Predictive justice in light of the new AI Act proposal

Citation for published version (APA):

DOI:
10.2139/ssrn.4286023

Document status and date:
Published: 14/12/2022

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Predictive justice in light of the new AI Act proposal

Chiara Gallese
17th June 2022

Abstract: In the latest years, there has been an increasing trend for police forces and judicial authorities to employ predictive profiling technologies in criminal justice, posing major risks to fundamental rights of citizens. The new AI Act proposal, by introducing a new legal basis for the processing of special categories of personal data without appropriate safeguards, creates a source of potential issues detrimental to data subjects’ rights and freedoms.

1 Introduction

In the latest years, there has been an increasing trend for police forces and judicial authorities to employ predictive profiling technologies in criminal justice, posing major risks to fundamental rights of citizens [7, 9]. Although there are not Courts in which decisions are solely made by an algorithm, in many countries decisional forecast is found in judicial proceedings (see Fig. 1) and in other cases related to law enforcement.

The regulatory framework is fragmented across the EU as both GDPR and the Law Enforcement Directive (LED) leave substantial space to each Member State to regulate the subject under internal law [11]. However, on the 21st of April 2021 the European Commission published a proposal for a regulation on Artificial Intelligence (AI Act) that will harmonize the legal framework regarding the use of AI devices across Europe.

The new proposal introduces a risk-based approach, dividing AI devices in different categories: no risk, low risk, medium risk, and high risk. According to Article 6, par. 2, and Annex III, point 6, predictive justice algorithms will be classified as high-risk, not only when the assessment is made about a criminal offender, but also when it is aimed at evaluating the risk for potential victims.

The new categorization will introduce a number of measures aimed at decreasing the risks involved in the use of those devices, such as quality assessment and data government procedures [4]. This discipline, however, creates some conflicts with data protection law, since it claims that no new legal grounds for the processing of personal data will be introduced, but at the same time it provides for an additional basis authorizing the processing of special categories of personal data for bias monitoring purposes.
In the context of predictive justice, this may entail new risks for data subjects, particularly vulnerable groups, whose data will be processed without the guarantees of GDPR, since the LED discipline provides for a substantial restriction of data subjects rights [8]. In particular, recent cases have shown how difficult it is to obtain the deletion of personal data archived for law enforcement purposes even when the data subject has been fully rehabilitated.

This article explores the risks of using predicting justice devices based on AI models in light of the new AI Act and its relationship with data protection law. The first section describe famous cases in which tools of Predictive Justice have been used in real life scenarios. The second section explains the legal framework emerging from the novel provisions of the AI Act. The third section explores risks and conflicts with data protection law introduced by the new legal framework.

2 Predictive Justice actual implementation and its risks

When Philip K. Dick published the short story The Minority Report, he certainly could not imagine that only few decades later the technological progress would have been so developed to implement AI systems for Predictive justice in actual courtrooms. Yet, algorithms have been employed in many countries, such as the US, Spain, and China.

Famous cases have cast shadows on how predictive justice models are designed and employed, such as COMPAS [2, 5], a tool designed to predict the likelihood of re-offense in convicted criminals, which was proven to be biased.
towards minorities and women. Many scholars have warned that predictive algorithms may perpetuate systemic biases leading to adverse consequences for society [1].

In the US, the case of COMPAS has become notorious among scholars, but other cases are relevant too, such as the Spanish VioGén [3], aimed at the prediction and handling of recidivism risk in the context of domestic violence[6].

Predictive justice is not the only domain in which automated means have been employed in the judicial field causing damages to citizens. For example in 2015 an error in a software aimed at calculating financial terms in divorces cases caused thousands of inaccurate calculations, leading to settle for unfair terms.

3 The new AI Act proposal and the protection of personal data

As noted by [12], one of the characteristics of the use of AI systems in the judicial field is the intense power exercised by the States towards its citizens. In the context of criminal justice, it is particularly important to take in consideration that citizens’ fundamental rights are at stake, such as freedom, the right to a due process, and the presumption of innocence.

