COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 12 May 2023 CDCJ-ADMIN-AI(2023)01 prov3

ADMINISTRATION AND ARTIFICIAL INTELLIGENCE
LIMITED WORKING GROUP CDCJ-ADMIN-AI

DRAFT CONCEPT NOTE CONCERNING THE USE OF ARTIFICIAL INTELLIGENCE (AI) FOR POLICING, THE ADMINISTRATION OF JUSTICE AND BORDERS/MIGRATION AND THE USE OF AI OR OTHER AUTOMATED DECISION-MAKING (ADM) FOR COMMUNICATION PURPOSES

Prepared by the limited working group ADMIN-AI for examination by the CDCJ

(item 5.4 of the draft agenda of the 100th plenary meeting of the CDCJ)

Directorate General Human Rights and Rule of Law – DGI

Background

At its 99th plenary meeting (23–25 November 2022), the CDCJ analysed the proposals made by the CDCJ-ADMIN-AI concerning possible future work of the CDCJ in the field of AI and administrative law and provided further guidance to the working group, instructing it to develop proposals on (a) using artificial intelligence (AI) for policing, the administration of justice and borders/migration and (b) using AI or other automated decision-making (ADM) for communication purposes, for their examination by the CDCJ at its 100th plenary meeting in May 2023.

The CDCJ-ADMIN-AI members proposed that the issues related to using AI or ADM for communication purposes be addressed while updating the handbook "The administration and you". This concept note further develops the proposal put forward earlier before the CDCJ by the CDCJ-ADMIN-AI on using AI for policing, the administration of justice, and borders/migration and explains how issues related to using AI or other ADM for communication purposes can be addressed while updating the handbook.

The concept note covers the following aspects:

1. Using AI for policing, the administration of justice, and borders/migrants:

    a) introduction, a brief rationale for the selection of the topic
    b) what are the possible gaps/risks provoked by the deployment of AI and ADM
    c) practical examples of the usage of AI for policing, the administration of justice, and borders/migration
    d) proposals for the CDCJ's future activities in this field.

2. How the issues related to using AI or other ADM for communication purposes can be addressed while updating the handbook "The administration and you".

1. Using AI for policing[1], the administration of justice, and borders/migration

a) Introduction, a brief rationale for the selection of the topic

1. The deployment of AI and ADM in the public sector is becoming more prevalent day by day and a number of tasks previously assigned to human beings are now (and will be even more in the future) carried out by inanimate machines. As a result, the idea that in the near future individuals will be more in contact with machines than with human beings when seeking state services is far from being unreal.

2. Currently, AI and ADM are used in various spheres in the public sector and their use has also reached the daily functioning of law enforcement agencies and the administration of justice.

3. Unlike human beings, who can be held accountable for their actions or inactions, machines cannot be held accountable nor punished for the errors observed in their functioning. Only the bodies (people) that implement or control them can be held responsible. It therefore requires additional safeguards to be in place when deploying AI and ADM in the public sector. This statement is even more true when speaking about the use of AI for policing, the administration of justice, and borders/migration where there is a stronger imbalance between the public authority and the persons affected. Additional safeguards can take various forms, such as impact assessments and ex-post reviews, aimed at ensuring transparency and accountability, reducing risks, and preventing and remedying in a timely manner any possible infringements.

[1] The notion of "policing" for the purpose of this concept note should not be understood as covering criminal law aspects of law enforcement but the management of data using AI and ADM.

4. The area proposed for the future work of the CDCJ comprises three subcategories: policing, administration of justice, management or control of borders/migration. These subcategories can be dealt with individually; however, their common denominator – in contrast to other proposed areas – is that these are areas in which official action is regularly based on a strong imbalance between the public authority and the persons affected, and accordingly, there is a higher risk of infringement of the rights of the concerned individuals. The relatively higher risk of violations of human rights and the absence of common policies, standards, or guidelines for national authorities are only some of the reasons why more measures aimed at securing human rights are required in this area.

b) What are the possible gaps and risks provoked by the deployment of AI and ADM

5. Many of the risks and limitations of using AI and ADM-based systems in policing, the administration of justice, and border/migration mirror the risks and limitations that apply to AI and ADM tools in a range of different contexts, both within the administration and in other areas such as healthcare, education, and consumer technologies. These can broadly be described as follows:

