eucrim
European Law Forum: Prevention • Investigation • Prosecution

Article

Reflections on Introducing Artificial Intelligence Tools in Support of Anti-Fraud

Bogdan Necula, Georg Roebling [*]

ABSTRACT

Over the coming years, new tools based on large language models (LLMs) and other artificial intelligence-based software are set to play an increasing role in many modern administrations, including in the anti-fraud domain. One might even argue that the prevention, detection, and investigation of fraud and associated illegal activities, which today involve processing and analysing an ever-growing volume of data of different types, are uniquely suited to the strengths of such tools. The authors of this article share some reflections on two particular challenges that authorities, which seek to harvest the potential of artificial intelligence for anti-fraud purposes, have to come to terms with: first, how to leverage the strength of artificial intelligence tools by identifying suitable use cases for the specific anti-fraud domain? Second, how to navigate the emerging regulatory framework considering in particular that the European Union's Artificial Intelligence Act has entered into force on 1 August 2024?

AUTHORS

Bogdan Necula

Deputy Head of Unit
European Anti-Fraud Office (OLAF)

Georg Roebling

Head of Unit
European Commission / European Anti-Fraud Office (OLAF)

CITATION SUGGESTION

B. Necula, G. Roebling, "Reflections on Introducing Artificial Intelligence Tools in Support of Anti-Fraud", 2024, Vol. 19(3), eucrim, pp206–214. DOI: https://doi.org/10.30709/eucrim-2024-019

Published in
2024, Vol. 19(3) eucrim pp 206 – 214

ISSN: 1862-6947

https://eucrim.eu

CC BY ND

I. Introduction

With the occasion of OLAF's 25th anniversary, the year 2024 has given us the opportunity to look back on the evolution of the European Anti-Fraud Office (OLAF) over the last quarter of a century through the prism of the Office's digital transformation.[1] The present article will complement that retrospective with a timid glimpse into the digital future.

Today, we can safely assume that new tools based on large language models and other artificial intelligence-based software are set to play an increasing role in many modern administrations in the future, including in the anti-fraud domain. One would even be tempted to say that the prevention, detection, and investigation of fraud and associated illegal activities, which today involve processing and analysing an ever-growing volume of data of different types, are uniquely suited to the strengths of such tools of artificial intelligence (AI). As we are prudently embarking on this journey ourselves, the purpose of this article is to share some of our own reflections and observations.

As promising as the potential of AI is without doubt for anti-fraud work, it is not always straightforward for public authorities to practically harvest this potential. There are many issues authorities need to come to terms with when it comes to practical implementation, three of which stand out. Addressing these issues decisively is likely to be key to the success of any such initiative.

First, public authorities need to identify for which anti-fraud-specific functionalities, or "use cases", in line with their own mandate they want to deploy an AI tool. To this effect, they need to conceptually link the strengths of AI tools to the specific requirements of making anti-fraud investigations more efficient and more effective. In other words, investigators and technical staff have to be on the same page. Authorities then also have to match and adapt existing AI technology to map the resulting use cases, which is likely to require some additional technical enhancements (such as fine-tuning and prompt engineering). They would also have to ensure adequate protection of confidentiality of any data handled, as required by the use case at hand. Section II below offers some initial thoughts on these conceptual foundations for any anti-fraud engagement with AI.

Second, public authorities will of course need to be scrupulous in ensuring compliance with the legal framework. The use of AI tools, especially in a context as sensitive as anti-fraud prevention and investigation, raises important ethical issues, even if the AI tool will always be limited to a mere support role. An effective protection of the rights of citizens, including notably those enshrined in the Charter of Fundamental Rights of the European Union, is imperative.[2] The legal framework has recently evolved with the adoption of the EU's AI Act.[3] Having that act now in force since 1 August 2024 is an important step forward in terms of legal certainty when deploying AI.[4] At the same time, some of the terms used in the AI Act are novel, and certain concepts are still to be fleshed out further by implementing and delegated acts and guidance. In addition, authorities wishing to deploy AI tools to support their anti-fraud work will need to be mindful of the applicable data protection regime – in the case of OLAF Regulation 2018/1725.[5] Some of the regulatory cornerstones of the emerging legal framework for AI tools relevant for anti-fraud work are summarised in Section III below.

