EUROPEAN LAW BLOG
NEWS AND COMMENTS ON EU LAW

ELB Blogpost 34/2024, 5 July 2024
Tags: predictive policing, AI Act, meaningful human intervention, human oversight, national security exemption.
Topics: Artificial Intelligence, Criminal Law, Data Protection and Digital Governance

Predictive Policing in the AI Act: meaningful ban or paper tiger?

By Jessie Levano

After years of anticipation, the final text of the Artificial Intelligence Act ('the Act') was approved by the Council on May 21st of this year. The landmark regulation, first of its kind, positions the EU at the forefront of the global effort to establish a comprehensive legal framework on artificial intelligence. The Act aims to safeguard fundamental rights and promoting the development of safe and trustworthy AI by adopting a risk-based approach, mandating stricter scrutiny for higher-risk applications. At the highest level of risk, the Act contains a list of "prohibited uses" of artificial intelligence (Article 5) due to their potentially detrimental consequences for fundamental rights and Union values, including human dignity, freedom, and equality (see Recital 28). While the Act prohibits the use of specific instances of AI predictive policing, we should seriously consider whether the ban will have meaningful effects in practice, or may become a mere instrument of symbolic politics. Leaning towards the latter, this blog cautiously implies that this concern reflects broader questions about the Act's commitment to developing "human-centric" AI and whether it effectively encompasses all individuals within its protective scope.

Predictive policing is not defined in the Act, but a leading definition provided by Perry et. al, is '*the use of analytical techniques to identify promising targets*' to forecast criminal activity. As highlighted by Litska Strikwerda (Dutch only), this may involve identifying potential crime locations (predictive mapping), as well as assessing the likelihood that an individual will either become a victim of a crime or commit a crime (predictive identification). While predictive identification has significant potential as a crime prevention tool, it has faced substantial criticism, particularly concerning potential human rights implications. For example, the extensive data collection and processing involved in

predictive identification raise serious concerns about data protection and privacy, including the correct legal basis for such data processing and the potential intrusion into individuals' private lives. Additionally, the discriminatory nature of algorithms can exacerbate existing structural injustices and biases within the criminal justice system. Another issue is the presumption of innocence, given that predictive identification approaches criminality from an almost entirely opposite perspective, labelling individuals as potential criminals before they have engaged in any criminal conduct. Recital 42 of the Act cites this concern in justifying the prohibition on AI based predictive identification.

Initially classified as a high-risk application of artificial intelligence under the Commission's proposal, predictive identification is now designated as a prohibited use of artificial intelligence under Article 5(1)(d) of the Act. This post seeks to demonstrate the potential limitations of the ban's effectiveness through a critical analysis of this provision. After providing a brief background on the ban, including the substantive lobbying by various human rights organisations after earlier versions of the Act failed to include predictive identification as a prohibited use, the provision and its implications will be analysed in depth. First, this post points out the potential for a "human in the loop" workaround due to the prohibition's reference to "profiling". Secondly, it will discuss how the Act's general exemption clause for national security purposes contributes to a further weakening of the ban's effectiveness.

The Ban in the Act

The practice of predictive identification has been under scrutiny for years before the final adoption of the AI Act. For example, following the experiments of "living labs" in the Netherlands, Amnesty International published an extensive report on the human rights consequences of predictive policing. The report highlights one experiment in particular, namely the "Sensing Project", which involved collecting data about bypassing cars (such as license plate numbers and brands) to predict the occurrence of petty crimes such as pickpocketing and shoplifting. The idea was that certain indicators, such as the type of car, could help identify potential suspects. However, the system disproportionately targeted cars with Eastern European number plates, assigning them a higher risk-score. This bias highlights the potentially discriminatory effects of predictive identification. Earlier that same year (2020), a Dutch lower court ruled that the fraud detection tool *SyRI* violated the right to private life under the ECHR, as it failed to fulfil the "necessary in a democratic society"-condition under Article 8(2) ECHR. This tool, which used "foreign names" and "dual nationality" as possible risk-indicators, was a key element in the notorious child benefits scandal in the Netherlands.