For this reason, the new AI Act proposal classified AI systems employed in the judicial field as "high-risk", and prescribed a number of safeguards to prevent their misuse, such as banning real-time face recognition in public spaces for law enforcement purposes.

According to the new legislation, high-risk systems must abide to strict obligations, such as having:

1. a Risk management system (similar to MDR and IVDR, article 9)
2. Data governance and data management practices (similar to MDR, IVDR, and GDPR, article 10)
3. Technical documentation (similar to MDR and IVDR, article 11)
4. a Logging system (article 12)
5. Transparency and information to users (similar to GDPR, article 13)
6. Human oversight requirements (similar to MDR and IVDR, article 14)
7. Accuracy, robustness and cybersecurity measures (similar to MDR, IVDR, and GDPR, article 15)
8. a Quality management system (similar to MDR and IVDR, articles 16 and 17)

Since its publication in 2021, thousands of amendments have been proposed by the Committee on the Internal Market and Consumer Protection and the
Committee on Civil Liberties, Justice and Home Affairs, such as adding a requirement of performing a human-rights assessment.

Additionally, throughout the whole text and in the recitals, it is specified that the new regulation is implemented without prejudice to GDPR, which, in case of conflicts, should prevail. It is also stated that the proposal should not introduce additional legal basis for the processing of personal data.

However, in the data management section, it is explicitly provided that the processing of special categories of data is permitted for the purpose of bias monitoring. Clearly, there is an incongruity at this regards, which we hope will be scraped from the final text of the regulation, as requested in the amendments.

The provisions regarding data governance fit into the large debate of AI auditing, a discipline that has expanded in recent years. In order to be considered trustworthy, AI systems need to be assessed, tested and audited not only by technical experts, but also by ethicists and data protection legal experts. The auditing is carried out with regards to the whole AI life cycle: data collection, data pre-processing, data processing, training, choice of the model and features selection, development of the model, testing, release into the market, and post-market monitoring.

Bias monitoring in the data sets means carrying out an assessment related to potential discrimination in the data, such as a gender imbalance, over-representation of a certain population, or disregarding a specific characteristic peculiar to certain minorities. The literature regarding bias in the data employed to train algorithms is extensive.

Clearly, to understand if a model will not be able to generalize in case it is employed on non-cisgender people, people with a certain neurodiversity, people of a certain ethnicity, and such, it is necessary to process the information that certain subjects hold that peculiar characteristic (e.g., being transgender, being autistic, being Jewish), that is, processing special categories of personal data (art. 9 GDPR). Without this information, it is impossible to create a balanced data set. This is the reasons behind the novel provisions regarding the processing of such sensitive data.

4 Novel risks and conflicts deriving from the new discipline

The AI Act proposal created the new legal basis for the processing of special categories of data taking in mind GDPR and algorithms developed in the private sector. However, the potential risks related to the handling of sensitive information in the public sector have not been adequately explored.

In particular, the discipline drawn by the Law Enforcement Directive, which applies to competent authorities processing personal data for law enforcement purposes, diverges from that of GDPR in many aspects, above all the provisions related to data subjects’ rights. In fact, according to article 14 and 16, right to access to personal data, and right to rectification or erasure, are limited.
It is important to keep in mind that the LED discipline applies also to police activities carried out prior to the commission of a criminal offense and prior to any conviction, for example the investigation on someone who, at a later stage, will be excluded from the suspects and cleared, that is a completely innocent person, which rights are restricted just because they happened to mistakenly be involved in a criminal investigation. It applies also to other activities, such as public security monitoring at demonstrations, sporting events, or even random checks on citizens.

Although article 6 prescribes to make a clear distinction between personal data of different categories of data subjects, such as victims of a crime, persons convicted of a criminal offense, witnesses, etc., this is not a mandatory element, as it is subject to discretion.

