues with the ML-based fraud detection system used by PayPal. The arbitration clauses and class-action waiver clauses used by PayPal cause most disputes to never reach a public court. Furthermore, the user agreement used by PayPal gives PayPal broad contractual discretion to freeze or reverse transactions,²⁹⁴ making lawsuits based on false positive fraud detection extremely unlikely to succeed. The absence of successful cases against PayPal for false positive fraud detection and non-detection of fraud is therefore not evidence of the absence of harm done by the ML-based fraud detection system used by PayPal.

The impact of a failing fraud detection system can have become clear when a programming error in a local German specific update of PayPal caused PayPal's fraud detection systems to fail.²⁹⁵ The failing of PayPal's fraud detection systems caused widespread payment disruptions in Germany and led to German banks halting almost 10 billion euros in PayPal

²⁹¹ *Mehmet v. PayPal, Inc.*, No. C-08-01961 RMW (N.D. Cal. Aug. 12, 2008); *Chen v. PayPal, Inc.*, 61 Cal. App. 5th 559 (2021).

²⁹² DCCA Hawaii, 'Release: Hawai'i office of consumer protection announces \$6 million settlement with Paypal' (DCCA Hawaii 22 December 2025) <<https://cca.hawaii.gov/ocp/release-hawaii%CA%BBi-office-of-consumer-protection-announces-6-million-settlement-with-paypal/>> accessed 21 April 2026; *Al-Ramahi v. PayPal, Inc.*, No. 5:22-cv-03632, complaint (N.D. Cal. filed June 21, 2022).

²⁹³ 'Paypal user agreement' (Paypal 26 January 2026) <https://www.paypal.com/us/legalhub/paypal/useragreement-full?country.x=US&locale.x=en_US#s6-agreement-arbitrate> accessed 26 April 2026; *Evans v. PayPal, Inc.*, No. 22-cv-00248-BLF (N.D. Cal. June 2, 2022); *Nezri v. PayPal, Inc.*, No. CV 22-2112 DSF (JPRx) (C.D. Cal. June 13, 2022); *Norkunas v. PayPal, Inc.*, No. 1:22-cv-05695-KMW-EAP, slip op. (D.N.J. July 12, 2023).

²⁹⁴ 'Paypal user agreement' (Paypal 26 January 2026) <https://www.paypal.com/us/legalhub/paypal/useragreement-full?country.x=US&locale.x=en_US#s6-agreement-arbitrate> accessed 21 April 2026.

²⁹⁵ 'Technical error disrupts PayPal transactions in Germany' (RTL today 3 September 2025) <<https://today.rtl.lu/news/luxembourg/technical-error-disrupts-paypal-transactions-in-germany-2333463>> accessed 21 April 2026.

payments as a precaution.²⁹⁶ Due to German banks quickly freezing payment transactions, damages caused by the failing of PayPal's fraud detection system remained limited.²⁹⁷ Still, these freezes had an impact on online shops, supermarkets and retail stores, as during these freezes consumers were unable to use PayPal's services.²⁹⁸

Another example of insufficient ML-based fraud detection systems used by payment apps causing issues is Zelle.²⁹⁹ Zelle is a peer-to-peer payment app used to send or receive money directly from or to bank accounts. Zelle has become one of the largest peer-to-peer payment apps available on the US market with it being embedded in many bank and credit union accounts. It is estimated that in 2023 alone, 120 million consumers and small businesses used Zelle to transfer 800 billion dollar.³⁰⁰ Zelle was developed by Early warning Services, a fintech company that is owned by 7 of the largest banks in the United States: Bank of America, Truist, Capital One, JPMorgan Chase, PNC Bank, U.S. Bank and Wells Fargo. Zelle was seen as the response of these banks to other popular peer to

²⁹⁶ 'Technical error disrupts PayPal transactions in Germany' (RTL today 3 September 2025) < <https://today.rtl.lu/news/luxembourg/technical-error-disrupts-paypal-transactions-in-germany-2333463>> accessed 21 April 2026.

²⁹⁷ Tom Gerken, 'European banks hit by rogue PayPal payments worth 'billions' (BBC 27 August 2025) < <https://www.bbc.com/news/articles/cj6y42jggdvo>> accessed 21 April 2026; Alina Boger, 'Ten billion euros frozen? PayPal security systems down - German banks block payments in the billions' *Nordbayern* (27 August 2025) < <https://www.nordbayern.de/news-in-english/paypal-security-systems-down-german-banks-block-payments-in-the-billions-1.14811187>> accessed 21 April 2026.

²⁹⁸ Alina Boger, 'Ten billion euros frozen? PayPal security systems down - German banks block payments in the billions' *Nordbayern* (27 August 2025) < <https://www.nordbayern.de/news-in-english/paypal-security-systems-down-german-banks-block-payments-in-the-billions-1.14811187>> accessed 21 April 2026.

²⁹⁹ 'CFPB Sues JPMorgan Chase, Bank of America, and Wells Fargo for Allowing Fraud to Fester on Zelle' (cfpb Consumer Financial Protection Bureau 20 December 2024) < <https://wayback.archive-it.org/23481/20241226205242/https://www.consumerfinance.gov/about-us/newsroom/cfpb-sues-jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle/>> accessed 21 April 2026.

³⁰⁰ Louis DeNicola 'Here's What You Need to Know About Zelle'(Experian 20 February 2025) < <https://www.experian.com/blogs/ask-experian/heres-what-you-need-to-know-about-zelle-the-mobile-payment-app-that-rivals-venmo/>> accessed 21 April 2026.

peer payment systems like PayPal and Venmo.³⁰¹ According to the CFPB³⁰² the response of these banks was rushed, which caused a neglect of proper safeguards, leading to Zelle becoming a goldmine for fraudsters and leaving fraud victims defenseless.³⁰³ While the CFPB initially started a lawsuit against Zelle, it has recently dropped this case. The dropping of this case seems to be part of a wider trend under the Trump administration, during which the CFPB has scaled back a number of corporate enforcement cases.³⁰⁴ However, The attorney general of New York has partly revived this case and is seeking restitution by Zelle for the damages incurred by new Yorkers who are victims of fraud, with the total claimed amount reportedly reaching 1 billion dollars.³⁰⁵ Given the recent nature of these developments, it remains uncertain how the case will ultimately be resolved.