- **Bias and discrimination**: Discrimination constitutes a qualified unequal treatment of different persons in comparable situations, with the effect of placing people at a disadvantage. A distinction can be made between direct and indirect discrimination. Direct discrimination occurs when a legal action is directly linked to a sensitive characteristic without proper and sufficient justification. Indirect discrimination can be the consequence of a formally neutral legal regulation that does not, itself, contain any obvious disadvantages for members of specially protected groups. The unequal treatment only results from the practical effects of the regulation. There exist different sources of discrimination in AI systems.[2] Often, biases established in society are explicitly or implicitly transferred to technology. These may originate either in society or reflect the personal attitudes of individuals (e.g. customers or system designers) who have a significant influence on the design of the system. The discrimination is the result of a pre-existing bias in the training data that represents the current society. For example, an algorithm allocating credit scores penalises women over men because women have historically had less access to credit.[3] In the case of predictive policing systems, historical over-policing of certain communities or areas is likely to be reinforced and perpetuated by an algorithmic system based on historical policing data.[4] Biases can also arise from technical specifications; for example, when facial recognition systems perform poorly on black faces instead of white because of underrepresentation in the dataset.[5] The discrimination as a consequence is thus strongly related to the missing data when a population group is underrepresented in the training groups. However, certain forms of discrimination also arise only in the application, especially when the AI application can continue to learn dynamically. The bias typically develops after a system has been implemented because of changes in societal knowledge, changes in the population, or with respect to cultural values. Furthermore, discrimination can result from correlations linked to so-called proxies. Proxies are seemingly innocuous characteristics that may, however, correlate strongly with proscribed characteristics. For example, the place of residence may be related to

[2] Kordzadeh Nima/Ghasemaghaei Maryam (2022), Algorithmic bias: review, synthesis, and future research directions, European Journal of Information Systems, Vol. 31 No. 3, pp. 388–409.
[3] Richardson Sharon (2022), Exposing the many biases in machine learning, Business Information Review, Vol. 39 No. 3, pp. 82–89.
[4] Alikhademi Kiana et al. (2022), A review of predictive policing from the perspective of fairness, Artificial Intelligence and Law, Vol. 30, pp. 1–17.
[5] Bacchini Fabio/Lorussi Ludovica (2019), Race, again: how face recognition technology reinforces racial discrimination, Journal of Information, Communication and Ethics in Society, Vol. 17 No. 3, pp. 321–335.

ethnic background or social status.[6] The impression of causality is created, although such causality does not exist.

- **Invisible inaccuracies and spurious correlations**: Many AI and ADM-based systems use a form of probabilistic reasoning that generates results that identify correlative relationships to a certain degree of certainty, resulting in outcomes that are neither completely robust nor based on a causal relationship, and are thus often inaccurate or spurious. This limitation of present-day AI – which will be mitigated as systems improve over time – is compounded by the "black box" nature of AI and ADM systems, which prevents a user from interrogating the accuracy of output or understanding the nature of the reasoning that underpinned a decision. Several examples of the harms and risks of inaccurate AI and ADM-based systems have been seen in the field of welfare administration, where false allegations of fraud have been levelled at benefit claimants resulting in severe financial and personal damage.[7]

- **Privacy and the normalisation of surveillance**: AI and ADM-based systems are built on huge amounts of data that are often of low quality, unrepresentative, or of unknown provenance, and their deployment incentivises the collection and retention of greater amounts of data to further train models. Some AI systems, particularly embodied AI (in the form of robots or drones) or facial recognition systems, more overtly expand surveillance into public and private realms, thus normalising surveillance.

- **Human autonomy and dignity**: Human behaviours, interactions, and relationships are challenged by the introduction of embodied AI systems, such as robots and drones, and an understanding of how these new technologies interact, where notions of human dignity are still emerging. As AI and ADM-based systems take over tasks that have historically been performed by humans, there is a risk that humans adapt behaviours to meet the demands of automated systems, or that they have less options for autonomous choices available because the options have been predefined/recommended by an ADM system. The introduction of AI systems into human systems, in effect, changes those systems and may render them less, rather than more, effective.