An example is a recent case taking place in Italy. The Court of Nola, acting as a judge of the criminal record, rejected the request made in the interest of a data subject, aimed at obtaining the removal from the criminal record of the criminal conviction decree issued by the Judge for Preliminary Investigation, following the intervening rehabilitation ordered by the Court of Surveillance of Naples in 2019.

In 2020, the Italian Court of Cassation, with the judgment no.35548 of 11st of December 2020, upheld the decision, arguing, among other reasons, that “under the currently applicable national and supranational legislation, the processing of data related to judicial functions is subject to special regulation. It is permitted - under the current GPDR (EU Regulation No. 2016/679) - to process them also in relation to sensitive data (Article 9) par. 2(f), the right to erasure does not apply to them (Article 17(3)), the law of the Union and individual states may provide for specific limitations to the rights of data subjects for the purposes of justice, and the national sectoral authorities are not competent to review them (Article 55(3)). Regarding the rights of data subjects, Article 2-duodecies of the Privacy Code, introduced by Article 2 of Legislative Decree No. 101 of 2018, regulates, in application of Article 23(1)(f) of Regulation (EU) 2016/679, the limitations of the rights of data subjects set forth in Articles 12 to 22 and 34, for the needs of safeguarding the independence of the judiciary and judicial proceedings”.

It also added that “the disclosure of data by means of the criminal record certificate requested by the person concerned is then excluded outright since convictions for which there has been rehabilitation do not appear in it. Similar limitation concerns the certificate requested, in the cases specifically provided for, by the employer. Public administrations and managers of public services may, on the other hand, request a general certification from which, in hypothetical cases, convictions for which there has been rehabilitation may also appear, but this is a request justified by the needs related to the exercise of their functions, so that it appears justified and proportionate in light of the tasks performed by these entities”.

A fully rehabilitated person, then, is not permitted to erase their past and start a new life, as their convictions will be kept by the public bodies. However, this is not the most concerning case, but just an example of how, even when
taking legal action, it is extremely difficult to obtain the erasure of sensitive
data held by the public sector under the current LED discipline.

With the introduction of the new discipline, if the text is not changed, vulnerable
groups will be at risks of being abused and exposed by the law enforcement
authorities, which are not known for being the most sensitive to instances of
minorities, especially in the most conservative countries.

There is a real risk that, even when performing random checks on the pop-
ulation, when receiving reports of domestic abuse from hospitals, or when iden-
tifying participants of a public demonstration (e.g., trade unions strikes, G20,
LGBT Pride), authorities might collect the most sensitive information and store
them indefinitely on their servers, without granting data subjects the possibility
of changing, updating, or deleting it.

5 Predictive justice and special categories of data

Although bias monitoring is a noble purpose, it is not balanced with citizens’
rights and freedom, and appropriate safeguards are not provided at the EU
level. In addition, with regards to biometric data, the recent Clearview AI
case[10] has shown how law enforcement authorities have misused AI systems
notwithstanding the limitations provided by law.

In this context, AI systems employed for predictive justice purposes could
be even greatly misused if a large number of citizens’ art. 9 data are collected,
stored, and used by public authorities.

Firstly, even from an ethical point of view, it cannot be considered accept-
able to employ such data to train predictive algorithms without data subjects’
consent, while their rights are heavily limited, and by just giving them a mere
notice (as requested by LED).

Secondly, the discipline creates a discrimination between people who were
unlucky enough to come across police, even if completely innocent and without
fault, and people whose data were never collected.

Thirdly, as shown in section 2, both the building of predictive algorithms
and their actual implementation have been the source of discrimination and
violation of citizens’ right and freedom, therefore the provisions of the AI Act
only increase the risk of incurring in such problems, without really providing
for safeguards and mitigation measures.

6 Conclusion

In this paper we explored the consequences of introducing a new discipline with-
out taking into account risks to data subjects’ rights, in particular for vulnerable
groups, and conflicts with data protection discipline.

The combined provisions of the LED directive and the AI Act proposal
creates relevant issues that will introduce relevant problems for citizens.
We advocate for a change to the current text of the proposal, as proposed by Brando Benifei’s Draft report.

References


