PANOPTYKON FOUNDATION

BEYOND DATA PROTECTION LAWS:
THE AI ACT MUST PROTECT PEOPLE FROM ALL HARMFUL USES OF AI SYSTEMS[1]

The AI Act, as proposed by the Commission, predominantly focuses on requirements for providers of AI systems and does not sufficiently address users (deployers) of AI. In the explanatory statement,[2] the Commission argues that focusing on the development stage of AI systems is a deliberate policy choice motivated by the belief that existing legislation, in particular data protection laws, sufficiently addresses the deployment stage. It is not disputed that the EU data protection framework – the General Data Protection Regulation (GDPR) and the Law Enforcement Directive (LED) – fully applies to the processing of personal data in the context of AI. However, the data protection framework alone is insufficient to address all scenarios related to the use of AI systems.

This paper explains why an additional layer of protection in the AI Act is needed to guarantee effective protection of fundamental rights and argues that the AI Act should:

1. Extend protection to AI systems with real-life impact whose use will not fall under data protection laws.
2. Complement data protection laws in the context of machine learning.
3. Facilitate public scrutiny over broader societal impacts of AI.

***

1. The AI Act should extend protection to AI systems with real-life impact whose use will not fall under data protection laws

Some AI systems will **escape requirements stemming from data protection laws** because either they do not process personal data, or process data about people other than those that are impacted by the system's operation. This will notably be the case for AI systems which do not make predictions and assessments about individuals but produce outputs in relation to trends, future events, physical objects or groups of people. Even when the outputs of such systems are not used for individual decision-making, their outcomes can have real-life impact on rights and freedoms of individuals or entire groups.

> ***Example 1: ROBORDER[3].*** *An autonomous border surveillance system will be able to detect boats moving offshore, making it possible for border authorities to prevent people on the boat from seeking asylum, regardless of their personal circumstances. Even though individual decisions are not made about them by the system, its use impacts their right to apply for international protection and presents a severe risk to the right to non-discrimination.*

[1] This paper was drafted with contributions from European Digital Rights (EDRi), Algorithm Watch, European Center for Not-for-Profit Law (ECNL), European Disability Forum, Fair Trials, and Prof. Douwe Korff.
[2] https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206
[3] https://roborder.eu/the-project/aims-objectives/

Example 2: Crime Anticipation System[4]. *A location-based predictive policing system in Amsterdam identifies areas at a high risk of occurrence of crime based on statistical demographic and income data and historical crime data. One of the indicators of high-risk is the number of non-Western people, arbitrarily defined to exclude Japanese and Indonesian people as well as all European nationalities except for Turkish individuals. Although processing mostly anonymised data, the system presents risks to the right to non-discrimination by presupposing a correlation between certain ethnicities and crime. Over-surveillance of certain areas paired with underdetection of crime in others can further embed existing bias in policing data to the detriment of marginalised groups.*

The impact and use of these systems will be not be possible to comprehensively challenge under data protection laws as they do not process data of people who are affected but rather rely on observations of objects or other people. As a result, users (deployers) of systems which may produce adversarial and discriminatory effects for entire groups of people, and which will have clear impacts also on individual members of these groups, could evade the obligation to assess the consequences of the system in a data protection impact assessment (DPIA) or to inform affected people about its operation.

If the goal of the AI Act is to prevent abuses of fundamental rights and incentivise the development of trustworthy AI, it must ensure that <u>all</u> AI systems which have real-life impact on people are used in a transparent and accountable way, regardless of whether or not their use falls under the GDPR or the LED.

2. The AI Act should complement data protection laws in the context of machine learning

In principle, all GDPR and LED requirements and rights apply to the processing of personal data in machine learning systems. But in practice, both controllers (organisations using AI) and data subjects struggle to apply GDPR concepts to AI reality. Key safeguards and rights (such as the right of access to data, informed consent, increased protection of sensitive data) often do not stand a chance to become reality, in particular due to the technical nature of machine learning systems and how they produce inferences.