Third, public authorities must check the – internal or external – availability of the relevant technical skills to carry out AI projects. This aspect may well influence the degree to which an anti-fraud authority engages with AI. We will not further explore the practical challenges linked to the availability of skills in this article. At this point, we would just like to mention the fact that OLAF, on behalf of the European Commission, annually awards grants to national authorities to build up their anti-fraud capacities to protect the Union's financial interests, in implementation of the Union Anti-Fraud Programme. Supporting Member States' digital capabilities is a stated priority,[6] which would naturally include building up AI expertise.

II. The Potential Use of AI Tools for Anti-Fraud Work

The dramatic leap forward in AI development in recent years has been transforming many industries, and its potential to revolutionize the anti-fraud domain is equally evident. AI developments could considerably facilitate certain steps in fraud prevention, detection, and investigation, particularly those that require an analysis of large volumes of data. Moreover, the power of AI tools cannot only make anti-fraud work more efficient, but also more effective. For example, AI tools may well pick up certain patterns in large data sets which can easily escape the human eye.

The following outlines some potential use cases of AI for anti-fraud preventive and investigative work from the perspective of natural language processing and image analysis. There will be a particular focus on how these technologies leverage large data sets to improve investigations.

1. Potential AI scenarios for anti-fraud work

One of the primary ways in which large language models (LLMs) can assist is through the analysis of text-based data. When pursuing anti-fraud investigations, investigators often deal with an enormous volume of text, including forensically acquired media, financial records, communication records, open sources data, and project-related documentation. As it stands, LLM technologies can contribute to automating the analysis of this data, extracting key information, identifying trends, and flagging suspicious communication. However, such analysis will have to be carefully reviewed by investigators in all cases for the reasons explained in section 2.b.

Considering an investigation's timeline, there are two main activities that define the world of anti-fraud: a) the prevention and pro-active detection of fraud and b) the reactive part, which is the actual investigation.

a) Use cases in the field of prevention

From a technical perspective, preventive tasks are dominated by risk analysis – a field in which advanced AI is already making good progress and is actively being tested by many software vendors. The risk analysis domain is technically quite complex due to the challenges surrounding data availability and the number of variables to be taken into account; hence, having AI assistance could generate additional insights.

Risk analysis on its own is already a conceptual challenge, simply when it comes to deciding on the scoring and the weights assigned to each risk and the calculations for the overall system. Here, the new AI technology can come into play by adding an understanding of qualitative risks. Furthermore, in light of the latest developments (especially the agentic approaches in which AI systems can carry out certain technical tasks autonomously, with minimal human intervention), a promising avenue would seem to be to test risk scoring systems in an automated manner with the help of **agentic systems**. This potential implementation presents the opportunity to run multiple risk approaches and, based on known true positives, to decide on the efficiency of the system.

Moreover, the field of prevention also includes verification of deliverables. In many cases, project deliverables are documents. Until now, the focus of these checks has been mostly on **plagiarism**, which is a complex issue. With the advent of generative AI, it has become easier for ill-intended individuals to alter text; as a consequence, traditional plagiarism checkers that focus on similarity will fail in flagging potentially copied

https://doi.org/10.30709/eucrim-2024-019

texts. However, the same tools that serve the fraudster can be used to apply detection and indicate text similarity approximation.

b) Use cases in the field of investigation

From an investigative perspective, the use cases that benefit from advanced AI utilisation are already much clearer and well formulated.

For example, AI can be used to sift through numerous elements in forensically acquired media for **keywords or phrases indicative of fraudulent intent**. By processing large volumes of text, in combination with various helper techniques, LLMs can spot anomalies or unusual patterns of communication that may signal criminal intent. Additionally, AI-driven text analysis tools can be used to identify connections between seemingly unrelated elements. For instance, by analysing language and terminology used in certain email content, AI systems may discover **patterns in the modus operandi** of fraudsters.

Another potential application of AI in text analysis is **automated summarisation**. By using AI tools, investigators can generate summaries of large reports, saving valuable time in reading and analysing documents. For example, investigators are enabled to quickly review summaries of investigation reports, witness statements, or intelligence analysis reports, allowing them to focus on verification and decision-making – rather than the manual task of reading lengthy documents to extract relevant information. This can significantly enhance the speed of investigations and response times, especially when trying to gain an overview of the state of a case.

Object detection systems are also becoming increasingly sophisticated, allowing AI to **identify and track items**. As an example, customs is facing significant challenges in building efficient analytics for the quick aggregation of various data that appears in a normal customs workflow. It is standard for a customs investigation to deal with customs declarations, either in digital or scanned formats, images of containers and lorries, images of the contents of the containers, etc., on a regular basis. In many instances, this wealth of data must be aggregated and queried for an efficient investigation. By exploiting machine learning and optical character recognition, users can extract some information available in these images in some of the situations.