(b) GDPR

As Zelle is not active on the European market, this analysis primarily focuses on PayPal. Should Zelle enter the EU market in the future, the considerations outlined in this section would apply equally to its operations.

Art. 6(1)(f) GDPR provides a plausible lawful basis for the processing of personal data for fraud detection purposes for PayPal. The recitals of the GDPR explicitly determine that

³⁰¹ Hugh Son ‘Zelle payments top \$1 trillion in 2024 as network’s growth outpaces rivals including PayPal’ (CNBC 12 February 2025) < <https://www.cnbc.com/2025/02/12/zelle-payments-top-1-trillion-in-2024>.> accessed 21 April 2026.

³⁰² The CFPB is the consumer financial protection bureau. It is an independent U.S. agency created to protect consumers in the financial market place.

³⁰³ ‘CFPB Sues JPMorgan Chase, Bank of America, and Wells Fargo for Allowing Fraud to Fester on Zelle’ (cfpb Consumer Financial Protection Bureau 20 December 2024) < <https://wayback.archive-it.org/23481/20241226205242/https://www.consumerfinance.gov/about-us/newsroom/cfpb-sues-jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle/>> accessed 21 April 2026.

³⁰⁴ Krause (n 276), 12; Laurence Darmiento, ‘Consumer agency drops Zelle lawsuit against big banks in latest legal pullback’ *Los Angeles Times* (5 March 2025) < <https://www.latimes.com/business/story/2025-03-05/consumer-financial-protection-bureau-zelle-lawsuit>> accessed 21 April 2026.

³⁰⁵ ‘Attorney General James Sues Company Behind Zelle for Enabling Widespread Fraud’ (Office of the New York state attorney general 13 August 2025) < <https://ag.ny.gov/press-release/2025/attorney-general-james-sues-company-behind-zelle-enabling-widespread-fraud>> accessed 26 April 2026.

fraud prevention can constitute a legitimate interest.³⁰⁶ As I explained in chapter 2.1.2.1.1 of this thesis, a division between fraud prevention and fraud detection is not made in this thesis. Art. 6 (1) (c) GDPR can also provide a lawful basis for the processing of personal data for PayPal when the processing is necessary for the purposes of preventing money laundering and terrorist financing.³⁰⁷

For PayPal to rely on art. 6(1)(f) GDPR as a lawful basis for the processing of personal data a balancing exercise must be held between the interest in fraud prevention and the fundamental rights and freedoms of the data subjects.³⁰⁸ Furthermore, PayPal has to assess if the prevention of fraud can not be reasonably achieved just as effectively by other means less restrictive of fundamental rights and freedoms of data subjects.³⁰⁹

PayPal also must comply with the transparency obligations under Articles 12–15 GDPR and does so via a privacy statement that can be found on their website.³¹⁰ In this privacy statement the types of personal data that PayPal collects for fraud detection purposes are clearly explained, which allows PayPal users to determine in advance what the scope and consequences of personal data processing by PayPal for fraud detection purposes entails.³¹¹ This privacy statement does however not provide much information about the inner workings of the ML-based fraud detection system used by PayPal. As was mentioned in chapter 2.1.2.3 of this thesis, trade secret considerations and the idea that transparency could enable fraudsters to game the system are considerable barriers for transparency for ML-based fraud detection systems. While the ‘right to explanation’ would likely allow data subjects being confronted with the fraud detection system used by PayPal to gain more information about the inner workings of the ML-based fraud detection system, exercising

³⁰⁶ GDPR, rec. 47.

³⁰⁷ AMLR, art. 76 (1).

³⁰⁸ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para 31.

³⁰⁹ European Data Protection Board, ‘Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR’, para 29.

³¹⁰ ‘Paypal Privacy Statement’ (Paypal.com 2 April 2026), <https://www.paypal.com/us/legalhub/paypal/privacy-full?country.x=US&locale.x=en_US> accessed 21 April 2026.

³¹¹ ‘Paypal Privacy Statement’ (Paypal.com 2 April 2026), <https://www.paypal.com/us/legalhub/paypal/privacy-full?country.x=US&locale.x=en_US> accessed 21 April 2026.

the right to explanation is a heavy burden for data subjects,³¹² making this thus an ineffective tool in this context for transparency.

Furthermore, Article 22 GDPR may be relevant where the ML-based fraud detection system of Uber leads to decisions such as the blocking of transactions, freezing of accounts, or refusal of payments, as these measures may produce legal or similarly significant effects for data subjects. Although payment providers may argue that such outcomes are not “solely automated” due to the involvement of human review, the assessment must focus on whether this human involvement is meaningful in substance rather than merely formal. As clarified by EDPB guidance³¹³ and SCHUFA,³¹⁴ Article 22 may still apply where human intervention does not substantially alter the automated outcome. Automation bias³¹⁵ can cause humans to exercise less scrutiny when making a fraud determination.

PayPal will however in certain cases be able to rely on EU law to justify decisions based on automated processing.³¹⁶ As explained in chapter 2.1.3.2 of this thesis, art. 76 of the AMLR allows entities who fall under the scope of the AMLR to adopt decisions resulting from automated processing when this is strictly necessary for the purposes of preventing money laundering and terrorist financing. PayPal clearly falls under the scope of the AMLR, as it can be considered as a financial institution as defined by the AMLR.³¹⁷ Art. 76 (5) (b) AMLR however still requires human intervention in many cases when the decision has an impact on a consumer, limiting the impact of this provision.

(c) AI Act

While there is only limited information available about the exact workings of the ML-based fraud detection systems used by PayPal and Zelle, it is likely that these detection systems will be considered AI systems under the AI Act as it is known that these systems make use of machine learning techniques.³¹⁸

³¹² Edwards (n 74), 51.

³¹³ Article 29 Data Protection Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’, 21.

³¹⁴ *UF AB v Land Hessen* (n 77).

³¹⁵ Ruschemeier (n 242), 1.

³¹⁶ GDPR, art. 22 (2) (b).

