Probation services in light of the European Artificial Intelligence (AI) Act

An appreciation of the impact
for The Confederation of European Probation - April 2024

Ministry of Justice and Security (Government of the Netherlands)
Directorate-General for Sanctions and Protection (DGSenB)
Department Artificial Intelligence
Contact: ai@minjenv.nl

What is the AI Act?

The Artificial Intelligence Act (AI Act) is the first comprehensive legislative framework that regulates AI systems that are placed on and used in the European market.[1] The aim of the new rules is to foster trustworthy AI in the EU and beyond, by ensuring that AI systems respect fundamental rights, safety, and ethical principles by addressing risks of impactful AI systems. The framework contains rules for multiple entities, such as the developers and deployers of AI systems. These could be either private or public entities.

The AI Act introduces a risk-based approach that classifies AI systems into different risk categories. The more risk is associated with the AI system, the more obligations that will apply. The AI Act roughly categorises AI systems into three categories: 1) *prohibited* AI systems, 2) *high-risk* AI systems and 3) AI systems with a *limited risk*.

How does the AI Act impact the work of probation?

Within probation, a significant amount of the work involves assessing the likelihood of individuals committing offences or repeating certain behaviours, alongside offering tailored guidance to mitigate the risk of future offences. To assist these assessments and recommendations, currently algorithmic systems are being used. Some see opportunities to responsibly deploy machine learning AI systems for these assessments.

Under the AI Act, certain AI systems are classified as *high risk*, subjecting their development, deployment and use to stringent criteria of review and documentation. This includes, e.g., the implementation of a risk management system, conducting bias assessments, ensuring robust cybersecurity measures, and performing assessments on the impact on fundamental rights.

AI systems that are intended to be used by law enforcement authorities or relevant bodies for assessing the risk of a natural person of (re-)offending – that is not solely based on profiling or to assess personality traits and characteristics – are classified as *high risk* under the AI Act.

In addition to this, individual risk assessments for criminal offences based solely on profiling or on personality traits or characteristics, are classified as a *prohibited* AI system under the AI Act. This prohibition shall not apply to AI systems 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. No evident examples were found so far in the Netherlands that would fall within the definition of this prohibition.

The Dutch Forensic Care Quality programme (KFZ) and the University of Twente explored the possibilities of AI within the Dutch probation services (3RO).[2] The goal of this research was to provide the probation services with a better understanding of what AI is, how it can be responsibly deployed and how probation service workers can work with AI systems. KFZ

[1] The AI Act is developed around the EU New Legislative Framework (NLF) that outlines the general structure that pieces of EU product legislation follow, and the tools new legislation has at its disposal.
[2] https://kfz.nl/projecten/ai-binnen-3ro-toepassing-en-toekomst

concludes that there are possibilities for the probation in using AI systems that go beyond risk taxation instruments, such as unburdening employees in processing information. The research takes a first step in building a new framework on how AI should be applied within probation. It is therefore useful to examine the AI Act in a broader perspective of probation services than only focusing on risk assessment tooling.

Example of a high risk AI system within the domain of probation

Deploying a quantitative recidivism risk assessment is an example of a high risk AI system. These assessments frequently utilize machine learning or statistical methods. Machine learning is more focused on forecasting outcomes or making decisions from sample data (training data), autonomously identifying and learning the patterns through different algorithms, whereas statistical methods rely on predefined mathematical principles to discern certain patterns in data as given by the researcher.

Dutch example of a high risk algorithmic system within the juvenile justice system

In the juvenile justice system, the 'National Set of Instruments Juvenile Justice System' (in Dutch: het LIJ (Landelijk Instrumentarium Jeugdstrafrechtketen)) supports chain partners in collecting and analysing information about young offenders and their living environment. Using the instruments in the LIJ, the chain partners receive and give each other insight into what is needed to prevent recidivism under young offenders. The goal of the LIJ is to lead them to appropriate interventions. The LIJ is based on the principles of the Risk-Needs-Responsivity (RNR) model.

