Bergen Journal of Criminal Law and Criminal Justice • Volume 10, Issue 2, 2022, pp. 29-50.

AI-based Law Enforcement Online: The Impact of the European Artificial Intelligence Act (AIA)

INGER MARIE SUNDE*

Abstract

The article addresses provisions of the AIA affecting law enforcement authorities ('LEA'), particularly the implications on crime prevention and investigation online. The analysis demonstrates that online, LE officers have fewer means for meaningful presence and intervention against crime than in the physical domain. This calls for a nuanced approach in the regulation of LEAs' use of AI, to ensure that the AIA does not impede LE online. The proposal by IMCO/LIBE of the European Parliament, for a blanket prohibition against predictive policing thus seems too sweeping. The analysis further criticises that the restriction on LE use of biometric identification systems in publicly accessible spaces is not applicable to publicly accessible spaces online. It concludes that the provisions relevant to LE need improvement to ensure legal certainty and that needs for effective LE online are catered for.

Keywords

AIA, AI, Artificial Intelligence, PrevBOT, Online Policing, Law Enforcement Online, Predictive Policing, Algorithmic Policing, AI categorisation, AI identification.

* The author is Professor at the Norwegian Police University College (Politihøgskolen). This article was made in capacity of professor II in the research project 'Police and Prosecution Law' (2017-2022), financed by the Trond Mohn Foundation, the Norwegian Police Directorate and the Faculty of Law of the University of Bergen (https://www.uib.no/politiogpatalerett). I would like to thank the peer reviewer for constructive comments. In addition, warm thanks are extended to PhD-student Carlos Jose Calleja Ahmad of the Norwegian Police University College, for very helpful comments.

This is an Open-access article under the terms of the Creative Commons Attribution 3.0 Unported License (http://creativecommons.org/licenses/by/3.0/), permitting all use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The proposed European regulation on artificial intelligence ('AI'), the 'Artificial Intelligence Act' ('AIA'),[1] has stirred the concerns of the European Police Chiefs who warn that 'the rules formulated could seriously affect police work'.[2] The AIA is generally applicable to all public and private AI systems that represent more than a low or minimal risk to citizens' safety and fundamental rights.[3] It does, however, also regulate some *specific* AI practices, including practices of law enforcement agencies ('LEA'). Two provisions address biometric identification systems, a technology useful to LEAs in a variety of contexts (Art. 5(1)(d) and Annex III.1).[4] In addition, seven provisions in Annex III to the AIA ('A.III') specifically address AI systems 'intended to be used' by LEAs (A.III.6 paragraphs a – g), some of which concern the use of person-based AI in LE practices to predict crime.[5]

This article seeks to understand the meaning and scope of the provisions, and how they play out in practice, particularly with a view to LE online. To focus on the digital domain is interesting, as AI may play a different role to LE online compared to in the physical domain. Obviously, one aspect of AI is that it can be more powerful and efficient than a human (more power with less human resources). In this case, AI is essentially performing a task that also the LE officer could perform. A different aspect concerns AI performing tasks that *cannot* be performed manually, in which case AI adds capabilities that could not be performed by LE officers' personal faculties, or the technological artefacts they already possess.[6] If these

1 The European Commission, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts,* 21 April 2021, COM (2021) 206 final. References to legal provisions in this article are to those of the AIA unless otherwise indicated.

2 The European Police Chiefs, *Joint Declaration*, Berlin, 24 May 2022, p. 1.

3 Further explained in section 2 below. The scope of the AIA does not extend to encompass AI systems developed or used exclusively for military purposes (Art. 2(3)).

4 See e.g., Fuster, *Person identification, human rights and ethical principles – Rethinking biometrics in the era of artificial intelligence.* Scientific Foresight Unit, European Parliamentary Service, December 2021, PE697.191.

5 The legal relevance of Annex III follows from Art. 6, 'AI systems referred to in Annex III shall be considered high-risk.'

6 The two aspects correspond to the distinction in extension theory between technology that replicates or augments human faculties, and technology that introduce qualitatively new capabilities. See Brey, Theorizing technology's role in crime and law, in *The Routledge Handbook of Technology, Crime and Justice,* eds. McGuire and Holt (Routledge 2017), pp.17-34, p. 23.

capabilities are suitable to the mandate of LEAs,[7] and respect human rights, the AIA should not impede them. Thus the "AIDA-report" of the European Parliament '[s]tresses the importance of law enforcement agencies' ability to identify and counter criminal activity, aided by AI technology'.[8]

The provisions addressed in the present analysis concern use of person-based biometric identification and categorisation technologies, and profiling of persons. With one notable exception, the provisions do not distinguish between the use of AI in the physical and digital domain. The question is then whether they are sufficiently context-sensitive to cater for the needs of LE in the countering of online crime.[9]

Use of biometric identification and categorisation technologies as well as profiling of natural persons, involve automated processing of personal data. Thus, in the context of LE, they fall under the scope of the Law Enforcement Directive ('LED') as transposed into the Member States' legal systems.[10] The practices are furthermore subject to rules, procedures and safeguards that guarantee respect for fundamental rights, such as (but not limited to) rights to privacy, non-discrimination, freedom of speech and assembly, and to a fair trial. The AIA adds an additional layer to existing law and regulation, that is, with the objective to ensure that AI systems in the EU are safe and respect fundamental rights, and to enhance the effective enforcement of fundamental rights (including data protection rights).[11] The AIA does not provide legal basis for personal data processing. Rather, it creates a legal framework on top of existing data protection rules, aiming to deal with special risks associated with personal data processing by AI systems, and practices based on the output of such processing. Hence to be lawful the processing must *per se* have legal basis in existing

7 The mandate of LEAs is for instance described in *The European Code of Police Ethics*, Rec (2001) 10 of the Committee of Ministers, 19 September 2001, Title 1, section 1: 'The main purposes of the police in a democratic society governed by the rule of law are: – to maintain public tranquillity and law and order in society; – to protect and respect the individual's fundamental rights and freedoms as enshrined, in particular, in the European Convention on Human Rights; – to prevent and combat crime; – to detect crime; – to provide assistance and service functions to the public.'