6. Although these drawbacks are common amongst AI and ADM-based systems, the potential harms that might flow from AI and ADM-based systems in the particular sub-fields of policing, the administration of justice, and border/migration are considerably severe, as they pertain to an individual's rights to a fair trial, freedom of movement, and freedom from arbitrary treatment and deprivation of liberty, among other rights. As such, despite the seemingly advantageous aspects of applying AI and ADM-based systems in these fields, in terms of cost savings and expeditiousness, the threshold for implementing them consistently with human rights is much higher.

c) Practical examples of using AI for policing, the administration of justice, and borders/migration and (b) using artificial intelligence

Policing (and law enforcement)

7. AI – or ADM-based systems – can be used in various forms for policing and law enforcement purposes. These include for example:

[6] Prince Anya E.R./Schwarcz Daniel (2019), Proxy Discrimination in the Age of Artificial Intelligence and Big Data, Iowa Law Review, Vol. 105 No. 1–5, pp. 1257–1318.
[7] Heikkilä Melissa (2022), Dutch Scandal serves as a warning for Europe over risks of using algorithms, Politico,https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

- Infringement detection: Processing of large amounts of data (web-scrapping) might be used to detect hate speech, terrorist threats, or child pornography content (a task mainly assigned to private platforms by the Digital Services Act, Oct. 2022), as well as social benefit fraud and tax fraud.
- In 2019, for example, it became known that the Dutch tax authorities had used a self-learning algorithm to create risk profiles to detect fraud in childcare services. Based on the system's risk indicators, families were fined on mere suspicion of committing fraud.[8]
- The French tax authority has been using AI to fight tax fraud since 2014. A law adopted in 2020 even authorised the tax authority to use data mining on social networks to track down fraud. Since 2022, France has used artificial intelligence to detect more than 20,000 undeclared swimming pools.[9]

8. Artificial intelligence (SATIKAS) from the Estonian Agriculture and Registration Agency uses satellite photos from the European Space Agency to detect whether beneficiaries have mowed their lawns before the deadline. All recipients will be checked. However, there would be no automatic sanction, although those who have not mowed their lawns will be dealt with by an official.[10] As regards predictive policing, AI-systems can be used by the police and other competent law enforcement agencies to predict and/or prevent crime. AI software can analyse large amounts of data. The software can be used to identify trends, patterns of behaviour, and other correlations much faster and more accurately than human beings are able to. In the international context, AI-based systems can be found in both person-based predictive policing and space-based predictive policing. On the one hand, this involves a prognosis about the dangerousness or endangerment of a person, and on the other hand, it involves a spatially based prognosis regarding the probability of criminal acts. In the first case, it is asked who could become dangerous or be endangered, while in the second case, where a certain danger could occur.

9. Such systems are used in the United States of America. For example:

- *Strategic Subject List* in Chicago (in force before the year 2020): This is person-based predictive policing. Data on people's social contacts is used to calculate the risk that a person could be involved in gang crime, for example.[11]
- *Predpol,* established in several cities such as Los Angeles, Chicago, Seattle, and Boston: This system is designed to predict the time and location of potential threats and provide appropriate police patrols.[12]
- *Risk Terrain Modelling (RTM)* is a geospatial crime analysis tool that is designed to examine environmental risk factors associated with crime and to identify the areas where their spatial influence is linked with vulnerability to criminal behaviour.[13]
- *HunchLab, used in several police administrations, for example in Chicago and Philadelphia:* HunchLab integrates various factors such as crime rates, near repeat patterns, socioeconomic factors, temporal factors, and social events into its analysis. This information is processed by a machine learning algorithm and updated regularly. Through training and testing the crime data, HunchLab creates predictions, which are