Another use case, also part of the challenges related to vision, is using **geo-located data**, such as aerial images of places of interest. AI models are becoming more and more efficient at identifying the typology of images and thus facilitating comparison between existing labelled data sets and the image of interest. One of the most relevant benefits is that AI-based object/area recognition greatly reduces the human effort and potentially the number of false positives for manual review.

Financial transaction analysis is another domain that AI may impact in a significant manner. Data sets of hundreds of thousands of lines of transactions appear to be the ideal environment for AI, with the purpose of identifying fraudulent behaviour. In everyday work, an analyst would have numerous tools and methods available to sift through these data sets and try to pinpoint financial flows, anomalous transactions, matching amounts, relevant details within a transaction description, etc. Thus, LLMs might not be the first tool designed to handle financial transactions. However, initial results in this field indicate that AI capabilities could be of great benefit, especially when dealing with the transaction description[7] from a natural language understanding perspective.

Last but not least, the **pre-processing and visualisation of data** is one of the biggest daily challenges of many operational intelligence analysts. LLMs can significantly enhance tasks such as entity recognition, entity resolution, co-reference resolution, and building network graphs, which are critical in complex data analysis for fraud investigations. **Entity recognition** involves identifying key entities like people, organisations, and locations within unstructured text. **Entity resolution** is the process of determining whether different mentions refer to the same real-world entity, which is especially useful in fraud investigations where names or identifiers may vary across data. The term **co-reference resolution** involves linking different mentions of the same entity within a text (e.g., resolving “he” or “the company” to the correct entity), allowing for a more coherent understanding and tracking of entities across documents. Once entities and their relationships have been identified, LLMs can assist in constructing network graphs that visually represent the connections between entities. These graphs enable investigators to uncover hidden relationships, visualise fraud patterns, and detect suspicious networks more effectively.

2. Challenges, limitations, and potential solutions

The previous section only sketched out some of the possible ways in which AI tools are likely to support anti-fraud prevention and investigations in the near future. Many more use cases will almost certainly appear over the coming months and years. Yet as tempting as the power of these AI tools will be for many anti-fraud authorities struggling with scarce resources, employing this technology also has limitations, such as notably the imperative to systematically and critically review the AI output by humans. This section explores some key challenges and limitations whilst at the same time attempting to point to potential solutions.

a) One of the most critical aspects of using AI in investigations, especially when working with LLMs for tasks like text analysis, is **prompt engineering**. This term refers to the process of designing specific inputs or prompts that guide AI models, particularly LLMs, to produce desired outputs. In practice, the concept of prompt engineering involves understanding how to effectively communicate with AI models to generate accurate, relevant, and context-specific outputs. To develop skills in prompt engineering, agencies may focus on understanding the AI model’s capabilities and limitations – i.e., how it works, what data it was trained on, etc. – iterative testing and comparing the results, and researching prompt libraries and tools.

To enable successful prompt engineering in the context of an investigation, it is also important that the AI tool is familiar with **domain-specific language**. For example, as mentioned, AI might be used to summarise various documents, such as intelligence analysis reports. However, the quality and relevance of the output depend heavily on how the input data is framed. If the prompts are not carefully constructed, the AI system might produce misleading or irrelevant results.

One of the key challenges of prompt engineering is ensuring that LLMs can understand and process the nuances of specific language. Data often contains jargon, abbreviations, or domain-specific terms (e.g. procurement), that may not be easily interpretable by AI models without specific contextual guidance. Moreover, both commercially available and open source LLMs are trained on general data sets and might not fully comprehend the domain-specific knowledge required for anti-fraud investigations.

To overcome this, prompt engineering requires deep collaboration between AI developers and professionals in the field. For instance, a well-engineered prompt might ask the AI tool to summarise reports by focusing on specific details like fact descriptions, modus operandi, or location. If designed correctly, prompt engineering can guide LLMs to provide accurate and contextually relevant insights.

b) Another major issue with LLMs, especially when applied to specialised fields like investigations, is the phenomenon of “**hallucinations**”. This term refers to instances where AI models generate plausible-sounding but inaccurate or entirely fabricated information. For an investigation, relying on inaccurate data could have serious consequences. Hallucinations in LLMs arise because these models are often trained on broad data sets that do not always include the specific, factual information required for legal or investigative tasks. As a result, when asked to generate text based on prompts, the model might "fill in the gaps" with information that sounds reasonable but is not grounded in reality. As a consequence, we need to be cautious when using

LLMs, ensuring that AI outputs are always verified by human experts to avoid the risks associated with incorrect information.

c) One emerging technique that helps mitigate some of the limitations of LLMs is **retrieval-augmented generation (RAG)**. RAG is a hybrid approach that combines the generative capabilities of LLMs with retrieval-based methods. In this system, instead of relying solely on the AI's pre-trained knowledge, the model first retrieves relevant information from a structured database or external knowledge source before generating a response.