8 AIDA-report. *Report on Artificial Intelligence in a Digital Age* (2020/2266(INI)), Special Committee on Artificial Intelligence in a Digital Age. European Parliament, 5 April 2022 (A9-0088/2022) para. 270.

9 The exception is Art. 5(1)(d) regarding the use of real-time biometric identification systems in publicly accessible spaces by LE, which only applies to physical places, see sections 2.1.b and 3.3.1 below.

10 The LED is Directive 2016/680.

11 *Explanatory Memorandum* to the AIA (the Commission's proposal, 2021), section 1.1, p. 2.

law.[12] Bryson's observation in her discussion of AI regulation, that '[a]ll human activity (…) occurs in the context of some sort of regulatory framework', therefore holds true in the present context.[13] In line with this, the question then is 'how to continue to optimize this framework in light of the changes in crime and LEAs' capacities introduced by AI and ICT more generally'.[14] It is of interest to LEAs that the joint IMCO/LIBE committee of the European Parliament, in its response to the Commission's Proposal, proposes to prohibit two LE practices deemed as 'predictive policing', which in the Proposal are categorised as lawful.[15] As will be explained, a prohibition may severely impact LEAs' use of AI online.

For LEAs, the AIA represents an important qualitative difference from the legal regime established by the LED, because, as a regulation the AIA has direct effect, meaning that, like the GDPR, it automatically becomes part of the national legal system.[16] In contrast, LED is a *directive* that must be transposed into national law to become effective.[17] A directive establishes goals to be achieved, not detailed rules. Acknowledging 'the special nature' of the field of LE, the LED was created as a parallel legal instrument to the GDPR, as the freedom afforded by transposition was believed suitable for catering to these needs.[18] The AIA deviates from this approach, by imposing detailed rules on LEAs with direct effect.

The structure in the following is first to give an overview of the relevant provisions within the risk-based framework of the AIA. Then follows an analysis of the provisions' scope and meaning, before the article zooms in on their applicability to LE online.

12 Personal data processing must have legal basis (LED Art. 8). This is in accordance with fundamental data protection principles, cf. LED Art. 4(a), and the General Data Protection Regulation ('GDPR') (Regulation 2016/679) Art. 5(1)(a).

13 Bryson, The Artificial Intelligence of the Ethics of Artificial Intelligence – An Introductory Review for Law and Regulation, in *The Oxford Handbook of Ethics of AI*, (OUP 2020), Kindle location 370.

14 *Ibid*. The original quote is 'how to continue to optimize this framework in light of the changes in society and its capacities introduced by AI and ICT more generally'. Kindle location 403.

15 IMCO/LIBE, *draft report on the proposal for [an AIA]*, Committee on the Internal Market and Consumer Protection ('IMCO'), Committee on Civil Liberties, Justice and Home Affairs ('LIBE'). European Parliament, 20 April 2022. (COM2021/0206(COD).

16 The Treaty of the Functioning of the European Union ('TFEU'), consolidated version, 7 June 2016, C 202/47, Art. 288, 2nd para.

17 TFEU Art. 288, 3rd para.

18 The Preamble to the LED, Recital 10 ('PR' 10).

2. Placing the police provisions in the risk-based legal framework

2.1 The risk-based approach

The AIA applies a risk-based approach. By application of the principle of proportionality, the level of obligations and restrictions imposed on AI practices seeks to reflect their risk to fundamental rights and citizens' health and safety.[19] Of four risk categories that have been identified, the AIA encompasses three. [20]

a) Police practices representing unacceptable risk – Article 5(1)(d)

Article 5 concerns 'prohibited practices',[21] i.e., practices that contradict 'respect for human dignity, freedom, equality, democracy, and the rule of law, and (…) fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child'.[22] Article 5(1)(a) – (c) prohibits practices considered to be 'manipulative, exploitative, and social control practices'.[23] In addition, paragraph d) prohibits the use of

> 'real-time' biometric identification systems in publicly accessible spaces for the purpose of law enforcement.

This provision is only applicable to publicly accessible spaces in the physical domain. This follows from the definition of 'publicly accessible space' in Art. 3(39), which only includes 'any physical place accessible to the public'. The Preamble further underlines that online spaces are not covered 'as they are not physical spaces'.[24]

To characterise paragraph d) as a 'prohibition' is misleading, as a close reading reveals the under-communicated fact that it *permits* such use, however only in three narrowly defined situations (Art. 5(1)(d)(i) – (iii)), *if* authorised in national law, conditions of necessity and proportionality are met, *and* judicial or administrative permission is granted (Art. 5(2) – (4)).[25] Put differently; LEAs' use of 'real-time'

19 *Explanatory Memorandum* to the AIA (the Commission's proposal, 2021), section 2.3, p. 7 and PR 14. This is in accordance with the strategy laid down in the Commission's *White Paper On Artificial Intelligence - A European approach to excellence and trust*, 19 February 2020, COM(2020) 65 final, p. 17-18. See also the AIDA-report, 2022, para. 130-134.

20 The risk categories are set out in Art. 1(a)-(c).

21 Article 5 is located in the AIA Title II 'Prohibited Artificial Intelligence Practices'.

22 PR 15.

23 *Ibid.*

24 PR 9.

25 See also PR 18-23.

biometric identification systems in publicly accessible spaces is not prohibited if the Member State has authorised the practice in its national law according to the conditions set out in Art. 5. In such case, the practice is categorised as 'high-risk' (see point b below).

Article 5(1)(d) – (4) should be read in conjunction with A.III.1, which generally categorises AI systems intended to be used for *biometric identification without consent of the exposed person,* as 'high-risk'. The provision governs all (AI based) biometric identification systems irrespective whether the use takes place in 'real-time' or later ('post'), or who makes use of the systems, hence applies also to LEAs.[26] Thus the 'prohibition' in Art. 5 is but a special instance of such 'high-risk' systems, imposing stricter conditions for use *when performed in 'real-time' by LEAs in publicly accessible spaces.* Corresponding use in publicly accessible spaces by other actors (private or public), falls under the scope of A.III.1.