³¹⁷ AMLR, art. 2 (6) (a) and 3 .

³¹⁸ Paypal Editorial Staff, ‘Harnessing machine learning fraud detection technologies’ (Paypal.com 5 November 2024) <<https://www.paypal.com/us/brc/article/payment-fraud->

The AI Act will however have a limited impact on ML-based fraud detection systems used by private entities for the purpose of detecting fraud in the offering of financial services, as the recitals of the AI Act determine that these should not be considered as high-risk AI systems.³¹⁹ Similar to the Connect-system, ML-based fraud detection systems used in the payments sector will rarely fall under the third risk-category as these systems rarely engage directly with the consumer.³²⁰ ML-based fraud detection systems will thus fall under the fourth risk category. As AI systems falling under this risk-category are not regulated, the AI Act would not contain any obligations for PayPal or Zelle.³²¹

(d) Conclusion

The experiences of PayPal and Zelle illustrate that the use of ML-based fraud detection systems in the payment sector entails certain risks. These risks include the risk of false positives, non-detection of fraud and non-transparency.

As was shown via the PayPal case study, false positives can have a considerable financial impact on merchants, with some experts even estimating that the financial impact of false detection is higher than the impact of fraud.³²² Furthermore, false positives may result in customers' accounts being frozen, leading to a complete lock-out from the payment service.³²³ Zelle, however, demonstrated that the risk goes both ways, as non-detection of fraud can also have considerable consequences.³²⁴ The payment industry thus has to walk

detection-machine-learning> accessed 21 April 2026; 'Xceed Zelle P2P Real-time' (Nice actimize xceed) <https://www.niceactimize.com/Documents/xceed_zelle_p2p_datasheet.pdf> accessed 21 April 2026.

³¹⁹ AI Act, rec. 58.

³²⁰ AI Act, art. 50

³²¹ European Commission, 'Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)', para. 2.

³²² Wallny (n 286), 1572.

³²³ *Mehmet*, No. C-08-01961; *Chen*, 61 Cal. App. 5th 559; DCCA Hawaii, 'Release: Hawai'i office of consumer protection announces \$6 million settlement with Paypal' (DCCA Hawaii 22 December 2025) < <https://cca.hawaii.gov/ocp/release-hawaii%CA%BBi-office-of-consumer-protection-announces-6-million-settlement-with-paypal/>> accessed 21 April 2026.

³²⁴ 'CFPB Sues JPMorgan Chase, Bank of America, and Wells Fargo for Allowing Fraud to Fester on Zelle' (cfpb Consumer Financial Protection Bureau 20 December 2024) < <https://wayback.archive-it.org/23481/20241226205242/https://www.consumerfinance.gov/about-us/newsroom/cfpb->

a delicate line between over-enforcement, which risks unjustified exclusion and financial harm to legitimate users, and under-enforcement, which exposes both consumers and institutions to significant fraud-related losses.

Non-transparency is another risk identified via the PayPal case study. While the privacy policy of PayPal provided information about the personal data that is processed and for which purposes these are processed, the privacy policy remains silent workings of the ML-based fraud detection system employed by PayPal.³²⁵ As was shown via the Dutch child benefit scandal case study, non-transparency can cause violations of data subjects' rights to go unnoticed. Automation-bias³²⁶ can also have an impact on this as it causes humans reviewing the decision made by the ML-based fraud detection system with less scrutiny.

2. Uber fraud detection

Labour platforms like Uber use algorithms to manage their workers.³²⁷ It has been shown that this algorithmic management often leads to platform workers facing financial, psychological, and physical harm due to opaque algorithmic decisions.³²⁸ While the algorithmic management systems used by Uber are not limited to just ML-based fraud detection systems, fraud detection does play an important role in the algorithms that Uber uses. Uber faces many different types of fraud, including payment fraud and promo abuse.³²⁹ This section will however focus on the ML-based fraud detection systems used by Uber to detect fraud committed by their platform workers. Common types of fraud committed by Uber platform workers include the use of fake GPS systems, either to trick the system into believing that a driver is closer to a customer than he actually is which

[sues-jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle/](https://www.jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle/)> accessed 21 April 2026.

³²⁵ 'Paypal Privacy Statement' (Paypal.com 2 April 2026), <https://www.paypal.com/us/legalhub/paypal/privacy-full?country.x=US&locale.x=en_US> accessed 21 April 2026.

³²⁶ Ruschemeier (n 242), 1.

³²⁷ Emma McDaid, Paul Andon and Clinton Free, 'Algorithmic management and the politics of demand: Control and resistance at Uber' (2023) 109 *Accounting, Organizations and society* 1, 1-2.

³²⁸ Varun N. Rao and others, 'Rideshare Transparency: Translating Gig Worker Insights on AI Platform Design to Policy' (2025) 9 *Proceedings of the ACM on Human-Computer Interaction* 1, 3.

³²⁹ Kanishka Hewageegana and others, 'A Survey on Graph Neural Networks for Fraud Detection in Ride Hailing Platforms' [2024] 7th *International Conference on Artificial Intelligence and Big Data* 168, 169.

allows drivers to gain an unfair advantage over other drivers, or to trick the system into believing that a driver picked a customer up or dropped him off in another place than he actually did. Other common fraud tactics include the use of long hauling which allows drivers to earn more by deliberately taking longer or less efficient routes, the use of ride collusion which works by a fake-customer and a driver working together to trick the app into believing that a service was actually provided while that was not the case and the use of hire conversions which works by the driver getting to a location with a customer, cancelling the drive and then directly negotiating with the customer, often offering him a lower price than the Uber app.³³⁰

A prominent ML-based fraud detection system used by Uber to detect driver fraud is risk entity watch.³³¹ Risk-entity watch uses machine-learning and anomaly detection to identify entities who deviate from normal patterns. Instead of Risk-entity watch evaluating a single transaction, it analyzes the historical behavior of entities and generates an anomaly score for these entities. Human analysts then review these anomaly scores together with a by Risk-entity watch generated explanation describing what behaviors of the flagged entity diverged from normal patterns.³³²

The use of ML-based fraud detection systems to detect worker fraud has not been without controversy, as fraud can be a cause of termination of these workers.³³³ As will be explored in the following chapter of this thesis, this has led to lawsuits to arise as workers challenge the transparency and art. 22 compliance of these systems.³³⁴

(a) GDPR

Uber will rely on Art. 6 (1) f GDPR as a lawful basis for the processing of personal data in the context of detecting fraud by active drivers on their platform. As was already

³³⁰ Hewageegana (n 329), 169.