2.1. <u>Sensitive inferences are not protected</u>

The very purpose of ML systems is to look for hidden correlations in thousands of data points and generate inferences on that basis. Even though recent case law and EDPB guidance leaves little doubt that inferences are indeed personal data[5], enforcing the GDPR in relation to correlations and inferences is very difficult in practice because **ML systems do not label them in human-understandable categories**. As a result:

- data controllers cannot verify if inferences reveal sensitive data and if they should fulfill their obligations under Article 9 of the GDPR,

[4] S. Oosterloo, G. van Schie, *The Politics and Biases of the "Crime Anticipation System" of the Dutch Police, http://ceur-ws.org/Vol-2103/paper_6.pdf*

[5] See EDPB Guidelines 8/2020 on targeting social media users or the decision of the Polish DPA which confirmed that a marketing profile created on the basis of cookies constitutes personal data: https://gdprhub.eu/index.php?title=UODO_(Poland)_-_ZSPR.440.331.2019.PR.PAM

- data subjects who do not have access to the "inside" of the system find it virtually impossible to challenge unlawful processing of sensitive data.

The sensitivity or the discriminatory effect of processing in the context of AI will usually arise *not* from the types of input data used but from inferences. This is because sensitive data or otherwise protected personal characteristics can be deduced from seemingly innocuous information. The American philosopher and computer scientist Brian Christian illustrates this in his book "The Alignment Problem: Machine Learning and Human Values" with the following example:

> *To the extent that, say, men and women tend toward different writing styles in general—subtle differences in diction or syntax—word2vec will find a host of minute and indirect correlations between software engineer and all phrasing typical of males. It might be as noticeable as football listed among their hobbies, rather than softball, as subtle as the names of certain universities or hometowns, or as nearly invisible as a slight grammatical preference for one preposition or synonym over another.*

Machine learning systems usually produce inferences through **proxies whose sensitive nature is *experienced* by the affected person but *invisible* to the algorithm, unless a human manually labels them.** Some organisations that use AI rely on this supposed ignorance to complicate the exercise of data subjects' rights.

> ***Example:*** *The EU Passenger Name Record (PNR) agreements with third countries caution against the use of the processing of sensitive data but at the same time most of the EU PNR agreements include a "free text" category where airlines could enter any types of information, sensitive or not. There is no predictability for passengers as to which information may be processed and how authorities may be using it.*

What's more, **it is often virtually impossible to isolate and name the specific correlation which reveals sensitive data**. First, algorithms often rely on a multitude of signals which might amount to a sensitive characteristic only when taken together (e.g. the analysis of a large number of various types of small activities in the use of Twitter, rather than a specific behaviour, makes it possible to infer whether a user is blind[6]). Second, as ML systems constantly learn and evolve, correlations they produce change dynamically which makes it difficult to capture and investigate them. And finally, while some information has already been known to serve as proxies (e.g. income can act as proxy for gender), data-rich AI systems are likely to produce new and counterintuitive proxies that will be much more difficult to detect[7].

2.2. Data protection laws do not adequately protect people from consequences stemming from the use of other people's personal data

In the era of big data, decisions about individuals are never based only on their personal data (which is covered by the GDPR), but also on data about other people which is used to define the parameters and the logic of the system. The GDPR does not offer a clear answer to the dilemma of how an impacted individual could challenge an automated decision when the harm stems not from

[6] M. Morris et al., *Understanding Twitter's Evolving Accessibility to Blind Users*, https://dl.acm.org/doi/10.1145/2858036.2858116

[7] S. Wachter, B. Mittelstadt, C. Russel, *Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3547922

the incorrect interpretation of their personal data, but from unfair optimisation logic defined by the user (deployer) based on the behaviour of other people.