The European Data Protection Board ('EDPB') and the European Data Protection Supervisor ('EDPS') have voiced strong opposition against biometric identification of persons in publicly accessible spaces, and call for a general ban on 'any use of AI for an automated recognition of human features in publicly accessible spaces - such as of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals - *in any context*' (italics added).[27] IMCO/LIBE do, however, not object to the provision.[28] The view of the EDPB-EDPS has, for its part, broad support among civil liberties groups, as expressed in the European Citizens Initiative from 2021.[29]

b) 'High-risk' police practices – Annex III(6)

The category 'high-risk' comprises AI systems 'that have a significant harmful impact on the health, safety and fundamental rights of persons in the Union'.[30] High-risk AI systems may only be put on the market or used if they comply

26 'Real-time' and 'post' biometric identification systems are defined in Art. 3(37) and (38). The Preamble mentions the risk that use of biometric identification systems 'can lead to biased results and entail discriminatory effects,' hence 'real-time' and 'post' biometric identification systems should be classified as high risk' (PR 33).

27 EDPB-EDPS *Joint Opinion on the proposal for an Artificial Intelligence Act*, 5/2021, para. 32, pp. 11-12.

28 IMCO/LIBE, 2022, p. 54.

29 'Civil society initiative for a ban on biometric mass surveillance practices', 7 January 2021.

30 PR 27. High-risk practices are defined in Art. 6, which apart from describing some fixed criteria, refers to the list in A.III. The list can be amended according to powers and procedures set out in Art. 7, cf. Art. 73.

with the mandatory requirements set out in the AIA Title III. These amount to a comprehensive set of obligations pertaining to the AI systems as such, the providers and users, oversight and more.

High-risk uses of AI by LEAs are listed in A.III.6. IMCO/LIBE call attention to use of AI in 'predictive policing' (paragraphs a and e), and propose to prohibit the practice as it 'violates human dignity and the presumption of innocence, and (…) holds a particular risk of discrimination'.[31] The position has strong support, voiced, for instance, by the international NGO Fair Trials, who in terms of technology, takes a broader approach that includes automated decision-making systems along with AI. Fair Trial asserts that the use of such systems 'reproduce and reinforce discrimination on grounds including but not limited to race, socio-economic status, and nationality, as well as engage and infringe fundamental rights, including the right to a fair trial and the presumption of innocence, the right to private and family life, and data protection rights'.[32] The European Council appears however to agree with the Commission in that the practice should be deemed as 'high-risk'.[33]

The discussions in this article concern the provisions in A.III.6 that address LEAs' use of AI for profiling natural persons to predict their relation to crime in the past, present or future, i.e., paragraphs a), e) and f).[34] The topic involves use of biometric identification and categorisation, which can be seen as special instances of profiling. This activates A.III.1 as well. Paragraph g) which concerns 'crime analytics regarding natural persons' could also be relevant, but differs from the other provisions in that the 'analytics' possibly are performed on aggregate level, as opposed to AI generated output on individual level. Yet, the scope of paragraph g) is not quite clear, as it mentions 'unknown patterns' and 'hidden relationships' that can be uncovered by search of 'complex related and unrelated large data sets' relating to natural persons. There will at least be a grey zone in the interface between paragraph g) and paragraphs a), e) and f). The Council however has proposed to delete paragraph g) from the list in A.III.6.[35]

31 IMCO/LIBE, 2022, p. 54.

32 Fair Trials, *Automating Injustice: The use of Artificial intelligence & Automated Decision-making systems in Criminal Justice in Europe,* 9 September 2021, p. 4. The term 'ADM' 'encapsulates all systems which process data and other inputs and produce outputs, which influence or assist with human decisions, to different degrees', p. 6.

33 Presidency of the Council of Europe, *Compromise Text, Proposal for an Artificial Intelligence Act*, 14278/21, 29 November 2021, p. 98.

34 Practices falling under the scope of A.III.6 paragraphs b), c) and d) fall outside the scope of the analysis. These concern the use of AI 'as polygraphs and similar tools or to detect the emotional state of a person', 'to detect deep fakes' and 'for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences'.

35 *Presidency Compromise Text*, 2021 para 3 p. 5, and p. 98.

c) Limited risk – AI systems subject to transparency obligations – Art. 52.

AI systems in this category are subject only to transparency obligations (Art. 52), not having to comply with the more demanding conditions mandatory to high-risk systems.[36] Examples are AI systems that interact with humans (chatbots, for instance), emotion recognition systems, biometric categorisation systems, and AI systems that generate or manipulate image, audio or video content ('generative AI'). According to Art. 52, the transparency obligation does not apply to AI systems used for crime detection, prevention, investigation or prosecution, obviously because it could jeopardise these purposes at the cost of societal interests in safety and law enforcement.

d) Low or minimal risk

AI uses deemed to represent low or minimal risk are not regulated by the AIA.[37] Some LE uses of AI fall into this category, e.g., AI generated predictions regarding *objects* or *places* (as opposed to persons). Moreover, if paragraph g) is deleted from A.III.6, as proposed by the Council, also crime analytics regarding natural persons will be out of scope of the AIA.[38]

2.2 The interplay between the risk categories

The interplay between the risk categories is not entirely straightforward. For instance, to conclude that LEAs' use of chatbots or biometric categorisation systems are of low or minimal risk because they are exempted from transparency obligations, seems unreasonable. In accordance with the underlying proportionality principle and to realise the purpose of the risk-based approach, it makes more sense to attribute each AI practice to the strictest risk category applicable, based on the description most loyal to the purpose of the AIA. For instance, to assess a LE chatbot only in relation to Art. 52 because it mentions 'chatbot', leads one to overlook that the chatbot's purpose in a concrete case could be to single out possible criminal offenders online. This should obviously be deemed as high-risk, and an example is discussed in section 3.3 below. Hence to the extent that LE practices that fall under the scope of Art. 52 are comprised by high-risk provisions as well, they must comply with the mandatory obligations pertaining to high-risk AI systems. They may, however, be used without any obligation to be transparent to the affected person.