This approach is particularly useful for anti-fraud tasks, where accurate and up-to-date information is crucial. For instance, instead of relying on the LLM to generate an answer from general knowledge, RAG-enabled systems are able to first retrieve relevant data from internal databases. AI then uses this specific information to generate a more accurate and contextually informed output. This minimises the risk of hallucinations and enhances the reliability of AI-generated insights.

d) **Reinforcement Learning from Human Feedback (RLHF)** is another emerging approach that combines traditional reinforcement learning with direct human input to improve the behaviour and performance of AI systems. This technique allows AI models, particularly LLMs, to learn more effectively from human preferences, judgments, and corrections, leading to more aligned, accurate, and user-friendly outputs. RLHF is especially valuable in areas where human interpretation, ethics, or nuanced decision-making play a critical role, making it a key tool in refining AI systems for real-world applications.

At its core, reinforcement learning (RL) involves training an AI agent by rewarding desired behaviours and penalising undesirable ones. In RLHF, humans play an active role by providing feedback in the form of rewards or corrections to guide the AI model's learning process. Instead of relying solely on predefined rewards from a static environment, RLHF allows humans to directly assess the outputs of AI and intervene when AI produces incorrect, unethical, or suboptimal results. This human feedback becomes a part of the reward mechanism, refining AI in its actions and decisions over time.

RLHF addresses several challenges that traditional AI training methods face, particularly in areas where objective measures of success are difficult to define. For example, in language models, it can be hard to quantify what constitutes a "good" response, as quality often depends on context, tone, and user intent. Human feedback provides the nuance that purely automated systems might lack. In practical terms, human annotators may review AI outputs and rank them based on quality or relevance, enabling AI to adjust its future responses based on this feedback. This iterative process continues until the AI system becomes more aligned with human expectations.

e) Although not strictly connected to advanced AI, the **security of data** manipulated in an AI framework should continue to be a top concern for practitioners. Many of the existing tools employ API (Application Programming Interfaces) and services in clouds to serve AI-generated content to users. In general, the terms of use can bring some piece of mind to concerned users. However, the general recommendation whenever such tools are used for investigative purposes is to build systems in protected environments, ideally segregated from the internet and with models and software that can be installed locally without additional resources.

f) Apart from the technical aspects, anti-fraud authorities planning to engage with AI may also wish to, from the outset, reflect on how to deal with **staff attitudes** towards this new technology. An informal (and not necessarily representative) survey at a recent conference with anti-fraud practitioners from the Member States and the Candidate Countries showed that the attitudes of those present fell into two groups of comparable size: Whilst respondents in one group highlighted the potential and benefit of AI for anti-fraud work, another group had reservations about such AI use, notably on account of privacy and ethical concerns. Some respondents were also wondering how AI would affect their current job.

Authorities may thus consider developing a training strategy to upskill staff as well as a parallel one on communication and awareness raising to pro-actively engage with staff on their legitimate questions and concerns. And of course, since key components of the emerging regulatory AI framework are precisely designed to address some of those questions, attention to full regulatory compliance may be a part of the answer.

III. Key Elements of the Emerging Regulatory Framework

This section explores some of the basic regulatory parameters which govern the use of AI by public authorities in the anti-fraud domain today. Adhering to these parameters is a precondition of deploying AI tools in full compliance. But in addition their existence may also in itself influence which AI use cases an authority may wish to pursue based on a cost-benefit analysis.

As explained in the AI Act, the use of AI systems by law enforcement raises particular concerns. This is notably due to what the Union legislator perceives as a power imbalance, and on account of the grave consequences that law enforcement action can have, such as surveillance, arrest, or the deprivation of a natural person's liberty.[8] In law enforcement, any possible discriminatory or in other ways unethical bias on the part of an AI tool could lead to unacceptable outcomes. Moreover, the use of an AI tool – with its autonomously generated, not totally predictable outcomes – is inevitably somewhat at odds with a law enforcement context where, according to Recital 59 of the AI Act, "accuracy, reliability and transparency is particularly important to avoid adverse impacts, retain public trust and ensure accountability and effective redress."