> ***Example:*** *Amazon uses an AI system to monitor and assess the productivity of warehouse employees[8]. If they do not achieve the minimum threshold, the system produces recommendations for managers to reprimand or dismiss them. The productivity threshold against which all employees are assessed is increased every month to match efficiency rates of the fastest working employees[9], leading to physical exhaustion and increased risk of dismissal, even when the employee was meeting productivity targets in the past.*

2.3. Article 22 of the GDPR fails to truly protect people from consequences of decisions supported by AI systems

The AI Act does not include any specific transparency obligations for people affected by AI, nor the right to explanation of decisions taken by or with assistance of an AI system. However, the GDPR is insufficient to help people understand and challenge such decisions because:

- an enforceable individual right to receive information about the logic of the system only exists for solely automated decisions defined in Article 22 of the GDPR as decisions made without meaningful human involvement[10],
- the GDPR does not create a fully-fledged right to explanation which would describe how the general logic of the system was applied in the individual case[11].

It is **very likely that no high-risk AI system will fall under the scope of Article 22 of the GDPR** because such systems will usually involve a human (especially when read together with the obligation of human oversight envisioned in Article 14 of the AI Act). As a result, people affected by high-risk AI systems will not have the right to find out about the logic of the system, nor the right to explanation how this logic applied in their particular case.

> ***Example[12]:*** *Parents of a small child receive a rejection letter from the kindergarten which uses AI to allocate limited places. Recommendations produced by the system are reviewed by an employee. In this case, the parents do not have the right to information about the general logic of the system (e.g. indicating that priority was given to older children, parents who had been on the waiting list for a longer time, who are employed, and who have a low income), nor a specific right to explanation why their application was rejected in the context of these criteria.*

Admittedly, to prevent circumventing the law, the EDPB clarified that "meaningful" human involvement means that the human reviewer does not merely rubber stamp the system's

---

[8] https://www.washingtonpost.com/technology/2021/12/02/amazon-workplace-monitoring-unions/

[9] https://krytykapolityczna.pl/swiat/ue/zarzadzanie-algorytmiczne-w-magazynach-amazona-w-europie-srodkowej/

[10] Art. 13(2)(f) of the GDPR.

[11] S. Wachter, B. Mittelstadt, L. Floridi, *Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation*, https://academic.oup.com/idpl/article/7/2/76/3860948.

[12] This example is inspired by the scenario described by Prof. Christiane Wenderhost in her analysis of the AI Act for the Austrian Federal Ministry of Social Affairs, Health, Care and Consumer Protection, *The Proposal for an Artificial Intelligence Act COM(2021) 206 from a Consumer Policy Perspective*, https://www.sozialministerium.at/dam/jcr:750b1a99-c5af-47bd-906a-7aa2485dabbd/The%20Proposal%20for%20an%20Artificial%20Intelligence%20Act%20COM2021%20206%20from%20a%20Consumer%20Policy%20Perspective_dec2021__pdfUA.pdf

decisions but is capable to critically assess and reverse the system's recommendations if needed. However, the extent to which this is possible for complex AI systems is questionable due to well-researched problems of automation bias or the lack of ability of human operators to interpret and assess the system's recommendations. In addition, from the perspective of a data subject, who does not have access to the "backstage" of decision-making, proving that the role of the human reviewer was not meaningful is practically impossible.

Similarly, the notion of "solely automated decision" is unclear in the light of decision-making processes with multiple stages (multi-stage profiling), some of which might not be automated. For example, if a credit decision is made by a human analyst, but on the basis of automated credit scoring[13].

2.4. Data protection rules on biometric data do not sufficiently protect people from AI harms

The EU legal regime related to the protection of biometric data, including the GDPR and the LED, is **not clear enough to prevent unacceptable surveillance, discrimination and manipulation** that arise in the specific context of the use of AI systems.

First, even though the GDPR and the LED treat biometric data as sensitive and their processing in-principle prohibited, certain conditions allow the use of these data on the basis of Union or Member State law. Article 9(2)(g) of the GDPR sets basic standards that these laws must meet (e.g. the use of biometric data must serve a substantial public interest, be necessary and proportionate for this aim, respect the essence of the right to data protection and provide safeguards for fundamental rights). But inevitably, details related to when such processing is permitted differ depending on the country.