[8] Heikkilä Melissa (2022), Dutch Scandal serves as a warning for Europe over risks of using algorithms, Politico,https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/
[9] Euronews with AFP (2022), France uses artificial intelligence to detect more than 20,000 undeclared swimming pools, Euronews, https://www.euronews.com/my-europe/2022/08/30/france-uses-artificial-intelligence-to-detect-more-than-20000-undeclared-swimming-pools
[10] European Association of Remote Sensing Companies (2021), A Case Study Grassland Monitoring in Estonia,https://earsc.org/sebs/wp-content/uploads/2021/05/Grassland-Monitoring-in-Estonia_vfinal.pdf
[11] Chicago Data Portal (2020), Strategic Subject List – Historical, https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List-Historical/4aki-r3np
[12] Predpol website : https://www.predpol.com
[13] Risk Terrain Modeling Actionable Spatial Analysis website : https://www.riskterrainmodeling.com

then utilised to suggest patrol allocation. To provide on-the-go support, the HunchLab platform is accessible through mobile devices, enabling police officers to view potential criminal hotspots in real-time. [14]

10. Similar systems exist in Europe as well. For example:

- *Burglary Predictor* in Utrecht (Netherlands): this tool uses machine learning algorithms to analyse data and identify areas where burglaries are most likely to occur for a given week. This creates models that can predict the likelihood of future burglaries in specific locations. Factors such as weather conditions, dates of public holidays or special events, as well as sociodemographic statistics of the neighborhoods, are taken into account.[15]
- *AIDA:* An algorithm used by the West Midlands Police, which predicts the likelihood of crimes occurring in specific areas.[16]

Administration of Justice

11. For the purposes of the administration of justice, AI and ADM-based systems are used for the following (the list is non-exhaustive):[17]

- Processing of large amounts of data (for fact-finding, e.g. by identifying suspects or detecting correlations and for identifying the legal rules applicable to a given dispute).
- Data processing and content generation (anonymisation, translation, decisions drafting).
- Process automation (relations with litigants and internal organisation of the courts; for instance, distribution of the cases among the judges).

12. For various reasons, in particular, due to the social unacceptability, the practice of replacing a judge with a machine to allow the full automation of judgments is in its infancy and very rarely experienced, whereas the provision of legal assistance through AI systems is more developed. For example:

- In the American judicial system, COMPAS is used to help judges decide on certain issues. COMPAS is a personal predictive analysis AI system that evaluates the risk of recidivism of a defendant, the risk of violent behaviour, and the risk of non-appearance in the absence of pretrial detention. The software, which illustrates the practice of evidence-based sentencing, therefore consists of deciding on sentences and its modalities according to the recidivism risk score, calculated from 137 input data. The algorithm was shown to discriminate against African-American populations, even though no ethnicity criteria were used. The cross-referencing of data, including the area of residence (which can reveal ethnicity), generated indirect discrimination. This is compounded by the fact that the system is driven by biased data, from court decisions themselves that reflect social stigmas in a country where African-Americans are already facing sentences that are on average 20% longer than Caucasians.[18]
- In South America, another tool known as Prométéa has been developed at the office of the public prosecutor in Buenos Aires (recommending solutions to disputes concerning the allocation of housing or social assistance). In Colombia, the PretoIA

[14] Chammah, Maurice (2016), Does Predictive Policing Lead to More Police in Black Communities?, The Marshall Project, https://www.themarshallproject.org/tag/hunchlab
[15] Burglary prediction for the municipality of Utrecht, Xomnia, https://www.xomnia.com/burglary-prediction-for-the-municipality-of-utrecht/.
[16] AIDA website : https://www.project-aida.eu/index.php/about-aida.
[17] Xu Zichun, Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities, Applied Artificial Intelligence, Vol. 36 No. 1.
[18] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

tool is used by the constitutional court to analyse *tutela* decisions (which allows any litigant to claim the protection of his or her fundamental constitutional rights threatened by the action or inaction of a public authority - decisions that are automatically forwarded to the constitutional court).

- Criminal justice: *HART* (Harm Assessment Risk Tool), which is implemented in Durham (UK), is a machine learning programme used by human custody officers. It assesses whether offenders brought into custody are at high, medium, or low risk of reoffending.[19]
- Data Processing in justice: *Datajust*, which is used in France, is a system which develops an algorithm that automatically extracts and evaluates data from court decisions on compensation for personal injuries. Specifically, it aims to capture the amounts claimed and offered by the parties in the instances, the valuations proposed in amicable dispute resolution procedures, and the amounts awarded to victims by the courts.[20]
- Estonian Courts are using the AI application "Salme". Salme is a court recording programme that includes a speech recognition software. The latter has been specifically designed to take into account the vocabulary used in hearings, a transcription of what was said, in addition to an audio recording. The speech-recognition transcript can be edited in real-time and the voices of different people present in the hearing can be better distinguished.[21]

Migration and Border Control

13. The tendency of increased use of AI and ADM-based systems in the public sector is observed in the field of migration and border control as well.