36 PR 70. IMCO/LIBE, 2022, has proposed deep fakes that resemble existing persons, or appear to be the authentic text of a person, also to be categorised as high-risk, p. 154, point 8a.

37 Cf. Art. 1(a)-(c).

38 *Ibid.,* fn. 35.

To summarise the location of the provisions in the present analysis; the use of biometric identification systems without consent of the targeted person is high-risk as per A.III.1, and when used in real-time by LEAs in publicly accessible physical places, it is subject to a special set of strict conditions (Art. 5(1)(d) – Art. 5(4)). Moreover, the use of profiling in certain LE practices is deemed as 'high-risk' as per A.III.6 paragraphs a), e) and f). Paragraph g) (crime analytics) is proposed to be deleted from the high-risk list, hence be out of scope of the regulation. LEAs' use of practices mentioned in Art. 52 is exempted from the transparency obligation towards the affected person.

3. Legal analysis of A.III.6 paragraphs a), e) and f)

3.1 Introduction

This section analyses paragraphs a), e) and f) in A.III.6. The paragraphs have in common that they concern person-based use of AI by LEAs. 'Person-based' means that the purpose for using the AI system is to generate predictions about a person through the processing of data about that person. 'Person-based' can be contrasted to 'object- and place-based' uses of AI, such as automatic number plate recognition (the number plate is an object) and geospatial analysis to detect criminal hot spots.[39] The analysis sets out to understand the provisions' meaning and scope. It then proceeds by examining their applicability to LE online.

3.2 The meaning and scope of A.III.6 paragraph a), e) and f)

Paragraphs a), e) and f) in A.III.6 read as follows:

'AI systems intended to be used by law enforcement authorities **or on their behalf** for (…)':[40]

> (a) (…) making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for **a natural person to become** a potential victim of criminal offenses.[41]

39 The purpose-oriented terminology applied in this analysis does not reflect a technical differentiation between AI technologies. Technology-wise, AI may treat human features as objects, e.g., automated fingerprint analysis and face recognition. To the AI system, fingerprints and faces are objects, comparable to stolen antiquities and firearms (the recognition of which could also be useful to LEAs).

40 Emphasis is added to text proposed to be included by the Council. *Presidency Compromise Text*, 2021, p. 98.

41 *Ibid*, p. 98.

> (e) (…) predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 [the LED], or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups.
>
> (f) (…) profiling of natural persons as referred to in Article 3(4) of [the LED] in the course of detection, investigation or prosecution of criminal offences.

Each provision describes a task for the AI system to perform. The task is to process personal data in order to produce an output. The term 'personal data' is defined in the LED Art. 3(1),[42] and forms part of the term 'profiling' applied both in paragraph e) first alternative ('alt.'), and paragraph f). 'Profiling' is defined in the LED Art. 3(4) as follows:

> any form of *automated processing* of personal data consisting of the use of personal data to *evaluate certain personal aspects* relating to a natural person, in particular to analyse or *predict* aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements (italics added).

That the processing entailed with paragraph a) and the second alternative of paragraph e) concerns personal data – just like paragraph e) first alt., and paragraph f) – follows implicitly from the stipulated task. In paragraph a) it concerns an '*individual risk assessment of natural persons*', and in paragraph e) second alt.: '*assessing personality traits and characteristics or past criminal behaviour*'. Obviously, to produce such assessments, processing of personal data is a necessity.

Data processing in an AI system is based on statistical computation. Hence the output is a probability statement which, by definition, entails uncertainty. For example, by using AI to assess the reliability of a person (reliability is a 'personality trait' as per paragraph e) second alt.), the output does not say for sure that the person is reliable (or not), rather offers an indication in one direction or another. Often probability statements are referred to as 'predictions', a term expressing the inherent quality of uncertainty.[43] Literally, 'prediction' means a statement about a future event, but as shown by the example, here it means a statement about the

42 'Personal data' means 'any information relating to an identified or identifiable natural person (data subject); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier, or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'.

43 See, e.g., Parker, What's the Difference Between AI & Predictive Analytics? *Technology Advisors*. 29 March 2021.

likelihood of a fact, and it is not required that the fact lies in the future. For example, it may concern whether a person is reliable, has a psychological condition, is in economic trouble, is likely to commit a crime, likely has committed a crime, and so on.

Paragraphs a) and e) do not only stipulate the task of the AI system, but the objective as well, that is, the purpose for which the AI system is used. In paragraph a) the objective is to predict *the risk of a person for (re)offending or becoming a victim of a crime,* and in paragraph e) to predict *the (re)occurrence of an actual or potential crime* (first alt.), or, to predict (assess) certain *personality traits and characteristics or past criminal behaviour* of individuals or groups (second alt.).

Futhermore, paragraph e) first alt. sets as a condition that the prediction must be 'based on' profiling, which as mentioned, is a defined term. In contrast, the kind of data processing is not specified in the second alternative, nor in paragraph a), as the words 'assessing/assessment' are not defined terms in European data protection law. This leaves a scope for interpretation in the individual case.

Paragraph f) differs from the others in that it does not set an objective for the data processing (profiling). It merely mentions the context, that is, 'in the course of detection, investigation or prosecution of criminal offence'. An example could be to create profiles similar to those of past offenders, or to understand the *modus operandi* in a series of cases sharing common features in the course of a criminal investigation.

Now, having identified the provisions' core elements, the variety of notions ('profiling' and 'assessing/assessment' (collectively referred to 'assessment')) strikes one as odd. 'Profiling' is an elastic term encompassing 'any form' of automated processing of personal data 'evaluating certain personal aspects' relating to a person. To exemplify, the definition includes a list of that which 'in particular' may constitute profiling. Predictions about a person's likelihood to (re)offend or become victim of a crime (paragraph a), or about a person's personality traits, characteristics or past criminal behaviour (paragraph e), may fall directly under 'evaluating certain personal aspects' of a person, and even be subsumed under 'behaviour' among the examples. This should lead to the conclusion that all the said provisions, despite the varied terminology, concern profiling.