1. The AI Act

To address these concerns, the AI Act introduces certain substantive and procedural guardrails. It is designed to improve the functioning of the internal market by laying down a uniform legal framework for the development, the placing on the market, the putting into service, and the use of AI systems in the EU in accordance with its values, and to promote the uptake of human-centric and trustworthy AI whilst ensuring a high level of protection of health, safety, and fundamental rights.[9]

To achieve these objectives, the regulatory approach taken in the AI Act is reminiscent of the risk-based regulatory layers familiar from product safety rules (the pyramid-shaped "hierarchy of hazard controls") that apply to some categories of goods placed onto the internal market. In this spirit, the AI Act in essence distinguishes between the following:

- The most harmful AI practices, which will be prohibited (Art. 5);
- High-risk AI systems to which rather stringent regulatory requirements apply (Art. 6(2) in combination with Annex III); and
- Less risky AI systems which remain largely unregulated.

a) Application of the AI Act for anti-fraud projects

The first question that needs to be clarified is of course whether an envisaged AI project that would support fraud prevention or investigation would actually fall into the scope of the AI Act.

(aa) *De ratione temporis*, the AI Act has been in force since 1 August 2024. However, its main provisions will only be **phased in progressively**: the prohibitions set out in Art. 5 will apply as of 2 February 2025[10], and the rules on high-risk AI systems referred to in Art. 6(2) in combination with Annex III only apply as of 2 August 2026[11]. High-risk AI systems already on the market prior to that date will in principle only have to comply with the AI Act if they are subject to significant changes in their designs.[12] However, public authorities that are deploying AI tools that were on the market before the cut-off date will nevertheless have to comply with the AI Act by 2 August 2030 at the latest.[13]

(bb) Today, many anti-fraud authorities already deploy a variety of analytical tools that operate on the basis of advanced algorithms, for example for fraud detection. This can sometimes give rise to doubts as to whether those systems would – possibly retroactively – fall under the AI Act. It is therefore important to delineate its scope of application *de ratione materiae* as well. Art. 3(1) of the AI Act contains the relevant definition in that regard: The Act applies, as a matter of principle, only to **machine-based systems which infer**, from the input they receive, how to generate output such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. The meaning of the decisive key term “infer”, which arguably suggests some degree of *autonomous* output generation, will without a doubt be further elaborated in the future.

(cc) It should also be noted that **any research, testing, and development** activity regarding AI systems before these are put into service do not fall into the scope of the AI Act, as long as no testing under real-world conditions is undertaken (e.g., experimenting with live data from a database).[14] Special rules, including a pre-authorisation or registration process, apply where the testing of high-risk AI systems is carried out under real-world conditions.[15] The subjects of such testing should also give their informed consent prior to the tests.[16]

b) Prohibition of a project?

If an AI tool to be developed were to, as a matter of principle, lie within the scope of the AI Act, it is of course imperative to ascertain early on whether such a tool would fall into the **prohibited categories** set out in Art. 5 of the AI Act (see above). For the present purposes, the prohibited practice which arguably comes closest to typical anti-fraud work concerns an AI-based assessment of the risk of natural persons committing a criminal offence.[17] However, that clause only applies if two conditions are fulfilled: (i) where the assessment is based solely on the profiling of a natural person, and (ii) where the AI system is not only used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity.

Prima facie, many of the risk analysis systems operated by anti-fraud authorities to detect expenditure or revenue fraud would not typically meet these conditions. In particular, in many cases those systems do not focus on natural persons, but on undertakings. In addition, it is difficult to imagine that these systems would be based exclusively on the profiling of a natural person. Moreover, they usually link their evaluation to objective and verifiable (but not necessarily verified) facts, such as previous infringements, or suspicious shipping routes. What is more, the assessment of whether a person is ultimately involved in a criminal activity will always be reserved for a human being, and never be automated – therefore the second of the two conditions above would not be met. Last but not least, Recital 42 of the AI Act adds further clarity in that regard: According to this section, the prohibition does not apply to AI systems using (i) risk analytics to assess the likelihood of financial fraud by undertakings on the basis of suspicions transactions, or (ii) risk analysis tools to predict the likelihood of the location of narcotics or illicit goods by customs authorities, for example on the basis of known trafficking routes. Against this background, the prohibitions of the AI Act should not typically apply to the well-established risk analysis systems operated by many agencies (if ever those systems were to be classified as AI tools based on their advanced features; see above).

c) High-risk project?

