

International Standards for Use of AI in Criminal Justice

Pia Puolakka, Team Leader, Prison and Probation Service, Finland

International Network for Criminal Justice

Webinar 4th of June 2024



Council of Europe Treaty on AI

- 17th of May 2024: Council of Europe Adopts First International Treaty on Artificial Intelligence!
<https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence>



Council of Europe Treaty on AI

- The first-ever international legally binding treaty for the use of AI systems
 - A risk-based approach to the design, development, use, and decommissioning of AI systems, which requires carefully considering any potential negative consequences of using AI systems
 - The outcome of the Committee on Artificial Intelligence (CAI), which brought together to draft the treaty the 46 Council of Europe member states, the European Union and 11 non-member states, as well as representatives of the private sector, civil society and academia, who participated as observers
 - Will be opened for signature on 5th of September
- ➔ **The recommendations for Prison and Probation Services are aligned with this general convention and are to be launched after the general convention is opened for signature**

Council of Europe's Council for Penological Co-operation (PC-CP)

- 'Ethical and Organisational Aspects of the Use Of Artificial Intelligence and Related Digital Technologies by Prison and Probation Services'
- Expert group (2021-2023)
 - Pia Puolakka, Team Leader, Prison and Probation Service, Finland
 - Håkan Klarin, CIO, IT Director, Kriminalvården, Sweden
 - Fernando Miró Llinares, Professor of Criminology and Criminal Law, University Miguel Hernández of Elche, Spain



Member States

- Recommends that governments of member states
 - Be guided in their legislation, criminal policy and practice by the principles and rules contained in the Recommendation
 - Ensure that this **Recommendation** and its **explanatory report** are translated and disseminated as widely as possible including judicial authorities, prosecution, police, prison, probation and juvenile justice services, as well as among private companies which design and provide AI in the framework of the criminal justice system
- The public authorities in charge of prison and probation services remain fully responsible for ensuring respect for the principles in the Recommendation. They should also ensure that the private companies which design, develop, provide, use and decommission such technologies, follow the same principles as stated in the Recommendation.

Definitions

- **Artificial Intelligence (AI)** means a machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment.
- **Related digital technologies** is a generic term that refers to all electronic devices, automatic systems and technological resources that generate, process or store information and data which are being used by AI.

Artificial Intelligence and Related Technologies in Corrections

- **Natural Language Processing**
 - Decision support systems, Analysis & Security
 - **Voice and Speech Recognition**
 - Voice recognition, Speech recognition & Natural Language Understanding
 - **Image Recognition**
 - Facial recognition & Security
 - **Data Analysis**
 - To analyse large data sets for patterns, correlations, and categories imperceptible to human minds
- **Prediction**
 - Behaviour & Optimization
 - **Recommender and Expert Systems**
 - Offender Management Systems (OMS)
 - **Virtual reality (VR) and Augmented Reality (AR)**
 - Rehabilitation and Learning
 - And more...

Principles of Recommendations

- AI and related digital technologies should be used legitimately and proportionately when and if they:
 - I. Contribute to the rehabilitation and reintegration of offenders
 - II. Do not replace prison and probation staff but assist them in their everyday work
 - III. Help the criminal justice system, the execution of penal sanctions, measures & the reduction of recidivism
- **30 Principles**
 1. Basic principles
 2. Data protection and privacy
 3. Use for safety, security and good order
 4. Use for offender management, risk assessment, rehabilitation and reintegration
 5. Use for staff selection, management, training and development
 6. Research, development, evaluation and regular revision

1. Basic Principles

- Respect for **human rights** and dignity of all persons impacted by use of AI
- **Aim for least negative impact on human rights:** should be only done if strictly necessary
- In conformity with the international standards and defined by national law: **liability** for any unlawful harm caused by AI should be ensured
- **Biases** should be prevented: to ensure equality and to prevent or resolve the creation or intensification of any discrimination or inequality between individuals or groups of individuals
- **Transparent** to public scrutiny, **monitored** on a regular basis, the logic behind and the outcomes of its use should be sufficiently **explainable**

1. Basic Principles

- When a decision is based on the use of AI and affects human rights, **human supervision** and **effective complaint procedures** should be put in place
- **Data** should be **accurate** and **sufficiently representative** of the general population and minority groups, including the target groups that might be impacted
- Preserve and promote **positive and beneficial human relations** between staff and offenders
- All AI users **should understand the basics** regarding what this use implies, including how and for what purpose and what are the ethical rules to be respected