Yet, a difference between 'profiling' and 'assessment' can still be pointed out. 'Profiling' is by definition 'automated processing' of personal data. This entails that profiling as a task can only be performed by computer systems. It follows that 'profiling' is a concept that excludes human judgement. An 'assessment' can however be made by a human, in which case paragraphs a) and e) second alt., may

be interpreted as concerning the joint product of the AI generated predictions and human judgement. Put differently, 'assessment' may refer to use of the AI system as decision support in broader assessments made by LE officers. Such assessments may include more information than the AI generated output, available to that person.

However, the prohibition against decisions 'solely' based on automated processing of personal data, including profiling, speaks against an interpretation that differentiates between 'profiling' and 'assessment' in the provisions analysed here. The prohibition is laid down in the LED Art. 11, and concerns decisions that produce 'an adverse legal effect' on the targeted person or 'significantly affect' him or her.[44] Decisions made by LEAs concerning individuals, whether for purely preventative intervention or in a criminal investigation, often affect that person in a significant manner. Thus, profiling as mentioned in paragraph e) first alt., and paragraph f) cannot lawfully be used to produce decisions detached from meaningful human assessment which includes other information at hand. In fact, this point is rather fundamental to the regulatory approach, where a concern is that humans 'place such confidence in AI that they trust it more than their own judgement'.[45] To protect human dignity, the AIA thus imposes safeguards to ensure that decisions affecting individuals effectively are made by humans.[46] This point is of course highly relevant to decisions made by LEAs.

This leads to the overall conclusion that there is no difference in meaning between 'profiling' and 'assessment' in the said paragraphs, and that assessment in paragraph a) and e) second alt., must be interpreted as special instances of profiling. As shown by the discussion, the varied terminology gives rise to interpretative questions that could be avoided by a consistent use either of 'profiling' or 'assessment' throughout. This is a weakness giving rise to legal uncertainty. This should be remedied before the AIA becomes final.

Another question to be raised is whether there is any overlap between paragraph e) first alt., and paragraphs a) and f). Recalling that paragraph e) first alt., concerns profiling of a person to predict the (re)occurrence of an actual or potential crime, one may ask if this simply mirrors paragraph a) which – as just concluded – concerns profiling of a person to predict the risk of that person for (re)offending. Crime is committed by persons, and it is hard to see the difference between profiling to calculate the likelihood that *a crime* will be committed by that person

44 For an understanding of the provision, see Mendoza and Bygrave, The Right not to be Subject to Automated Decisions based on Profiling, *University of Oslo Faculty of Law Legal Studies-Research Paper Series* (2017) no. 20. The article deals with Art. 11's counterpart in the GDPR Art. 22.

45 AIDA-report, 2022, para. 57.

46 Cf. Art. 14 about human oversight.

(paragraph e), and that *that person* will commit a crime (paragraph a). The problem of discerning a meaning distinctive to each paragraph is a further source of legal uncertainty. However, Paragraph a) includes also an alternative concerning a probable victim of a crime. As this alternative does not have a pendant in paragraph e) it could be dealt with in a separate paragraph.

The question of overlap between paragraph e) first alt., and paragraph f), stems from the expressions 'actual … criminal offence' (paragraph e) and 'detection … of criminal offences' (paragraph f). Both expressions comprise crime in the present, for instance ongoing organised crime.[47] How to draw a line between these alternatives is not clear from the wording. A possibility is to lay emphasis on the ultimate objective of the profiling. To clarify: paragraph e) states that the immediate purpose of the profiling is 'predicting the (re)occurrence of an actual … criminal offence'. This does however not say anything about the objective for which the information gained by the profiling, shall be used. With respect to 'potential' crime (also mentioned in paragraph e) first alt.), the objective is possibly to intervene proactively in accordance with LEAs' mandate to prevent and avert crime. Profiling to predict 'actual' crime would however call for action to stop it and open a criminal investigation. This comes close to the objective for use of profiling for 'detection … of criminal offences' as mentioned in paragraph f).

An option is to seek recourse to the distinction between crime prevention and criminal prosecution, reserving paragraph e) for the former and paragraph f) for the latter. However, a LEA cannot bind itself in advance as to the response it will apply to indications of an actual crime. An initial aim of solving a crime problem by preventative intervention may be overtaken by an obligation to open a criminal investigation. In the end, a decision about the action to be taken depends on the concrete circumstances of the situation.

Furthermore, crime 'detection' in paragraph f) is not a defined concept in criminal procedural law. This stands in contrast to 'investigation' and 'prosecution' (also mentioned in paragraph f). Investigation is a purpose-oriented, evidence gathering activity, aimed at clarifying whether a criminal offence was committed and who did it. Prosecution is the process of holding the perpetrator to justice in criminal proceedings. In both cases, the law clarifies the procedural status of the activity of the LEA. Thus, a criminal investigation presupposes a decision made by a competent person on basis of information providing 'reasonable suspicion/reasonable ground'

47 Of course, crime detection may also concern crime in the past, but use of AI for that purpose does not raise a question of overlap.

to believe that a crime was committed.[48] Prosecution demands a criminal charge (indictment). Crime 'detection' may however refer to intelligence activity aimed at uncovering crime that later could be investigated and prosecuted. Use of profiling in crime detection in support of an ongoing criminal investigation, could be covered by 'investigation' in paragraph f). If so, to have an independent meaning, crime 'detection' must encompass profiling outside the scope of a criminal investigation. As one can see, this comes close to paragraph e) first alt., as explained above.