Despite widespread concerns, a ban on predictive policing was not included in the Commission's initial proposal of the Act. Shortly after the publication of the proposal,

several human rights organizations, including Fair Trials, started intensive lobbying for a ban on predictive identification to be included in the Act. Subsequently, the IMCO-LIBE report recommended prohibiting predictive identification under Article 5 of the Act, citing its potential to violate the presumption of innocence, human dignity, and its discriminatory potential. Lobbying efforts continued vigorously throughout the negotiations (see this signed statement of 100+ human rights organizations).

Eventually, the clause was incorporated in the Parliament's resolution and is now part of the final version of the Act, reading as follows:

> [The following AI practices shall be prohibited:] *the placing on the market, the putting into service for this specific purpose, or the use of an AI system(s) for making risk assessments of natural persons in order to assess or predict the likelihood of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics.* […] *This prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity.* (Article 5(1)(d)).

The "Human in the Loop" Problem

The prohibition applies to instances of predictive identification based solely on profiling, or on the assessment of a natural person's personality traits and/or characteristics. The specifics of these terms are unclear. For the definition of "profiling", the Act (Article 3(52)) refers to the definition given in the GDPR, which defines it as any automated processing of personal data to evaluate personal aspects relating to a natural person (Article 4(4) GDPR).

The first question that arises here relates to the difference between profiling and the assessment of personality traits and characteristics. Inger Marie Sunde has highlighted this ambiguity, noting that profiling inherently involves evaluating personal characteristics. A difference between "profiling" and "assessing" may lie in the degree of human involvement. While profiling implies an (almost) entirely automated process with no meaningful human intervention, there is no clear indication on the level of human involvement required for "assessing".

A deeper concern lies in the question as to what should be understood by "automated processing". The test for a decision to qualify as solely-automated, including profiling, is that there has been no meaningful human intervention in the decision-making process. However, the exact meaning of "meaningful" here has not been spelled out. For example, the CJEU in the SCHUFA Holding case confirmed automated credit scoring to be a solely automated decision (in the context of Article 22 GDPR), but did not elaborate on the

details. While it is clear that the human role should be active and real, not symbolic and marginal (e.g. pressing a button), a large grey area remains (for more, see also here). In the context of predictive identification, this creates uncertainty as to the extent of the human involvement required, opening the door for a potential "human in the loop"-defense. Law enforcement authorities could potentially circumvent the ban on predictive identification by demonstrating "meaningful" human involvement in the decision-making process. This problem is further aggravated by the lack of a clear threshold for the definition of "meaningful" in this context.

The second paragraph of the prohibition on predictive identification in the Act states that the prohibition does not apply to AI systems supporting human assessment of criminal involvement, provided this is based on "objective and verifiable facts directly linked to a criminal activity". This could be understood as an instance of predictive identification where the human involvement is sufficiently "meaningful". Nevertheless, there is room for improvement in terms of clarity. Additionally, this conception of predictive identification does not reflect its default operational mode – where AI generates predictions first, followed by human review or verification – but rather the opposite scenario.

In the event that an instance of predictive identification does not fit the definition of a prohibited use, this does not result in the entire practice being effectively free from restrictions. Other instances of predictive identification, not involving profiling or the assessment of an individual's personality traits, may be classified as "high-risk" applications under the Act (See Article 6 in conjunction with Annex III 6(d)). This distinction between prohibited and high-risk practices may hinge on whether the AI system operates solely automatically, or includes meaningful human input. If the threshold for meaningful human intervention is not clearly defined, there is a risk that predictive identification systems with a degree of human involvement just beyond being "marginal and symbolic" might be classified as high-risk rather than prohibited. This is significant, as high-risk systems are simply subject to certain strict safety and transparency rules, rather than being outright prohibited.

In this regard, another issue that should be considered is the requirement of human-oversight. According to Article 14 of the Act, high-risk applications of AI should be subject to "human-oversight" to guarantee their safe use, ensuring that such systems are used responsibly and ethically. However, as is the case with the requirement of "meaningful human intervention", the exact meaning of "human oversight" is also unclear (as explained thoroughly in an article by Johann Laux). As a consequence, even in instances where predictive identification does not classify as a prohibited use under Article 5(1)(d) of the Act, but is considered high-risk instead, uncertainty about the degree of human involvement required remains.