14. Images and footage from cameras installed in airplanes, helicopters, or drones, allow the detection of instances of illegal border crossings. These AI systems have the advantage of being able, compared to human controls, to continuously analyse a larger amount of data, at night, at times in difficult weather conditions, and at a lower cost.

    - The Estonian Police and Border Guard Board has purchased nine ELIX-XL drones, with the support of European Union funding, to monitor the daily situation on the eastern border and to respond to rescue and border incidents. The drones are part of an overall border construction project to ensure that the external borders of several states are securely protected.[22]
    - For two years (2021 and 2022), a European Maritime Safety Agency (EMSA) unmanned aerial vehicle (UAV) has been deployed in Estonia for a few months to help patrol the maritime border, carry out search and rescue operations, and detect marine pollution.[23]

---

[19] Oswald Marion et al. (2018), Algorithmic risk assessment policing models : lessons from the Durham HART model and "Experimental" proportionality, Information & Communications Technology Law, Vol. 27 No. 2, pp. 223–250, Barnes (2022), AI can predict reoffending; university finds, Durham Constabulary,https://www.durham.police.uk/News/News-Articles/2022/January/AI-can-predict-reoffending-university-study-finds.aspx; Fair Trials (2022), FOI reveals over 12,000 people profiled by flawed Durham police predictive AI tool, https://www.fairtrials.org/articles/news/foi-reveals-over-12000-people-profiled-by-flawed-durham-police-predictive-ai-tool/.

[20](2020), Traitement automatisé de données à caractère personnel, Justice.fr, https://www.justice.fr/donnees-personnelles/datajust

[21] Oyetunde, Blessing (2022), Introducing Salme; Estonian courts' speech recognition assistant, e-Estonia, https://e-estonia.com/introducing-salme-estonian-courts-speech-recognition-assistant/

[22] LETA/TBT (2018), Estonia's police authority to showcase drones purchased for guarding eastern border,The Baltic Times,https://www.baltictimes.com/estonia_s_police_authority_to_showcase_drones_purchased_for_guarding_eastern_border/

[23] (2022), Drones launched from Saarema monitor pollution and ships in Baltic Sea, News, https://news.err.ee/1608743026/drones-launched-from-saaremaa-monitor-pollution-and-ships-in-baltic-sea

15. AI and ADM are also used for asylum-related issues. For example:
    - Germany's immigration authority, the Federal Office for Migration and Refugees, has piloted the use of digital tools, including facial and dialect recognition, to help verify personal identities within the asylum determination process when asylum seekers arrive in the country.[24]
    - "iBorderCtrl is an experimental artificial intelligence (AI) system, funded since 2016 by the European Union and used in at least three border crossings in the Schengen area (Greece, Hungary and Latvia). It assesses the "reliability" of the person wishing to enter the European area: it is a "facial recognition lie detector" which then redirects the traveller either into fast queues or, on the contrary, towards thorough controls. The system is able to discern, among 38 "micro-movements", the supposedly untrue statements of a person. These "micromovements" include, for example, the angle of the head or the movement of the eyes. This virtual border guard asks the traveller questions (name, country of origin, length of stay, reason for travel, etc.). Those whose answers are deemed honest by the system are given a code authorising them to cross the border. The others are directed to physical border guards.

Technological solutions

Biometric Recognition

16. Biometric recognition software is being used by the police and other competent law enforcement agencies to identify individuals based, for example, on their faces, voices or gait. The software relieves them from checking identity documents across different databases manually. Apart from recording an actual image, most of these software applications also collect and process biometric data and can thereby identify people. One of the most widespread uses is to identify offenders (e.g. *Traitement des antécédents judiciaires* (TAJ) in France[25], Interpol).