³³¹ Becky Hui and others, 'Risk Entity Watch – Using Anomaly Detection to Fight Fraud' (Uber Blog September 28 2023) < <https://www.uber.com/us/en/blog/risk-entity-watch/>> accessed 21 April 2026.

³³² Becky Hui and others, 'Risk Entity Watch – Using Anomaly Detection to Fight Fraud' (Uber Blog September 28 2023) < <https://www.uber.com/us/en/blog/risk-entity-watch/>> accessed 21 April 2026.

³³³ 'Fraud activities' (Uber) < <https://www.uber.com/be/en/drive/driver-app/fraud-activities/>> accessed 21 April 2026.

³³⁴ Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1020; Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018.

mentioned previously in this thesis, fraud can constitute a legitimate interest.³³⁵ Uber will however only be allowed to process as much personal data as is necessary to achieve the legitimate interest of fraud prevention and must balance the fundamental rights of the workers with the legitimate interest pursued. Only if the fundamental rights of the workers do not override the legitimate interest of Uber to prevent fraud will the personal data processing comply with the GDPR.³³⁶ Furthermore, Uber will have to comply with the transparency provisions found in art. 12-15 of the GDPR and is unable to make decisions based solely on automated processing.³³⁷

A 2021 case before the Amsterdam District Court concerned claims by four Uber drivers who were terminated for alleged fraud. These four drivers argued that Uber had failed to comply with the GDPR's transparency requirements and had infringed their right not to be subject to decisions based solely on automated processing.³³⁸ While the transparency claims were partly successful, as the court held that two of the four drivers were insufficiently informed of the reasons for their termination, leaving them unable to check the accuracy and lawfulness of the termination decision,³³⁹ the art. 22 GDPR claim was less successful as the court determined that there was no infringement of art. 22 due to there being sufficient human involvement.³⁴⁰ While the initial flagging by the ML- based fraud detection system did temporarily block the access of the driver to the Uber app, the court determined that this temporary block was not sufficient to be considered as producing similarly significant effects.³⁴¹ When considering if art. 22 of the GDPR was applicable, the court did have attention for the human involvement actually being meaningful, but ultimately the court decided that the human involvement in this case was sufficient to avoid an application of art. 22 GDPR.³⁴² Ultimately, this case illustrates that the GDPR's 'right to explanation' can enable data subjects to obtain insight into the

³³⁵ GDPR, rec. 47.

³³⁶ GDPR, art. 6 (1) (f).

³³⁷ GDPR, art. 22.

³³⁸ Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018, para. 4.4.

³³⁹ Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018, para. 4.29.

³⁴⁰ Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018, para. 4.17-4.26.

³⁴¹ Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018, para. 4.25.

³⁴² Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018, para. 4.17-4.26.

functioning of ML-based fraud detection systems, if they are able to invest the necessary time and resources to exercise their rights.³⁴³

(b) AI Act

Uber's risk entity watch will likely be considered as an AI system under the AI Act³⁴⁴ as it generates risk-scores autonomously, is able to adapt itself throughout its lifespan and infers due to the machine-learning approach it uses.

As was explained in chapter 2.2.2.2 of this thesis, ML-based fraud detection systems used by labour platforms to detect fraud of their workers will be considered as high-risk AI systems due to these systems falling under art. 4 (b) of annex III. Risk entity watch will thus have to comply with the obligations determined by the AI Act for high-risk AI systems. In particular, Uber will be required to implement a risk management system,³⁴⁵ ensure the quality and representativeness of the data used,³⁴⁶ and comply with the documentation obligations.³⁴⁷ Moreover, the ML-based fraud detection system allow for effective human oversight,³⁴⁸ and meet requirements of accuracy, robustness, and cybersecurity.³⁴⁹

The exact impact that these new obligations will have on the ML-based fraud detection system used by Uber is of course unknown. However, similar to the risk-classification model used in the Dutch child benefit scandal, I believe that the documentation obligations³⁵⁰ and obligation to implement a risk-management system³⁵¹ will improve transparency about the inner workings of the fraud detection system used. This greater transparency would be able to stop violations of data subjects' rights earlier and would thus mainly have a preventive purpose.

³⁴³ Edwards (74), 51.

³⁴⁴ AI Act, art. 3 (1).

³⁴⁵ AI Act, art. 9.

³⁴⁶ AI Act, art. 10.

³⁴⁷ AI Act, art. 11-12.

³⁴⁸ AI Act, art. 14.

³⁴⁹ AI Act, art. 15.

³⁵⁰ AI Act, art. 11-12.

³⁵¹ AI Act, art. 9.

(c) Conclusion

The main risk exemplified by Uber's risk entity watch is the risk that non-transparency poses. Uber's risk entity watch is deployed in a sensitive context, as drivers' fraud can cause a driver to be terminated from the Uber platform and thus lose their income. Key in making sure that terminated drivers are capable to understand why this termination has happened is transparency. While the GDPR does currently allow drivers to gain an explanation of a termination decision via the right to an explanation,³⁵² the exercise of this right requires significant resources.³⁵³ The obligations implemented by the AI Act for high risk systems will expand this transparency, allowing for a partial shift of the burden away from individuals by imposing ex ante documentation³⁵⁴ and accountability requirements³⁵⁵ on providers and deployers.

IV. Comparison

While ML-based fraud detection systems are subject to different legal regimes under the AI Act and under the GDPR depending on whether they are deployed in the public or private sector and depending in which domain they are deployed (e.g. welfare vs taxation, payments vs labour platforms), the case studies demonstrate that the underlying risks and benefits they generate are remarkably similar.