Part of the LIJ is the *Ritax*: an actuarial risk assessment instrument. With the instrument, the risk and protective factors of and any other signs of concern for juvenile suspects are mapped out, which are based on scientific research about factors that increase or decrease the risk of recidivism of criminal behaviour. The outcome of the Ritax is a Dynamic Risk Profile (DRP). The DRP provides an overview of the dynamic risk factors that influence a juvenile suspects' likelihood of re-offending. For each domain, a score is calculated that expresses the severity of the risk factors within this domain - and thus the relationship with the likelihood of recidivism. The LIJ provides automated suggestions for appropriate behavioural interventions, which the professional working with the juvenile in question can adhere to or can deviate from if necessary. In summary, with the Ritax, information is collected and analysed to support decision making with regard to achieving an appropriate intervention, realising risk management measures to protect the young offender against others and the offender self, and realising care for the young offender that may be additionally needed.

The Ritax is largely based on the Washington State Juvenile Court Assessment (WSJCA) that was developed in 2004 in the United States of America (Van der Put, 2021). The Ritax was updated based on a standardization study among the Dutch population. The Ritax was revised and improved in 2021 based on a Dutch standardization study by Van der Put et al (2021). The updated Ritax is put into use by council investigators of the Child Care and Protection Board and youth probation staff.

The Ritax is considered an 'algorithmic system' by the Netherlands Court of Audits, following an accountability audit of the Ministry of Justice and Security in 2022. In this audit, the Netherlands Court of Audits tested the Ritax using the Courts review framework for algorithms (in Dutch: 'Toetsingskader Algoritmes').

A crucial question is if and when these kind of algorithmic systems need to be considered high-risk AI systems under the AI Act, as they operate under an (albeit complex) 'if this-then-that-paradigm'. Recital 12 could be interpreted to imply that these are not AI-systems following the definition of the AI Act.

Further guidance under way

A European *Artificial Intelligence Office* will be established within the European Commission.[3] The AI Office plays a key role in the implementation of the AI by, i.a., preparing guidance and guidelines, implementing and delegated acts, and other tools to support effective implementation of the AI Act and monitor compliance with the regulation.[4] As a result, further guidance on the definition of AI systems is under way as it is not yet clear enough.

Nonetheless, the general political opinion in the Netherlands is that these algorithmic systems, as has been described above, have impact on citizens. Therefore, the position of the government of the Netherlands is that these impactful systems, where relevant, need to adhere to the provisions under the AI Act, for instance, through performing a fundamental rights impact assessment (FRIA).

The AI Act is a new legislative framework that will shift the balance in the ongoing digitalisation of our societies, and it will also have its impact on the field of probation. It is therefore advisable for partners in the domain to take the AI Act into consideration when developing and deploying algorithmic systems.

[3] https://digital-strategy.ec.europa.eu/en/library/commission-decision-establishing-european-ai-office
[4] https://digital-strategy.ec.europa.eu/en/policies/ai-office

Annex – Relevant source list

The full text of the last AI Act can be found here. The European Parliament has published a *Corrigendum* to the AI Act that was adopted in March 2024. The relevant articles (not exhaustive) in light of the probation services are:

Article 3 - Definitions

*(1) '**AI system**' means a machine-based system 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;*

*(2) '**profiling**' means profiling as defined in Article 4, point (4), of Regulation (EU) 2016/679;*

Article 5 – Prohibited Artificial Intelligence Practices

(d) the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support 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;

Annex III, paragraph 6, High-Risk AI Systems Referred to in Article 6(2)

(d) AI systems intended to be used by law enforcement authorities or on their behalf or by Union institutions, bodies, offices or agencies in support of law enforcement authorities for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural 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;

(e) AI systems intended to be used by or on behalf of law enforcement authorities or by Union 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.

A few relevant sources for the relationship between the AI Act and probation services

- Research article of Maastricht University: Predicting Recidivism Meets AI Act
- Research 3RO: AI binnen de 3RO (Kwaliteit Forensische Zorg)

  "We found that AI is scarcely applied internationally within the probation setting, apart from risk assessment instruments. Using AI to organise information is an application probation workers would find valuable. When these AI systems function properly, they can potentially ease the workload of the probation workers. However, downsides of AI systems are potential prejudice in data, not knowing how the algorithm makes a decision and accepting the decisions of AI-systems by the users (both probation workers and convicts). It is clear that these AI systems need to be constantly monitored to prevent this. Within the probation AI needs to be able to decrease the workload of the probation workers in order to be implemented successfully. Keeping the human factor front and central in the probation is essential in implementing AI systems."

- Presentation IPS (Innovative Prison Systems) for the European Commission
- Standardization study DSP (Ritax 2.0, Dutch Probation Services)