17. In addition to being used for identification purposes, biometric recognition technology is also used to authenticate a person's identity by comparing the features of a person's face to those stored in a database. An image of the person's face is captured with a digital camera (creation of a biometric template) that is then processed to extract key features, such as the distance between the eyes, the position of the nose, or the shape of the mouth (feature extraction). The extracted features are compared with the use of an algorithm to those stored in the database. If the similarity score is given, the person is considered to have been successfully authenticated. For example, facial recognition authentication is used for border control at airport gates (Automated Border Control - ABC[26]).

18. Biometric recognition for identification purposes might be deployed as an instrument of (mass) surveillance in publicly accessible spaces for identification purposes. Mass surveillance is a way of monitoring the population with the processing of biometric data. The use of such systems in publicly accessible spaces can not only violate people's right to privacy but can also have chilling effects on people's enjoyment of other human rights, such as freedom of expression or association as it would deter them from participating in protests or assembling with others. This can also have discriminatory effects by overly affecting groups that are already facing discrimination.

[24] Forster, Madeleine (2022), Refugee protection in the artificial intelligence era, Chatham House, https://www.chathamhouse.org/2022/09/refugee-protection-artificial-intelligence-era/2-near-future-ai-and-asylum

[25] Traitement d'antécédents judiciaires (Taj), Service-Public.fr website : https://www.service-public.fr/particuliers/vosdroits/F32727

[26] European Commission, Migration and Home Affairs, Glossary Automated Border Control (ABC): https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/automated-border-control-abc_en

19. Real-time biometric recognition is used sporadically in Europe to monitor specific, geographically delimited areas for identification purposes. The scanned faces are checked against police or court watch lists.

    - For example, the London Metropolitan Police uses live facial recognition cameras to improve officers' ability to identify suspects.[27]
    - In France, facial recognition has been experimented during the carnival in 2019 in Nice. Prior consent from the volunteers to conduct the experiment was obtained.[28]

20. It must be noted that not only does real-time use bring with it the possibility of (mass) surveillance, but it implies greater risks of breaching human rights. It is often not relevant to know whether an individual could be identified through such a system in real-time or some days later.

Remote monitoring

21. Drones are increasingly being used for monitoring and patrolling purposes. The drone's aerial capability allows it to inspect structures that are difficult to reach from the ground. Researchers at the University of Maryland and the University of Zurich equipped a drone with event cameras and a sonar system to make it capable of detecting and dodging objects thrown at it. These cameras do not necessarily use biometric recognition technologies. They can, however, detect crowd movement, a person falling, a gun in a person's hand, or even count people wearing masks in public spaces, for example. These drones can be used in high-risk environments.

    - In Greece, for example, drones were used experimentally to verify information in tax returns in tourist regions. Using the data collected by the drones, the authorities were able to determine how many passengers were on board excursion boats and compare this data with the receipts declared by the taxpayers in their tax returns.[29]
    - A drone is being developed at the Academy of Internal Defence in Estonia, which will in the future be used by the police and rescue services to search for missing people in forests and landscapes. AI uses visual data to detect objects. At the end of 2022, an Estonian company announced that they started producing drones specifically for rescue services, with AI technology capable of detecting a forest fire or finding a missing person in the landscape.[30]
    - In France, drones have been used to monitor public space during lockdown in 2020, and, after that, to monitor mass protests.[31]

[27] UK Metropolitan Police Facial Recognition Technology Advice and Information : https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition; Dodd Vikram (2020), Met police to begin using live facial recognition cameras in London, The Guardian, https://www.theguardian.com/technology/2020/jan/24/met-police-begin-using-live-facial-recognition-cameras.

[28] Jasserand Catherine (2022), Experiments with Facial Recognition Technologies in Public Spaces: In Search of an EU Governance Framework, SSRN, https://ssrn.com/abstract=4204452

[29] De Hoon, Iven. Greece uses drones to find tax fraud, No More Tax, https://nomoretax.eu/greece-drones-tax-fraud/.