First, both in the public sector and in the private sector the risk of false positive fraud detection emerges as a central concern. In the public sector, the Dutch child benefits scandal illustrates how false positives can lead to severe and systemic harm, including wrongful accusations of fraud and financial recovery measures. In the private sector, similar errors manifest through account freezes (PayPal) or wrongful terminations (Uber). Secondly, the risk of automation bias is present irrespective of the institutional setting. While I was unable to find empirical research investigating the existence of automation bias specifically in a fraud detection context, research on automation bias in the medical

³⁵² Rb. Amsterdam 11 maart 2021, ECLI:NL:RBAMS:2021:1018.

³⁵³ Edwards (n 74), 51.

³⁵⁴ AI Act, art. 11-12.

³⁵⁵ AI Act, art. 9, 10, 14 and 15.

sector,³⁵⁶ self-driving cars,³⁵⁷ aviation sector³⁵⁸ and public sector³⁵⁹ seems to indicate that this bias is not contained to a specific field. There is nothing that indicates that this bias would not play a role in the context of ML-based fraud detection. The Dutch child benefits scandal demonstrates how automation bias in a fraud detection system can lead to an ML-based fraud detection system that is supposed to work as a decision-support system transforming into a de facto decision-maker.

Thirdly, opacity constitutes a structural risk common to all case studies. The limited transparency surrounding these systems restricts the ability of individuals to understand, challenge, or verify decisions affecting them. Fourthly, the risk of discrimination and bias exists across both the public and private sector, with transparency seemingly being the best way to prevent these biases from persisting.

Notwithstanding these risks, ML-based fraud detection systems also generate significant benefits across both the public and private sectors. First, they substantially enhance the efficiency of fraud detection by enabling the processing of large volumes of transactions or claims that would be impossible to review manually.³⁶⁰ As was demonstrated by the HRMC ‘Connect’ case study this allows entities to allocate resources more effectively, focusing human oversight on high-risk cases. Secondly, such systems can improve the accuracy of fraud detection compared to purely rule-based or manual approaches, particularly where sophisticated or evolving fraud patterns are involved.³⁶¹ Lastly, ML-based systems can enable more proactive and real-time intervention, allowing fraud to be detected and mitigated at an earlier stage.³⁶²

³⁵⁶ Kate Goddard, Abdul Roudsari and Jeremy C. Wyatt ‘Automation bias: a systematic review of frequency, effect mediators, and mitigators’ (2012) 19 *Journal of the American Medical Informatics Association* 121.

³⁵⁷ Lauren Kahn, Emilia S. Probasco and Ronnie Kinoshita, *AI Safety and Automation Bias The Downside of Human-in-the-Loop* (Center for Security and Emerging Technologies 2024) 10-13.

³⁵⁸ Kahn (n 357), 14-22.

³⁵⁹ Ruschemeier (n 242).

³⁶⁰ Okolo (n 8), 1198.

³⁶¹ Okolo (n 8), 1198.

³⁶² Benjamin Borketey, ‘Real-Time Fraud Detection Using Machine Learning’ (2024) 12 *Journal of Data Analysis and Information Processing* 189.

While many of the risks of ML-based fraud detection systems in the private and public sector are similar, the application of the GDPR and AI Act differs even within the public and private sector. The public-private dichotomy does not appear to influence the AI Act assessment of whether an ML-based fraud detection system qualifies as high-risk. However, this dichotomy does seem to have an impact on the application of the GDPR and this not only on the lawful bases on which entities can rely for the processing of personal data. The GDPR appears to allow public ML-based fraud detection systems to avoid many of the obligations that apply to private ML-based fraud detection systems. This is primarily due to exceptions permitted via national law. The GDPR allows member states to restrict via member state law the transparency obligations³⁶³ of the GDPR and art. 22 of the GDPR.³⁶⁴

V. Conclusion

This thesis set out to examine how ML-based fraud detection systems are used in both the public and private sectors, to identify their key risks and benefits, and to assess whether the GDPR and the AI Act are adequate to protect against the risks of ML-based fraud detection systems.

The analysis has shown that machine-learning techniques have allowed fraud detection to evolve from being rule-based to being ML-based, which has significantly increased efficiency and detection capabilities.

A central finding of this thesis is that, despite the legal differences between the public and private sectors, the underlying risks associated with ML-based fraud detection systems are largely similar. Issues such as false positives, automation biases, lack of transparency, risks of discrimination, arise in both contexts. The case studies examined in this thesis demonstrate that these risks are not theoretical, but can have profound real-world consequences.

Regarding the legal framework, while the GDPR provides a robust foundation, this thesis has shown that the GDPR struggles to fully capture the complexity of ML-based fraud detection systems. In particular, the distinction between fully automated and partially automated decision-making, as well as the use of multi-stage systems, creates significant

³⁶³ GDPR, art. 12-14.

³⁶⁴ GDPR, art. 23.

legal uncertainty. The functional approach adopted in the SCHUFA case partially addresses this issue, but ambiguities remain. Furthermore, due to the GDPR allowing the restriction of certain obligations via member state law, certain public ML-based fraud detection systems are able to operate under reduced transparency and accountability requirements, thereby weakening the level of protection afforded to data subjects.

The AI Act represents an important step forward by introducing a risk-based regulatory framework. Its classification of certain ML-based fraud detection systems as high-risk systems is a promising development. However, the exclusion of many ML-based fraud detection systems, especially in the financial sector and in the tax sector, raises concerns about regulatory gaps. Furthermore, the ongoing uncertainty regarding the interpretation and enforcement of the AI Act limits its immediate effectiveness.

Overall, this thesis concludes that while the current legal framework offers important safeguards, it is not fully adapted to the realities of ML-based fraud detection. Especially under the AI Act it is unclear why a distinction is made between certain sectors as the risks of ML-based fraud detection systems cut across these boundaries.

Future regulatory efforts should therefore focus less on formal distinctions between sectors and more on the actual risks posed by ML-based fraud detection systems. Greater emphasis should be placed on meaningful human oversight and transparency. Only by addressing these cross-cutting concerns can the law effectively respond to the challenges posed by ML-based fraud detection.

VI. Bibliography

Alabi M, 'Bias and Fairness in AI-Based Fraud Detection: An Ethical Assessment in Financial Applications' (unpublished manuscript 9 September 2023).

Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* (Amnesty International 2021).

Article 29 Working Party, 'Guidelines on transparency under Regulation 2016/679', (29 November 2017).

Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679', (11 April 2018)

Ariyibi K and others, 'Leveraging Artificial Intelligence for enhanced tax fraud detection in modern fiscal systems' [2024] GSC Advanced research and reviews 129, 131.

'Article 6 GDPR' (GDPRhub) <https://gdprhub.eu/index.php/Article_6_GDPR> accessed 31 March 2026.

‘Attorney General James Sues Company Behind Zelle for Enabling Widespread Fraud’ (Office of the New York state attorney general 13 August 2025) < <https://ag.ny.gov/press-release/2025/attorney-general-james-sues-company-behind-zelle-enabling-widespread-fraud> > accessed 26 April 2026.

Autoriteit persoonsgegevens, ‘boete belastingsdienst kinderopvangtoeslag’ (Autoriteit persoonsgegevens 7 December 2021) <<https://www.autoriteitpersoonsgegevens.nl/documenten/boete-belastingdienst-kinderopvangtoeslag>> accessed 31 March 2026

Barbosa S and Félix S, ‘Algorithms and the GDPR: An analysis of article 22’ [2021] *Anuário da Proteção de Dados*, 75.

Becker R, Volinsky C and Wilks A, ‘Fraud Detection in Telecommunications: History and Lessons Learned’ (2010) 52 *Technometrics* 20.

Binns R and Veale M, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’ (2021) 11 *International Data Privacy Law* 319.

Boger A, ‘Ten billion euros frozen? PayPal security systems down - German banks block payments in the billions’ *Nordbayern* (27 August 2025) < <https://www.nordbayern.de/news-in-english/paypal-security-systems-down-german-banks-block-payments-in-the-billions-1.14811187> > accessed 21 April 2026.

Borgesius F and Poort J, ‘Online Price Discrimination and EU Data Privacy Law’ (2017) 40 *Journal of Consumer Policy* 347.

Borketey B, ‘Real-Time Fraud Detection Using Machine Learning’ (2024) 12 *Journal of Data Analysis and Information Processing* 189.

Buelens S, Platteau and Torfs, ‘AI-Verordening. Inzoomen op de eerste verplichtingen’ [2026] *Nieuw Juridisch Weekblad* 2.

‘CFPB Sues JPMorgan Chase, Bank of America, and Wells Fargo for Allowing Fraud to Fester on Zelle’ (cfpb Consumer Financial Protection Bureau 20 December 2024) < <https://wayback.archive-it.org/23481/20241226205242/https://www.consumerfinance.gov/about-us/newsroom/cfpb-sues-jpmorgan-chase-bank-of-america-and-wells-fargo-for-allowing-fraud-to-fester-on-zelle/> > accessed 21 April 2026.

Charlotte B and Callaghan A, ‘Artificial Intelligence Approaches to Fraud Detection in Public Procurement’ (unpublished manuscript, 4 September 2025).

Cox K, Rooses D and Wuyts T ‘Een juridische verkenning van de AI Act’ (2024) 88 *Rechtskundig Weekblad* 283.

Council of the EU ‘Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI’ (European Council Council of the European Union 21 May 2024) <<https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial->

intelligence-AI Act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/> accessed 21 April 2026.

Daly S, 'Artificial intelligence, the Rule of Law and public administration: the case of taxation' (2024) 83 *The Cambridge Law Journal* 437.

Darmiento L, 'Consumer agency drops Zelle lawsuit against big banks in latest legal pullback' *Los Angeles Times* (5 March 2025) <<https://www.latimes.com/business/story/2025-03-05/consumer-financial-protection-bureau-zelle-lawsuit>> accessed 21 April 2026.

Davis P and Schwemer, 'Rethinking Decisions under Article 22 of the GDPR: Implications for Semi-Automated Legal Decision-Making' [2023] *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace* 1.

DCCA Hawaii, 'Release: Hawai'i office of consumer protection announces \$6 million settlement with Paypal' (DCCA Hawaii 22 December 2025) <<https://cca.hawaii.gov/ocp/release-hawai%CA%BBi-office-of-consumer-protection-announces-6-million-settlement-with-paypal/>> accessed 21 April 2026.

DeNicola L, 'Here's What You Need to Know About Zelle'(Experian 20 February 2025) <<https://www.experian.com/blogs/ask-experian/heres-what-you-need-to-know-about-zelle-the-mobile-payment-app-that-rivals-venmo/>> accessed 21 April 2026.

Dyson R, 'What does the taxman know about you, your finances and your lifestyle?' *The Telegraph* (25 June 2015) <<https://www.telegraph.co.uk/finance/personalfinance/tax/11697816/What-does-the-taxman-know-about-you-your-finances-and-your-lifestyle.html>> accessed 1 April 2026.

Ebers M 'Truly Risk-based Regulation of Artificial Intelligence How to Implement the EU's AI Act' (2025) 16 *European Journal of Risk Regulation* 684.

Edwards L and Veale M, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 *IEEE Security & Privacy* 46.

European Commission, 'Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)' COM(2025) 5053 final.

European Data Protection Board, 'Guidelines 2/2019 on the Processing of Personal Data under Article 6(1)(b) GDPR in the Context of the Provision of Online Services to Data Subjects', 2/2019 (8 October 2019).

European Data Protection Board, 'Guidelines 03/2022 on Deceptive design patterns in social media platform interfaces: how to recognise and avoid them', 03/2022 (14 February 2023).

European Data Protection Board, 'Guidelines 1/2024 on Processing of Personal Data Based on Article 6(1)(f) GDPR', 1/2024 (8 October 2024).

‘Feiten en cijfers’ (Dienst Toeslagen Ministerie van Financiën 3 April 2026) <<https://herstel.toeslagen.nl/dashboard-kinderopvangtoeslag/>> accessed 21 April 2026.

Finck M and Pallas F, ‘They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data under the GDPR’ (2020) 10 International Data Privacy Law 11

‘Fraud activities’ (Uber) < <https://www.uber.com/be/en/drive/driver-app/fraud-activities/>> accessed 21 April 2026.