LEAs may have an interest in such use of profiling, in order to produce information (intelligence) *that may establish* reasonable ground to open a criminal investigation. In the event that use of AI for predictive policing becomes prohibited, as suggested by IMCO/LIBE, it will become crucial to know whether also this kind of use will be prohibited. The critique by IMCO/LIBE concerns 'predictive policing'.[49] This is however hardly a settled concept, hence as noted by Egbert and Leese, it is necessary 'carefully to define what we are speaking about when we refer to predictive policing'.[50] Objections against predictive policing in the context of crime prevention may also be relevant to use of AI to gather intelligence aimed at informing LEAs about priorities to be made regarding which crime and which criminals that should be subject to further investigation and prosecution. If nothing else, the risk that use of profiling reinforces discriminatory LE practices against marginalised groups is a point in case. It is therefore not clear how far-reaching a prohibition against predictive policing will be. The uncertainty could impact the interpretation of paragraph f), perhaps limiting it only to concern crime detection in the course of a criminal investigation. Again, the issues arising from the wording of the paragraphs show a need for clarification. This is underlined by the direct effect of the AIA, which in order to attain its goal of horizontal legal harmonisation, must make use of unambiguous terms and set clear scope for the provisions.

48 The European Code of Police Ethics, 2001, section 47 says that 'Police investigations shall, as a minimum, be based upon *reasonable suspicion* of an actual or possible offence or crime' (emphasis added). The Norwegian Criminal Procedural Code ('NPCC') of 22 February 1981 no. 25, section 224, requires 'reasonable ground'.

49 Cited in section 2.1.b.

50 Egbert and Leese, *Criminal futures – Predictive policing and everyday police work* (Routledge 2021), p. 19. Fyfe, Gundhus and Rønn, Introduction, in *Moral Issues in Intelligence-Led Policing* (Routledge 2018), note that there is 'is far from being a single agreed definition of terms such as intelligence, predictions, risk assessment, pre-emption, prevention, etc.' p. 4.

3.3 Person-based AI predictions used by LEAs online

3.3.1 Introduction

This section analyses the provisions' applicability to person-based LE use of AI online. Challenges with crime detection, prevention and investigation online differ from the physical domain in some important ways that may call for different capabilities. The question is whether the AIA is adequate in this respect, thus enabling effective LE online within the framework of human rights. Admittedly, the regulators have a difficult task at hand, as no clear concept exist of crime prevention and LE presence, in online spaces. Such activities can be performed in various ways, both manually and automatically, and with different degrees of human presence.[51] Presence is about being in the situation, and as humans do not have a presence in online spaces that equals physical presence, the measures at LEAs' disposal might also have to be different.[52] 'Presence' should not be confused with 'human oversight', which always must be provided for.[53]

Information about online users is often reduced to a user profile and a chat, thus preventing LE officers from visual gathering of information about persons' appearances and identities the way they do in physical space. Use of anonymising technologies adds to the challenges special to cyberspace. This provokes a need for gathering information in new ways by new technology. The practices that thus attract interest are those representing capabilities that cannot be performed by humans, capable of piercing the veil of anonymity and enabling LE officers to 'see' a user's real appearance. According to the conceptual framework of extension theory, the use of AI for such purpose extends human cognitive function, introducing a qualitatively novel LE capability.[54] This perspective can be contrasted to the one applied in the Commission's White Paper on AI, which emphasises that 'AI can perform many functions that previously could only be done by humans'.[55] In that perspective AI merely replicates or augments human faculties.[56]

To make the analysis concrete the 'PrevBOT concept' is used for illustration. Two articles by Nina Sunde and Inger Marie Sunde describe an AI tool aimed at

51 Sunde, Patruljering på internett [Policing the Internet], in *Straff og frihet – til vern om den liberale rettsstaten. Festskrift til Tor-Axel Busch,* eds. Sæther et al. (Gyldendal 2019), pp. 597-608.

52 *Ibid.,* p. 604-608.

53 Human oversight is mandatory for 'high-risk' AI systems, cf. Art. 14.

54 Brey 2017, p. 23 and 26.

55 *White Paper on AI,* 2020, section 5.A, p. 11.

56 Brey 2017, p. 23.

countering the problem of child sexual exploitation and abuse ('CSEA').[57] The origin of many CSEA cases is encounters between adults and minors in publicly accessible chat forums online, where would-be offenders often pose with false user profile and give incorrect and misleading personal information about age and/or gender. PrevBOT – a chatbot – is set to operate in publicly accessible chat fora, and by real-time processing of the data in chat conversations, it may automatically (i) *categorise* conversations as sexualised or non-sexualised; (ii) *categorise* online participants into specific classes of age and gender; and (iii) *identify* previous CSEA convicts. The tool aims first to identify publicly accessible online forums that pose a risk of CSEA. The forums should be such that *de facto* are used by children, where there is sexualised speech, and adults who misrepresent themselves and seek contact with children. By monitoring these spaces more closely, the police may be able to intervene before CSEA is committed, by submitting a 'preventative warning' to the would-be offender. In addition, PrevBOT may identify former CSEA convicts who resume the criminal activity online. Often, the offenders have many victims. Hence, identification of a previous CSEA convict in this context may give reasonable ground to open a criminal investigation. Apart from leading to apprehension and prosecution, this may end the continued abuse of many children.

The question concerns the applicability of the AIA provisions to PrevBOT's functions. Here, a mapping of PrevBOT's functions and relevant legal notions is needed. PrevBOT's predictions are based on automated text-analysis applied to chat conversations. As the data of these conversations is personal data as per the LED Art. 3(1), PrevBOT's functions are realised through automated personal data processing.[58] The predictions serve three functions, one place-based and two person-based. The place-based function concerns place detection (spaces with CSEA risk), while the person-based functions are about categorisation and identification of persons. These functions are realised through processing of biometric data as defined in Art. 3(33):

57 The PrevBOT articles are published in *Nordic Journal of Studies in Policing* (open access) with the title 'Conceptualizing an AI-based Police Robot for Preventing Online Child Sexual Exploitation and Abuse'. Article 1: Part I – The Theoretical and Technical Foundations for PrevBOT, Sunde and Sunde, *NJSP,* 1/2021; Article 2: Part II – Legal Analysis of PrevBOT, Sunde and Sunde, *NJSP* (accepted for publication).

58 Data is 'personal' if the person is 'identifiable' directly or indirectly by use of information in that data. Hence the notion of 'personal data' is extremely broad (LED 3(1) is cited in fn. 42). See also Purtova, The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology* (2018), 10:1, pp. 40-81.

> personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which *allow or confirm the unique identification* of that person, such as facial images or dactyloscopic data (italics added).

The definition corresponds to the definition of 'biometric data' in the LED Art. 3(13), and the term forms part of the definitions of 'biometric identification system' and 'biometric categorisation system' set out in the AIA Art. 3(35) and (36).[59] However, there is a problem with the phrase 'allow or confirm the unique identification' (italicised above), as it is difficult to square with the notion 'biometric categorisation'. The latter is obviously not about identifying persons, rather assign them to categories based on information from their biometric data.[60] To solve the problem, IMCO/LIBE has proposed to add the related notion 'biometrics-based data' which includes the phrase 'may or *may not* allow or confirm the unique identification' of a person (italics added). In such case both terms will form part of the definition of 'biometric categorisation system'.[61]

For the present discussion, the important point is that biometric/biometrics-based data is derived from personal data through technical processing. This is exactly what PrevBOT does. The extracted data concern linguistic behaviour in the chat conversations. With respect to categorisation, the linguistic behaviour enables the AI system to predict the true age and gender of the participant in the chat. This can unmask adults who pretend to be children or misinform about their gender. This function of PrevBOT is possibly a 'biometric categorisation system' within the meaning of Art. 3(35), which is dealt with in Art. 52. However, as LEAs are exempted from the transparency obligation concerning such systems, the interesting questions arise in relation to the provisions in A.III.6.

With respect to identification, the linguistic behaviour may be unique to the person, i.e., a 'linguistic fingerprint', that may be compared to linguistic fingerprints of CSEA convicts in a LE database. This function of PrevBOT is a 'biometric identification

59 Cf. Art. 3(35) and (36).

60 Biometric categorisation system is defined in Art. 3(35) as 'an AI system for the purpose of assigning natural persons to specific categories, such as sex, age, hair colour, eye colour, tattoos, ethnic origin or sexual or political orientation, on the basis of their *biometric data*' (italics added). The Council (2021) has proposed to add the categories 'health' and 'personal traits'. IMCO/LIBE (2022) proposes in addition to add 'gender', 'mental or physical ability, behavioural or personality traits' (thus amending 'personal' to 'personality'). For a conceptual discussion of biometric categorisation, see Fuster 2021, pp. 16-19.

61 IMCO/LIBE, 2022, point 33a, p. 49, and point 35, p. 51.

system' as per Art. 3(36).[62] Both the linguistic fingerprint and the comparison that might generate a 'match' are predictions. To account for the uncertainty, the AIA requires two persons to verify and confirm the match prior to any 'action or decision' performed on basis of it (Art. 14(5)). Interestingly, the provision does not require the persons to be experts within the field.[63]

A first question is whether application of the biometric identification function in publicly accessible chat fora is restricted as per Art. 5(1)(d). Clearly, the provision only covers physical places (explained in section 2.1.b). The Preamble does not offer any explanation for this limitation, and it is hard to see why physical and digital spaces are treated differently in this respect.[64] However, it means that PrevBOT is not restricted by Art. 5. Nonetheless and reasonably enough, it is still 'high-risk', as per A.III.1.[65]

A second question is whether the biometric categorisation and identification functions fall under the scope of A.III.6 paragraph a), e) or f). If paragraphs a) and e) are converted into prohibitions, the question is then whether this unduly restricts LEAs' capability to prevent online crime. The place-based function does not seem to raise questions in relation to paragraph a), e) and f), as it is not oriented towards persons. For this function to be lawful, it is sufficient that it complies with the conditions of the LED as transposed into national law.[66]

3.3.2 Discussion of the applicability of A.III.6 paragraphs a), e) and f) to LE online

The question whether the use of PrevBOT's categorisation and identification functions fall under the scope of A.III.6 a), e) or f) raises a general issue concerning the relevance of contextual information. The point is that PrevBOT's predictions do not *per se* say anything about risk. Their value in that respect is totally dependent

62 The definition reads: An AI system 'for the purpose of identifying natural persons through the comparison of a person's biometric data with the biometric data contained in a reference data repository'.

63 It should be noted that uncertainty regarding a 'match' not only stems from the prediction, but also from the lack of evidence that characteristics believed to be unique to a person, really are unique (the 'black swan' problem). This should require the human controllers to be experts in the field. See Saks and Koehler, The Individualization Fallacy in Forensic Science Evidence. *Vanderbilt Law Review* (2008) 61 (1) pp. 198-219; Saks, Forensic Identification: From a faith-based 'Science' to a scientific science. *Forensic Science International* (2010) pp. 14-17.

64 Such a critique is fronted by the EDPB-EDPS, 2021, quoted in section 2.1.b above, which underlines that the prohibition should be applicable 'in any context'. EDPB-EDPS concludes that 'for consistency reasons, AI systems for 'large-scale remote identification in online spaces' should be prohibited under Article 5 of the Proposal', para 32, pp. 11-12.

65 See section 2.1.b above.

66 See Sunde and Sunde, *NJSP* (paper accepted for publication).

on the availability of supplemental (contextual) information. This stands in contrast at least to paragraphs a) and e) first alt., as both seem to require a causal relation between the AI generated output and a specific risk. To illustrate; the 'individual risk assessment' mentioned in paragraph a) comprises the AI generated prediction. The same is the case for the use of 'profiling' to predict the '(re)occurrence of an actual or potential crime' as per paragraph e) first alt. PrevBOT's prediction of age/gender or identity does however not have any ostensible relevance to the risk that the person commits a crime. For the prediction to have value for that purpose, it must be combined with other information at hand, which as a minimum involves a CSEA-risky place and the participant's initiative to contact children (including PrevBOT posing as a child). It is this context that allows misrepresentation of age/gender to be understood as indicating that a person whose identity is unknown, may commit a CSEA crime. The same is the case with identification of a former CSEA convict. Once a person is discharged from punishment, s/he is entitled to the same fundamental rights and freedoms as any other free citizen. This includes the right to be presumed innocent and be left in peace from LEAs. To respect this right, the context of a CSEA-risky place must be established prior to use of the identification function, and identification is only relevant in that context.