The regulatory requirements applicable to AI tools for anti-fraud purposes will then depend on whether a **high-risk** classification pursuant to Art. 6(2) of the AI Act is warranted. This provision refers to several specific categories of AI use cases set out in Annex III, which the Union legislator, in principle, deemed to present a higher risk. For the present purposes, Point 6 in Annex III dealing with the law enforcement area is the most relevant.

aa) Point 6 Annex III refers to certain activities by **law enforcement** authorities.

(1) The AI Act defines these authorities, as far as this article goes, as any public authority competent for the prevention, investigation, detection, or prosecution of a criminal offence or the execution of criminal penalties.[18] It is reasonable to assume that this **definition** focusing on criminal offences does not cover mere administrative authorities. This view is, in our opinion, supported by Recital 59, which clarifies that AI systems specifically intended to be used for “the administrative proceedings by tax and customs authorities” are not to be classified as high-risk AI systems. It should also be noted that Point 6 is not limited to law enforcement authorities, but equally addresses AI systems intended to be used by Union institutions, bodies, offices, and agencies supporting law enforcement authorities.

(2) Point 6 of Annex III AI Act goes on to categorise a number of AI systems with specific functionalities as high risk. These notably concern AI systems

- “to evaluate the **reliability of evidence** in the course of the investigation or prosecution of criminal offences” (Point 6c);
- “for assessing the **risk of a natural person offending or re-offending** not solely on the basis of the profiling of persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups” (Point 6d); or
- “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” (Point 6e).

As it is still early days, it is difficult to predict to what extent these categories will be practically relevant for the AI use cases which anti-fraud authorities may be considering at some point in the future. Suffice it to say that, first of all, from today’s perspective it is not easy to envision an AI system evaluating the reliability of evidence, but of course technologies are developing fast. Secondly, it needs to be underlined that the other two categories are limited to the profiling of natural persons for which the unlikelihood of relevance for anti-fraud AI tools has been already mentioned above under point 1b).

bb) However, even where an anti-fraud AI project to be evaluated could *prima facie* fall into one of the aforementioned three categories under point 6 of Annex III, the AI Act adds an important derogation of practical relevance: Pursuant to Art. 6(3), AI systems which perform certain types of **ancillary tasks** are not to be considered high risk. This relates in particular to AI systems intended to (i) perform a narrow procedural task, (ii) improve the result of a previously completed human activity, or (iii) perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III. However, before putting an AI system which the provider has concluded to not be high risk due to its ancillary nature into service, law enforcement authorities need to register it in a secured EU database.[19]

cc) Classifying an AI system as high risk would have important further regulatory consequences. **Regulatory requirements** for high-risk AI systems are set out, notably, in Arts. 8 to 27 AI Act. They include, for example, the need to establish a risk management system (Art. 8), to draw up technical documents (Art. 11), and to

https://doi.org/10.30709/eucrim-2024-019

keep records (Art.12). In addition, transparency obligations (Art. 13) and obligations for facilitating human oversight (Art. 14) need to be fulfilled. Under certain conditions, a fundamental rights impact assessment will also need to be carried out (Art. 27). Many public authorities are set to carefully examine the expected costs and benefits which deploying a high-risk AI system would entail. However, it is beyond the scope of this article to provide details of these requirements.

2. Data protection rules

Next to the necessary compliance with the AI Act, the use of AI tools for anti-fraud purposes must also adhere to the **applicable data protection regime**.[20] In the case of OLAF, this would be Regulation 2018/1725[21], applicable to EU institutions, bodies, offices, and institutions. It is aligned to similar provisions in the General Data Protection Regulation (GDPR)[22].

a) Application of the data protection regime and overlap with the AI Act

The data protection rules naturally only apply to the extent that personal data is actually processed by an AI tool. This means that where an AI tool is deployed using data sets not containing such personal data (for example, container numbers, or vessel movements, as long as those elements cannot be linked to a specific person),[23] the processing is out of **scope** of the applicable data protection regulation.[24]

On occasion there may be some **functional overlap** between the requirements of the applicable data protection rules and the AI Act. For example, where a data protection impact assessment needs to be carried out,[25] that analysis may in part address similar issues as those required as part of the Fundamental Rights Impact Assessment under the AI Act (see above 1 cc)). Likewise, the need for a data protection impact assessment depends on whether the processing of personal data as part of AI use is likely to result in high risks to the rights and freedoms of natural persons, taking into account the nature, scope, context, and purposes of the processing. The EU institutions would base their assessment on the Guidance and template for threshold assessment provided by the European Data Protection Supervisor.[26] It remains to be seen whether, in making that assessment, they might take into account the Union legislator's choice to exempt some ancillary AI from being considered high risk, pursuant to Art. 6(3) of the AI Act (see above 1 bb)).