‘Fraud Protection Advanced’ (Paypal Developer 28 August 2024) <<https://developer.paypal.com/docs/checkout/advanced/customize/fraud-protection/fraud-protection-advanced/>> accessed 21 April 2026.

Gerken T, ‘European banks hit by rogue PayPal payments worth 'billions' (BBC 27 August 2025) < <https://www.bbc.com/news/articles/cj6y42jggdyy>> accessed 21 April 2026

Gerken T, ‘HMRC admits using AI to scan social media for tax cheats’ BBC (12 August 2025) < <https://www.bbc.com/news/articles/cqjyedz202ko>> accessed 21 April 2026.

Goddard K Roudsari A and Wyatt J, ‘Automation bias: a systematic review of frequency, effect mediators, and mitigators’ (2012) 19 Journal of the American Medical Informatics Association 121.

‘Got Secrets? HMRC “Connect-System” Targets Tax Evaders Now!’ (eClear 10 May 2023) <<https://eclear.com/article/got-secrets-hmrc-connect-system-targets-tax-evaders-now/>> accessed 21 April 2026.

Grosan C and Abraham A, *Intelligent Systems: A Modern Approach* (Springer-Verlag Berlin Heidelberg 2011).

Gupta H, Soni P and Kumar R, ‘The Evolutionary Arms Race - Safeguarding Financial Stability through Advanced Fraud Detection Techniques’ in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)* (the Institute of Electrical and Electronics Engineers 2024)

Hadwick D and Lan S, ‘Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany’ (2021) 13 World Tax Journal.

Hadwick D, ‘slipping Through the Cracks, the Carve-outs for AI Tax Enforcement Systems in the EU AI Act’ (2024) 9 European Papers 936.

Häuselmann A and Custers B, ‘Substantive Fairness in the GDPR: Fairness Elements for Article 5.1A GDPR’ (2024) 52 Computer Law & Security Review (2024) 1.

Hewageegana K and others, ‘A Survey on Graph Neural Networks for Fraud Detection in Ride Hailing Platforms’ [2024] 7th International Conference on Artificial Intelligence and Big Data 168.

HM Revenue & Customs ‘Guidance HMRC appropriate policy document’ (GOV.UK 8 May 2019) < <https://www.gov.uk/government/publications/hmrc-appropriate-policy-document>> accessed 21 April 2026.

HM Revenue & Customs, 'Official Statistics 1. Tax gaps: Summary' (GOV.UK 19 June 2025) <<https://www.gov.uk/government/statistics/measuring-tax-gaps/1-tax-gaps-summary>> accessed 21 April 2026.

Hui and others, 'Risk Entity Watch – Using Anomaly Detection to Fight Fraud' (Uber Blog September 28 2023) < <https://www.uber.com/us/en/blog/risk-entity-watch/>> accessed 21 April 2026.

Ider K and Schmietendorf A, 'Data Privacy for AI Fraud Detection Models: A framework for GDPR compliant AI' in Mauri J and others (eds) *ICDS 2020 : The Fourteenth International Conference on Digital Society (International Academy, Research, and Industry Association (IARIA) 2020)*

Kahn L, Probasco E and Kinoshita R, *AI Safety and Automation Bias The Downside of Human-in-the-Loop* (Center for Security and Emerging Technologies 2024).

Kasirzadeh A and Clifford D, 'Fairness and Data Protection Impact Assessments' in Fourcade M and others (eds), *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI* (association for Computing Machinery 2021) 146.

Krause D, 'The Future of Consumer Financial Protection: Lessons from the Zelle Lawsuit Dismissal' (unpublished manuscript, 6 March 2026).

Kroll J, 'Accountable Algorithms' (PhD thesis, Princeton University 2015).

Krzysztofek M, *GDPR: Personal Data Protection in the European Union* (Wolters Kluwer 2023) 91.

Lampropoulos S, Thanasis G and Kontogeorga G, 'Fraud Detection in Banking Transactions with the Use of Artificial Intelligence and Anonymized Data' (2025) 24 *Journal of European economy* 656.

Letter from Sir Robert Devereux and Jon Thompson to Meg Hillier (10 June 2016) <<https://www.parliament.uk/globalassets/documents/commons-committees/public-accounts/Correspondence/2015-20-Parliament/PAC-Response-final-signed-copy-of-follow-up-letter-to-3rd-party-data.pdf>> accessed 21 April 2026.

Lu S, 'Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence' (2020) 23 *Vanderbilt Journal of Entertainment and Technology Law* 99, 116

Lu S, 'Data Privacy, Human Rights, and Algorithmic Opacity' (2022) 110 *California Law Review* 2087, 2103.

Malgieri G, 'Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations' (2019) 35 *Computer Law & Security Review*.

Malgieri G, 'The Concept of Fairness in the GDPR' in Hildebrandt M and others (eds.), *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020).

Marchini K and Van Dyke A, 'Overcoming False Positives: Saving the Sale and the Customer Relationship' (Javelin 21 September 2015) <<https://javelinstrategy.com/research/overcoming-false-positives-saving-sale-and-customer-relationship>> accessed 21 April 2026.

Maynard P and Harrington M, 'It's just not fair! The ambiguity of GDPR fairness principle' [2024] Privacy laws & business 20.

McCullagh K, 'Post-Brexit data protection in the UK - leaving the EU but not EU data protection law behind' in González G, Van Brakel R and De Hert P (eds) *Research handbook on privacy and data protection law* (Edward Elgar 2022) 37-38.

McCullagh K, 'post-Brexit UK Data Protection: Staying the Course or Charting a New Direction?' in Edoardo Celeste and others (eds) *Data Protection and Digital Sovereignty Post-Brexit* (Bloomsbury Publishing 2023), 24-25.

Nance R, Evans M and McBurney P, 'The Commission's guidelines on AI systems – what can we infer?' (Norton Rose Fulbright 13 February 2025) <https://www.dataprotectionreport.com/2025/02/the-commissions-guidelines-on-AI_systems-what-can-we-infer/> accessed 21 April 2026.