2. Data Protection and Privacy

- Offenders continue to enjoy their fundamental rights and freedoms, including the right to **respect for private life** and the **right to data protection**
- All key actors should comply with data protection law, be **transparent** and able to **demonstrate** to the individuals concerned that they comply with data protection principles and obligations
- Data should be stored in a form that allows a personal identification **for no longer than is strictly necessary** to fulfil the purposes for which it was initially collected
- Only the amount and type of personal data which are **strictly necessary to fulfil a specific task** should be collected, stored, transmitted or otherwise processed

3. Safety, Security and Good Order

- Use of AI for maintaining safety, security and good order should also allow for better **risk and crisis management**
- Prison and probation services should be consulted in order to identify and evaluate the **needs regarding the assistance of staff through AI** and related digital technologies in the execution of tasks related to safety, security and good order
- The use of AI in **electronic monitoring**, including **biometric recognition** technologies, should be proportionate to the purpose and used only when strictly necessary. It should be carried out under regular human control and should be human centered.

4. Offender Management, Risk Assessment, Rehabilitation and Reintegration

- For **offender management**, AI should be used to manage offenders' files and particular cases and to generate automatic alerts in cases of non-compliance if they improve monitoring and decision taking. The final responsibility for this remains with the professionals!
- When used for increasing the **accuracy and objectivity of risk assessment**, the challenges of algorithmic biases and quality and representativeness of data should be addressed
- The results of **risk assessment should be used only for risk management**
- The **rehabilitation** and **reintegration** of offenders, as well as their **social contacts** and **medical treatment** can be facilitated by use of AI, but it shouldn't replace face-to-face contact

5. Staff Selection, Management, Training and Development

- To **optimise** human and managerial **capacities** and **processes**, and to focus on supporting the staff's **professional development**
- Should assist managers in predicting future **organisational capacity**, including detecting especially problematic areas in staff resourcing
- A person should have the right to be **informed of the reasons for decisions** related to their selection, recruitment and professional development, taken on the basis of AI, and **should have the right to request their human review**

6. Research, Development, Evaluation and Regular Revision

- The design and development of, as well as research in, AI and related should be sufficiently well funded and supported, and should **help develop further the proper and efficient use of AI**
- AI and its use should **be evaluated at regular intervals** by independent and competent evaluators concerning their performance, their intended and unintended outcomes and the need for adaptations
- Procedures and resources should be in place to regularly monitor, identify, assess, prevent and mitigate **possible risks and adverse effects** resulting from the design, development and use of AI

To Be Reviewed Regularly!

- 30. “This Recommendation needs to be reviewed regularly and revised accordingly in order to continue working to protect the human rights and fundamental freedoms of its users and the safety and security of our societies.”



Some Finnish Examples

- **RISE AI***: Recommender and expert system for offender management -> risks, needs, services and placement to units
- **Aurora AI**: National AI programme by Ministry of Finance for public service recommendations based on the client input (questionnaire)
- **Training AI** as prison labour in three prisons
- **Online courses on basics of AI** by Helsinki University to increase cognitive and digital skills, and AI literacy of offenders
- **Chatbot** for customer service on the Prison and Probation Service's public website

Should We Worry About AI in Prisons and Probation?

- AI has many **benefits**: Faster processing of large data sets, effective use of resources, faster detecting of security risks and offenders' needs etc.
- AI has also **risk factors**: Biased data -> Biased conclusions, automation starts to orient processes too much instead of human decision-making
- Using AI to **assist human decision making with the intention of contributing to the rehabilitation and reintegration of offenders** is the key to appropriate and ethical use of AI
 - In the future **AI literacy** is a necessity for all corrections' staff and offenders!

References

- Council of Europe adopts first international treaty on artificial intelligence, 17.5.2024.
<https://www.coe.int/en/web/portal/-/council-of-europe-adopts-first-international-treaty-on-artificial-intelligence>
- Guidelines for the safe and ethical implementation of AI in prisons and probation. Joint Interview. Council for Penological Co-operation of the Council of Europe. *Justice Trends*, 10. <https://justice-trends.press/guidelines-for-the-safe-and-ethical-implementation-of-ai-in-prisons-and-probation/>
- Knight, V. & Van De Steene, S. (2020) The digital prison: Towards an ethics of technology. In: Birch, P., Sicard, L. (Eds.) *Prisons and Community Corrections*. London: Routledge.
- *Puolakka, P. (2020). *RISE AI: Reducing the Risk of Recidivism with AI*. Aalto Executive Education: Diploma in Artificial Intelligence. Unpublished.
- Puolakka P., & Van De Steene, S. (2021). Artificial Intelligence in Prisons in 2030. An exploration on the future of Artificial Intelligence in Prisons. *Advancing Corrections Journal*, 11, ICPA.



R/SE

PRISON AND PROBATION SERVICE
FINLAND

Thank you!

Pia Puolakka

pia.puolakka@om.fi