b) Implementation of key data protection principles

Given that this article can only outline the potential use of AI tools and the connected challenges in the anti-fraud area, and given the complex matter, this article cannot exhaustively discuss the application of the EU data protection regime to AI use by anti-fraud authorities. Hence, we wish to limit ourselves to highlighting certain **key principles** underpinning the applicable data protection regime, which should also be implemented when using AI for the kind of anti-fraud purposes described above.

First of all, when developing and deploying AI tools, it is essential to ensure that the processing of personal data is lawful, fair, and transparent.[27] **Lawful processing** requires that the anti-fraud authority has a valid legal basis for the processing of personal data, and that the personal data is collected for specified, explicit, and legitimate purposes and not further processed in a manner that is incompatible with those purposes. The processing must also be necessary for the performance of the task of the anti-fraud authority.[28] In addition, the authorities must implement appropriate technical and organisational measures to ensure the security and confidentiality of the data, including the use of encryption and access controls.

Anti-fraud authorities must be **transparent** about the use of AI tools in the processing of personal data. This includes providing clear information to individuals about the use of AI tools, the types of data being

https://doi.org/10.30709/eucrim-2024-019

processed, and the purposes of the processing. Individuals must also be informed about their rights, including the right to access, rectify, and erase their personal data.

Anti-fraud authorities using AI for purposes that involve personal data need to be mindful of the **data minimisation** principle.[29] When looking at the illustrative AI use cases presented in Section II above, limiting the exposure of personal data to the AI tool to only a small sub-set of data (e.g., one case file only), rather than a whole database, could be one of the possible means of implementing the data minimisation principle. Such a limitation, however, must be compatible with the intended use case.

Anti-fraud authorities will naturally also be very mindful of the fact that in the context of AI use, personal data is processed in a manner that ensures appropriate **security** of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction, or damage, using appropriate technical or organisational measures.[30] In particular, it can reasonably be expected that anti-fraud authorities would not normally work with internet-based AI tools created by third parties when confidential information, including personal data, is involved; instead, they would operate their AI tool in a more secure IT environment. In addition, the usual access control limitations familiar from the general IT system will often need to be applied.

Moreover, AI tools must not become a way to undermine **data access policies** based on a need-to-know principle by allowing the accidental or intentional disclosure via an AI output of data to which a user would not normally have access.

Since compliance with data protection rules is of fundamental importance to anti-fraud authorities planning to use AI on data sets containing personal data, they are well-advised to integrate this dimension into the design of their AI system right from the start (**data protection by design**).[31] The data minimisation and confidentiality principles mentioned previously are possible elements in such a design approach. Another possibility may be to focus on the design of the input interface. Where users of an AI tool can engineer prompts as they wish, there is always the hypothetical possibility that a rogue user might abuse the power of the AI tool for purposes not compatible with the mission of the public authority. Such abuse can be largely eliminated with a different design, in which the system administrator configures the user interface in such a way that only pre-defined prompts are available to regular users.

IV. Conclusions

The field of AI is developing at a fast, not to say furious, pace. New models with substantially expanded capabilities are being released by the major providers several times a year. Keeping up with these developments is a challenge to all actors, so there will inevitably always be some element of learning by doing.

Anti-fraud authorities are working with limited resources whilst the data volumes they have to deal with are growing exponentially. The processing power of especially the latest generative AI tools give hope that they can help authorities to stay on top of the game. To harvest this potential, authorities will, however, have to invest in the technical and intellectual infrastructure, i.e. to build up the relevant technical and user expertise. OLAF has begun supporting national authorities on this challenging but promising trajectory as concerns the protection of the Union budget.

At the same time, anti-fraud authorities need to be mindful of the limitations and constraints of AI tools. This applies both from the perspective of the inherent technological limitations of such tools (such as potential bias and hallucinations), and from a privacy perspective. For these reasons, it is clear that AI tools will always be limited to a support role in anti-fraud prevention and investigation. The objective of the prudent use of AI by anti-fraud authorities must be to render the decision-making of human anti-fraud investigators more efficient and effective, and never to replace it.