[30] Krattworks website : https://www.krattworks.com/

[31] La Quadrature Du Net (2020), France: First victory against police drones, EDRi, https://edri.org/our-work/france-first-victory-against-police-drones/; Rosemain, Mathieu (2021), French watchdog condemns police for unlawful use of drones to patrol lockdown, Reuters, https://www.reuters.com/article/us-health-coronavirus-france-drones-idUSKBN29J15Z

- Estonia did not have a full lockdown, but there was a 2+2 rule (distance) and a ban on mass gatherings. Drones were also used as a loudspeaker to warn the public of the ban.[32]

Robots

22. Robots are being increasingly used to monitor security in both low-risk and high-risk areas. They are used to patrol shopping malls and monitor power grids or other sensitive locations. They can reach areas not accessible or not conducive to human patrolling or monitoring. In addition, their use can increase efficiency and 24/7 coverage, and reduce risk to human security personnel. For example:

- United States: Robots are being used for security purposes in shopping malls and airports. In San Francisco, the police already use robots, which are remotely controlled and sent for reconnaissance when there are bomb threats, for example, or to check that the premises are safe before intervening. In the extension of this use, the idea is to go further and equip these robots with an explosive charge to be able to bring down barricades and to direct police toward an armed or dangerous suspect (December 2022).[33]
- Japan: Robots are being used for patrol duties in train stations. In Tokyo, Takanawa Gateway Station will install six types of robots capable of handling such tasks as guiding passengers, cleaning the station, and performing security duties.[34] This type of robot is already widely used in Japanese stations.

d) Proposals for the CDCJ's future activity

23. In view of the foregoing, the CDCJ-ADMIN-AI is convinced that the increasing use of AI and ADM using artificial intelligence (AI) for policing, the administration of justice, and borders/migration calls for the development of common policies, standards, or guidelines for national authorities.

24. The CDCJ-ADMIN-AI considers that the CDCJ as a standard-setting intergovernmental body of the Council of Europe, competent in the field of public and private law, could take actions within its competence to address the challenges posed by the AI and ADM deployment for policing, the administration of justice and borders/migration. The CDCJ-ADMIN-AI confirms that it is unaware of the intention of other organisations to work in this particular field. This, in turn, makes the possible actions of the CDCJ even more relevant.

25. CDCJ actions could help to ensure that AI and ADM are used responsibly for policing, the administration of justice, and borders/migration and with full respect to fundamental rights such as personal freedom, informational self-determination, and prohibition of discrimination.

26. The CDCJ-ADMIN-AI is of the view that developing a recommendation on the usage of AI and ADM for policing, the administration of justice, and borders/migration would be an appropriate way of meeting the possible challenges in this field, briefly touched upon in the concept note.

---

[32](2020), Police take to drones to enforce 2+2 coronavirus rule, News, https://news.err.ee/1075942/police-take-to-drones-to-enforce-2-2-coronavirus-rule

[33] Labeyrie, Isabelle (2022), Etats-Unis : à San Francisco, les futurs robots-tueurs de la police inquiètent la population, Franceinfo, https://www.francetvinfo.fr/replay-radio/le-monde-est-a-nous/etats-unis-a-san-francisco-les-futurs-robots-tueurs-de-la-police-inquietent-la-population_5491851.html

[34]Nahao, Riho (2020), Robots roam Tokyo's newest train station to patrol and sweep, Nikkei Asia, https://asia.nikkei.com/Business/Transportation/Robots-roam-Tokyo-s-newest-train-station-to-patrol-and-sweep

2. How the issues related to using artificial intelligence (AI) or other automated decision-making (ADM) for communication purposes can be addressed while updating the handbook "The administration and you"

27. At its 99th plenary meeting (23-25 November 2022), the CDCJ instructed the CDCJ-ADMIN-AI to further develop two of the proposals it made for possible future work of the Committee in the field of AI and administrative law. As regards the proposal on possible future work on using AI or ADM for communication purposes, the Committee discussed whether this issue could be covered while updating the handbook "Administration and You" or if it needed separate attention.

28. Following the discussions at the plenary meeting, the CDCJ-ADMIN-AI examined both options. CDCJ-ADMIN-AI members expressed the view that, to a large extent, the issues raised in using AI or ADM for communication purposes are comparable to the ones raised by the deployment of AI and ADM in public administration in general. As a consequence, the CDCJ-ADMIN-AI decided to propose to the CDCJ to cover the issues related to usage of AI or ADM for communication purposes while updating the Handbook. Accordingly, this part of the concept note explains how such issues can be addressed within the framework of the update.