It was concluded already that the paragraphs must be interpreted in light of the prohibition against automated decision-making in the LED Art. 11. With respect to PrevBOT, issuance of a preventative warning may constitute a decision that seriously affects the recipient, and the opening of a criminal investigation on basis of a linguistic fingerprint of a CSEA convict, may produce an adverse legal effect on that person, who from that point in time is deemed as suspect of a crime.[67] It therefore follows that PrevBOT's predictions may only lawfully be used as support in decisions made by LE officers.

Still, there is an anomaly in that PrevBOT's predictions do not relate directly to the objectives set by paragraph a) and e) first alt. This could lead PrevBOT's assessments and profiling to be deemed as falling outside the scope of the provisions. Thus, if the practices covered by these provisions were to be prohibited, it would not affect use of PrevBOT. In contrast, the second alternative of paragraph e) mentioning 'past criminal behaviour' is applicable to PrevBOT's identification function, whereas assessing 'personality traits and characteristics' could be relevant to predictions of age/gender, as they expose the willingness of an adult to seriously misrepresent him-/herself in order to succeed in getting in contact with children.

67 In order to target individual persons, PrevBOT processes personal data of everyone on the chat room. The data protection issues relating to persons whose data is processed without resulting in any alert that trigger further action by the LEA, are discussed in Sunde and Sunde, article accepted for publication, see fn. 61, *supra.*

Paragraph f) does not specify the objective of the profiling. It only states that it must take place 'in the course of detection, investigation or prosecution' of a criminal offence. In that context, to combine the prediction with other information available to the LEA, is obviously relevant and within the scope of the provision.

Insofar as use of PrevBOT is comprised by the provisions, it cannot be used if paragraphs a) and e) are converted into prohibitions, hence cannot be used for preventative purpose by detecting would-be offenders and submit a preventative warning. This could be a severe push-back against LEAs' efforts to counter widespread sexual abuse of children, where the need for proactive preventative initiatives seems clear. As explained, the challenges confronting LEAs online are different from in physical space where persons' appearances are visible. Moreover, in the case of a prohibition, it will be crucial to clarify whether PrevBOT still may be used for generating data relevant to the condition 'reasonable suspicion/ground', with a view to opening a criminal investigation. This depends on the scope of paragraph f) which was discussed in section 3.2. If such use falls outside the scope of crime 'detection' in paragraph f), and paragraphs a) and e) become prohibited, then neither this function may lawfully be put into use.

The result seems anomalous in relation to Art. 5(1)(d) which permits a Member State to authorise use of biometric identification systems in publicly accessible physical spaces by law enforcement. A corresponding possibility should be established for online use as well, thus at least opening for use of PrevBOT's identification function. One should even consider to go further and open for use of the categorisation function as well, for the LEAs to gain a better understanding of who is in a CSEA-risky place. A problem though is the borderless environment on the internet. It may lead to situations where LEAs in a Member State which has authorised such online uses of biometrics-based AI-tools, may come to act against individuals residing in a Member State where it is not authorised. This problem does not arise with respect to use of the technology in physical space, and causes new questions relating to suitable regulation of LE online.

Misgivings against predictive policing are mainly voiced in a context involving LE in physical space, and admittedly there is much evidence to suggest that the reliance on automated decision-making systems may lead to unfair practices.[68] However, blanket prohibitions may have unintended side effects with severe consequences for LE online. That the online dimension of crime is becoming increasingly serious, and in fact dominates how many crimes, including so-called traditional crimes, are committed in the 21st century, is not in dispute. As demonstrated by the analysis,

68 There is a wealth of literature covering predictive policing, but see for instance Fyfe, Gundhus and Rønn 2018; Egbert and Leese 2020; Brayne, *Predict and Surveil* (OUP 2021), and Fair Trials 2021.

in the cyber context, LEAs have a need for special tools to be able to intervene in a meaningful way to prevent crime. This should lead the EU law-making bodies to apply a nuanced approach ensuring that the AIA include provisions reflecting important differences in the challenges human LE officers are confronted with, depending on whether they work against crime online or offline. The example of PrevBOT shows that such use can be performed in a targeted manner without encroaching on the fundamental rights of citizens.[69] The AIA should therefore not hinder such use of AI by LEAs.

4. Summary and conclusion

The analysis has demonstrated that LE officers have fewer means for meaningful presence online and for intervening against online crime, than they have in the physical domain. This should call for a more nuanced approach to the provisions in the AIA, to ensure that they not impair the legal basis for reasonably effective LE online. Blanket prohibitions against A.III.6 paragraphs a) and e) thus seems as an overly broad measure to control problems of predictive policing.

Moreover, use of both 'profiling' and 'assessment' in A.III.6 creates legal uncertainty which should be remedied before the AIA becomes final. As it seems, 'assessment may be more suitable than 'profiling' since the latter so strictly is connected to automated data processing. In both cases, there also seems to be a problem that the provisions are not clear about the relevance of contextual information to the risk assessments. This is a point when the output of the AI system does not have any ostensible relevance to the risk to be assessed. As has been shown, lawful use of person-based AI for LE purposes must involve meaningful human judgement, which should be clearer expressed in the said provisions.

The analysis has further questioned why the restriction on LE use of real-time biometric identification systems in Art. 5(1)(d) does not comprise use of the technology in publicly accessible spaces online. With a reservation for jurisdictional issues relating to LE online, the legal differentiation between physical and online spaces in this respect seems unfounded. Finally, the analysis has demonstrated a possible overlap between paragraph e) first. alt., and paragraph a) and f). In addition, there is a grey zone between paragraph g) (person-based crime analytics) and the other provisions.

It must therefore be concluded that the AIA in its current form raises as series of important issues relevant to the application of its provisions to LEAs. These should be clarified, both for the sake of ensuring that needs for suitably effective LE online

69 See also Sunde and Sunde, *NJSP* (article accepted for publication).

are adequately dealt with, and for ensuring legal certainty. A possibility is to deal with issues special to LEAs in a separate legal instrument, the way it was done with respect to the GDPR and the LED.