---

1. See eucrim issue 4/2024 (forthcoming).↩
2. See also, from a general law enforcement perspective, the discussion of these matters in Europol, *AI and policing - The benefits and challenges of artificial intelligence for law enforcement,* 2024, available at: <https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf> accessed 9 December 2024.↩
3. 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 certain Regulations and Directives (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024.↩
4. See D. Hadwick, "'Error 404 – Match not found' – Tax Enforcement and Law Enforcement in the EU Artificial Intelligence Act", (2023) *eucrim* 55-60.↩
5. Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, 39.↩
6. See for example the 2024 Call for proposals for the Union Anti-Fraud Programme (EUAF), available at: <https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/euaf/wp-call/2024/call-fiche_euaf-2024-ta_euaf-2024-trai_en.pdf> accessed 9 December 2024.↩
7. The financial transaction typically has a description field that, from a data perspective, is a free text field. It usually contains details of the transactions, in many cases with valuable insights such as invoice number, contract reference, explanation for the payment, etc.↩
8. See Recital 59 AI Act. For the concerns, see also D. Kafteranis, A. Sachoulidou, and U. Turksen, "Artificial Intelligence in Law Enforcement Settings – AI Solutions for Disrupting Illicit Money Flows", (2023) *eucrim*, 60-66, in particular Chapter IV.↩
9. See Recital 1 AI Act.↩
10. Art. 113(a) AI Act.↩
11. Art. 113 AI Act.↩
12. Art. 111(2) AI Act.↩
13. Art. 111(2) AI Act, last sentence.↩
14. Art. 2(8) AI Act.↩
15. See Art. 60(4) AI Act.↩
16. Art. 61(1) AI Act.↩
17. Art. 5(1)(d) AI Act.↩
18. See the legal definition in Art. 3(45) AI Act.↩
19. See Art. 49(4) AI Act.↩
20. Art. 2(7) AI Act.↩
21. *Op. cit.* (n. 5).↩
22. 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), OJ L 199, 4.5.2016, 1.↩
23. See for example ECJ, 9 November 2023, Case C-319/22, *Autoteile-Handel*, para. 45 with reference to the earlier judgment of 19 October 2014, Case C-582/14, *Breyer*, paras. 43 *et seq.*↩
24. See Art. 2(1) Regulation 2018/1725.↩
25. See Art. 39 Regulation 2018/1725.↩
26. See European Data Protection Supervisor, "Data Protection Impact Assessment (DPIA)" <https://www.edps.europa.eu/data-protection-impact-assessment-dpia_en> accessed 9 December 2024.↩
27. Art. 4(1)(a) Regulation 2018/1725.↩
28. Art. 5(1)(a) Regulation 2018/1725.↩
29. Art. 4(1)(c) Regulation 2018/1725.↩
30. Art. 4(1)(d) Regulation 2018/1725.↩
31. Art. 27(1) Regulation 2018/1725.↩

* Authors statement

This article only reflects the authors' personal opinions and cannot be attributed to the Institution that employs them.

COPYRIGHT/DISCLAIMER

© 2024 The Author(s). Published by the Max Planck Institute for the Study of Crime, Security and Law. This is an open access article published under the terms of the Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0) licence. This permits users to share (copy and redistribute) the material in any medium or format for any purpose, even commercially, provided that appropriate credit is given, a link to the license is provided, and changes are indicated. If users remix, transform, or build upon the material, they may not distribute the modified material. For details, see https://creativecommons.org/licenses/by-nd/4.0/.

Views and opinions expressed in the material contained in eucrim are those of the author(s) only and do not necessarily reflect those of the editors, the editorial board, the publisher, the European Union, the European Commission, or other contributors. Sole responsibility lies with the author of the contribution. The publisher and the European Commission are not responsible for any use that may be made of the information contained therein.

About eucrim

eucrim is the leading journal serving as a European forum for insight and debate on criminal and "criministrative" law. For over 20 years, it has brought together practitioners, academics, and policymakers to exchange ideas and shape the future of European justice. From its inception, eucrim has placed focus on the protection of the EU's financial interests – a key driver of European integration in "criministrative" justice policy.

Editorially reviewed articles published in English, French, or German, are complemented by timely news and analysis of legal and policy developments across Europe.

All content is freely accessible at https://eucrim.eu, with four online and print issues published annually.

Stay informed by emailing to eucrim-subscribe@csl.mpg.de to receive alerts for new releases.

The project is co-financed by the Union Anti-Fraud Programme (UAFP), managed by the European Anti-Fraud Office (OLAF).

Co-funded by
the European Union