Natwhani K, *Artificial intelligence in automated decision-making in tax administration: the case for legal, justiciable and enforceable safeguards* (The institute for Fiscal Studies 2024)

Nilson Report 1023' (Nilson Report August 2013) <1023 - Nilson Report> accessed 21 April 2026.

Nilson Report 1118' (Nilson Report October 2017) < 1118 - Nilson Report> accessed 21 April 2026.

OECD, *Governing with Artificial Intelligence The State of Play and Way Forward in Core Government Functions* (OECD Publishing 2025).

Okolo J and others, 'Analyzing the Use of Machine Learning Techniques in Detecting Fraudulent Activities' (2025) 26 World Journal of Advanced Research and Reviews 1198.

Palumbo A, 'The Unexplored Potential of the Fairness Principle under the GDPR: Lessons from the Recent TikTok Case' (CiTiP blog, 7 December 2023) <https://www.law.kuleuven.be/citip/blog/the-unexplored-potential-of-the-fairness-principle-under-the-gdpr-lessons-from-the-recent-tiktok-case/> accessed 21 April 2026.

'Paypal Editorial Staff, 'Harnessing machine learning fraud detection technologies' (Paypal.com 5 November 2024) <<https://www.paypal.com/us/brc/article/payment-fraud-detection-machine-learning>> accessed 21 April 2026.

'Paypal Privacy Statement' (Paypal.com 2 April 2026), <https://www.paypal.com/us/legalhub/paypal/privacy-full?country.x=US&locale.x=en_US> accessed 21 April 2026.

‘Paypal user agreement’ (Paypal 26 January 2026) <https://www.paypal.com/us/legalhub/paypal/useragreement-full?country.x=US&locale.x=en_US#s6-agreement-arbitrate> accessed 26 April 2026;

‘Public Task’ (ICO) <<https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/public-task/>> accessed 31 March 2026.

Ranchordás S and Scarcella , ‘Automated Government for Vulnerable Citizens: Intermediating Rights’ (2021) 30 William & Mary Bill of Rights Journal 373.

Rao V and others, ‘Rideshare Transparency: Translating Gig Worker Insights on AI Platform Design to Policy’ (2025) 9 Proceedings of the ACM on Human-Computer Interaction 1.

Ruscheimer H and Bareis J, ‘Searching for Harmonised Rules: Understanding the Paradigms, Provisions, and Pressing Issues in the Final EU AI Act’ in Gsenger R and Sekwenz M (eds), *Digital Decade: How the EU shapes digitalization* (Nomos 2025).

Ruscheimer H and Hondrich L, ‘Automation bias in public administration – an interdisciplinary perspective from law and psychology’ (2024) 41 Government Information Quarterly 1

Sadgali I, Sael N and Benabbou F, ‘Human Behavior Scoring in Credit Card Fraud Detection’ (2021) 10 IAES International Journal of Artificial Intelligence (IJ-AI) 698.

Sanghrajka J, ‘HMRC’s Connect computer and investigations’ [2020] Taxation <<https://www.taxation.co.uk/articles/hmrc-s-Connect-computer-and-investigations>> accessed 21 April 2026.

Schaake M, ‘Trade secrets shouldn’t shield tech companies’ algorithms from oversight’ (Brookings Institution 2020) < <https://www.brookings.edu/articles/trade-secrets-shouldnt-shield-tech-companies-algorithms-from-oversight/>> accessed 21 April 2026.

Schweblin S, Enríquez M and Luiselli , ‘Explainable AI for Transparency in Automated Fraud Detection Systems’ (unpublished manuscript 6 December 2025).

Simonofski A, ‘Balancing Fraud Analytics with Legal Requirements: Governance Practices and Trade-Offs in Public Administrations’ (2022) 4 Data & Policy e14.

Simpson J, ‘Eight ways that the taxman is spying on you’ The Times (6 September 2025) < <https://www.thetimes.com/money/family-finances/article/eight-ways-that-the-taxman-is-spying-on-you-7ds87gs3q>> accessed 21 April 2026.

Son H ‘Zelle payments top \$1 trillion in 2024 as network’s growth outpaces rivals including PayPal’ (CNBC 12 February 2025) < <https://www.cnn.com/2025/02/12/zelle-payments-top-1-trillion-in-2024>> accessed 21 April 2026.

Surden H, ‘Machine Learning and Law’ (2014) 89 Wash L Rev 87.

Szostek D and Prabucki R, ‘High-Risk AI systems’ in Raposo V (ed), *The European Artificial Intelligence Act Promises and Perils?* (Springer 2025) 166.

Tamponi R and others, 'Lost in Vagueness: Towards Context-Sensitive Standards for Robustness Assessment under the EU AI Act' (unpublished, 19 November 2025).

Taylor E, 'Explanation and the Right to Explanation' (2023) 10 Journal of the American Philosophical Association 467.

'Technical error disrupts PayPal transactions in Germany' (RTL today 3 September 2025) < <https://today.rtl.lu/news/luxembourg/technical-error-disrupts-paypal-transactions-in-germany-2333463>> accessed 21 April 2026.

Tomić S and Štimac V 'Pinning down an octopus: towards an operational definition of AI systems in the EU AI Act', [2025] Journal of European Public Policy 1.

McDaid E, Andon P and Free C, 'Algorithmic management and the politics of demand: Control and resistance at Uber' (2023) 109 Accounting, Organizations and society 1.

Wachter S Mittelstadt B and Russell C, 'Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and ai' (2021) 41 Computer Law & Security Review 1.

Wedge R and others 'Solving the "false positives" problem in fraud prediction Automated Data Science at an Industrial Scale' in Ulf Brefeld and others (eds), *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part III* (Springer 2019) 372.

Wiedemann K 'Profiling and (automated) decision-making under the GDPR: A two-step approach' (2022) 45 Computer Law & Security review 1.

Wood H, 'What is HMRC's Connect AI System?' (TaxAssist Accountants 4 November 2025) < <https://www.taxassist.co.uk/resources/articles/what-is-hmrc-s-Connect-AI-system>> accessed 21 April 2026.

'Xceed Zelle P2P Real-time' (Nice actimize xceed) <https://www.niceactimize.com/Documents/xceed_zelle_p2p_datasheet.pdf> accessed 21 April 2026.