Finally, it should be noted that even if the AI would only have a complementary task compared to the human, another problem exists. It pertains to the potential biases of the actual "human in the loop". Recent studies suggest humans are more likely to agree with AI outcomes that align with their personal predispositions. This is a problem distinct from the inherent biases present in predictive identification systems (as demonstrated by, for example, the aforementioned cases of the "Sensing Project" and the Dutch childcare benefits scandal). Indeed, even the human in the loop "safeguard" may not offer requisite counter-balance to the use of predictive identification systems.

General clause on national security purposes

Further, the Act includes a general exemption for AI systems used for national security purposes. As national security is beyond the EU's competences (Article 4(2) TEU), the Act does not apply to potential uses of AI in the context of the national security of the Member States (Article 2 of the Act). It is uncertain to what extent this exception may influence the ban on predictive identification. National security purposes are not uniformly understood, although established case law has confirmed several instances, such as espionage and (incitement to- and approval of) terrorism to be included within its meaning (see this report by the FRA). Yet, given the degree of discretion granted to the Member States in this area, it is uncertain which instances of predictive identification might be excluded from the Act's application.

Several NGOs focusing on human rights (particularly in the digital realm) have raised concerns about this potential loophole, arguing that the exemption under the Act is broader than permitted under European law. Article 19, an advocacy group for freedom of speech and information, has argued that such a broad exemption contradicts European law, stating that '*the adopted text makes the national security a largely digital rights-free zone*'. Similar concerns have been raised by Access Now. The fear is that Member States might invoke the national security exemption to justify the use of predictive identification techniques under the guise of safeguarding national security. This could undermine the effectiveness of the ban in practice, allowing for the continued use of such technologies despite their potential to infringe upon fundamental rights. For example, the use of predictive policing in counter-terrorism efforts could disproportionately target minority communities and individuals from non-Western backgrounds. Combined with the existing concerns about biases and the potential for discriminatory outcomes in the context of predictive identification, this is a serious ground for concern.

Rather than a blanket exemption, national security considerations should be addressed on a case-by-case basis. This approach finds support in the case law of the ECJ, including its ruling in *La Quadrature du Net*, where it reiterated that the exemption is not by definition synonymous with the absolute non-applicability of European law.

Conclusion

While at first sight the ban on predictive identification appears like a significant win for fundamental rights, its effectiveness is notably weakened by the potential for a "human in the loop"-defence and the national security exemption. The human in the loop-defence may allow law enforcement authorities to engage in predictive identification if they assert human involvement, and the lack of a clear definition for "meaningful human intervention" limits the provision's impact. Additionally, the exemption for AI systems offering mere assistance to human decision-making still allows for human biases to influence outcomes, and the lack of clarity regarding the standards for "human oversight" for high-risk applications are not promising either. The national security exemption further undermines the ban's effectiveness. Given the broad and ambiguous nature of the exemption, there is significant scope for Member States to invoke this exemption.

Combined, these loopholes risk reducing the ban on predictive policing to a symbolic gesture rather than a substantial protection of fundamental rights. In addition to the well-documented downsides of predictive identification, there is an inherent tension between these limitations in the ban, and the overarching goals of the AI Act, including its commitment to safeguard humanity and develop AI that benefits everyone (see for example Recitals 1 and 27 of the Act). Predictive identification may aim to enhance safety by mitigating the threat of potential crime, but it may very well fail to benefit those already marginalised, for example minority communities and individuals from non-Western backgrounds, who are at higher risk of being unfairly targeted, for example under the guise of counter-terrorism efforts. Addressing these issues requires clearer definitions, stricter guidelines on human involvement, and a nuanced approach to national security exceptions. Without such changes, the current ban on this instance of predictive policing risks becoming merely symbolic: a paper tiger failing to confront the real challenges and potential harms of the use of AI in law enforcement.