29. There has been considerable progress in general purpose AI systems that generate content, including Large Language Models (hereafter, LLMs) installed in chatbots (hereafter, generative chatbots), leading to a debate about how and when such chatbots might be integrated into services across the economy as well as in public administration.

30. This recent wave of generative chatbots is technically and substantively different from what has historically been used by e-commerce sites and some e-government applications. Whereas most chatbots were previously just automated decision-trees with tightly defined parameters and outputs (and therefore with restricted capabilities), generative chatbots are integrated into a handful of LLMs, spreading AI systems that draw on huge amounts of data gathered from the internet and trained against millions of parameters. They are designed to mimic human speech and generate realistic and conversational answers to prompts.

31. Generative chatbots are experimental tools that are still only partially understood by those who built them, let alone by their users and deployers. They are likely to constitute only the first wave of a range of generative tools that will be built on LLMs in the years to come. They bring with them a range of risks, concerns, and drawbacks. Because they are trained on a dataset comprised, amongst other things, of online internet forums, copyrighted data from published works, news sites, and a wide range of other material, their outputs are unpredictable, occasionally abusive, or harmful, often inaccurate, and sometimes deceptive (a chatbot might write plausible-sounding answers but which are fabricated, false or non-sensical, known as "hallucinating"). Technical mechanisms for fine-tuning generative chatbots are being developed, and technological advancements are moving incredibly fast with the result that many concerns might be avoided in the coming months.

32. The long-term challenges posed by generative chatbots are more fundamental and harder to quantify. Generative chatbots may equal or better mimic human performance on a range of tasks, including summarising research or synthesising policy documents, but they also lack qualities of human scrutiny, discretion, and analysis, as well as accountability, comprehensiveness, and transparency. They may mimic human conversation so well that it may be difficult for their conversation partner to discern an AI. They may also be vulnerable to outside manipulation.

33. In the coming months and years, there will inevitably be a push towards integrating generative chatbots and other generative tools into public administration. For example, a Colombian judge has already been found to be using a chatbot. It is vital that these tools are considered as something fundamentally different from technological tools that have been used so far, and that caution prevails as to their integration into public administration.

34. While the deployment of generative chatbots brings a number of advantages, their use poses new challenges for public administration, their deployers, and the beneficiaries of their services. Accordingly, in its capacity as a standard-setting intergovernmental body of the Council of Europe, the CDCJ could provide further legal support to address these challenges.

35. A number of issues raised by the deployment of chatbots are very similar to the challenges posed by the use of AI and ADM in public administration in general. More specifically, the challenges posed by the deployment of AI-enabled chatbots are particularly relevant to the following principles:

- Principle 2, (**equality of treatment**) - chatbots should give the same information when dealing with the same kind of request
- Principle 3, (**objectivity and impartiality**) - chatbots should not “act” in a biased manner
- Principle 5, (**legal certainty**) - information provided through generative chatbots to advise citizens should be reliable
- Principle 6, (**transparency**) - citizens in contact with public administration through digital tools should be informed that the said tool is a generative chatbot and that no human being is involved
- Principle 7, (**privacy and data protection**) - generative chatbots can also process personal data
- Principle 10, (**right to be heard**), generative chatbots could be used when allowing citizens to express their views before an administrative decision is taken and when defining the limits of such an approach.
- Principle 13 (**form and notification of administrative decisions**)

36. Considering that the handbook is currently being revised and that the principles mentioned are also highly relevant for chatbots, the CDCJ-ADMIN-AI proposes to consider general purpose of AI systems in the context of the revision of the handbook. It will also have to be considered carefully whether the use of generative chatbots for the purpose of communicating information or solving administrative problems requires the elaboration of a new principle, given that depriving a person of the possibility of communicating with officials at all in the context of administrative processes, may give rise to questions related to human dignity.

37. In light of the above considerations, the CDCJ-ADMIN-AI proposes to take into account issues raised by the use of generative chatbots to communicate information or solve administrative matters when updating the handbook.