Second, the GDPR definition of biometric data also creates a grey area when the "purpose" of processing is not to uniquely identify a data subject, or when physical, physiological or behavioural data are used to profile people without uniquely identifying them.

> ***Example:*** *Netherlands-based company VisionLabs sells a system[14] which profiles the age, gender and emotions of shoppers in order to influence their purchasing behaviour, and argues that their systems "fully comply to personal data protection regulations for practically any country in the world" because they process face attributes rather than facial images.*

Third, even when the data in question are unambiguously biometric data used for the purpose of unique identification, the basis of consent has been systematically misused[15] as an exception, and other exceptions create serious margin for abuse as well.

Last but not least, whilst it has been argued that the AI Act should not further regulate biometrics in order to avoid overlapping with the GDPR or the LED, it is noteworthy that the GDPR in fact anticipates further restrictions on the processing of biometric data in Art. 9(4). The AI Act is the best way to achieve this without risking fragmentation. Such guidance could also help to avoid the problems that arise from the use of biometrics being regulated differently in the laws of the Member States - something that both seriously hampers the deployment of GDPR-compliant

[13] This very issue has recently been brought before the CJEU in Schufa II (C-634/21).
[14] https://visionlabs.ai/industries/retail
[15] https://edri.org/our-work/blog-ban-biometric-mass-surveillance/

biometric systems across borders within the Single Market and undermines the protection of data subjects in that regard. Only properly harmonised rules on biometrics – which can be implemented via the AI Act in order to enhance the protections of the GDPR - set at a sufficient level to properly protect our rights, can ensure safe deployment of such systems.

2.5. Data protection laws fail to protect people deprived of legal capacity from unwanted processing of sensitive data

Article 9(2)(c) of the GDPR makes it possible to process sensitive data, e.g. data about health, when "processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent". Given the prevalence of discriminatory and outdated guardianship laws which persist throughout Europe, many persons with disabilities, particularly those with intellectual and psychosocial disabilities, are deprived of legal capacity. This means they will not be protected against processing of their personal data by AI systems if this gap in the GDPR is not closed in the AI Act. The AI Act must therefore ensure that privacy and data protection of all persons, including those under substituted decision-making regimes such as guardianships, are protected when their data are processed by AI systems.

3. The AI Act should facilitate public scrutiny over broader societal impacts of AI

Finally, the reason why the AI Act should create an accountability framework not only for providers but also for users (deployers) of AI systems is that **existing regulations alone, in particular data protection laws, cannot achieve one of the key objectives of the AI Act**, namely to "ensure that AI systems in the EU are used in a way which respects fundamental rights and Union values". This is because the impact of AI systems spans much further than data or consumer protection to broader fundamental rights, socio-economic and political issues (such as more acute consequences for marginalised groups, the potential facilitation of structural inequality and over-surveillance, the militarisation of borders or expending of resources), and environmental concerns.

The ambitions of the AI Act to consider these broader impacts are reflected by the proposal's aim to prohibit certain uses of AI as incompatible with fundamental rights and Union values. But as long as a system does not fall into the prohibited category, it is not clear how these consequences are to be assessed. If the goal of the European legislators is really to build trust in AI and promote responsible and rights-respecting use of this technology in the EU, the AI Act should require organisations considering deployment of AI to analyse these broader impacts[16], especially when AI is used to pursue public policy goals, as well as create the conditions for effective public scrutiny over such uses.

***

[16]A similar recommendation was put forth by the Council of Europe Ad Hoc Committee on Artificial Intelligence in December 2021 according to which states should introduce a mechanism for assessing the impact of AI systems on the enjoyment of human rights, the functioning of democracy, and the observance of the rule of law: https://rm.coe.int/possible-elements-of-a-legal-framework-on-artificial-intelligence/1680a5ae6b.

In line with our amendments, the AI Act should therefore:

- require users of high-risk AI systems to examine the envisioned impact of the AI system on fundamental rights, equality, accessibility, public interest and the environment – this assessment should be performed prior to deployment and at regular intervals afterwards;
- ensure that the results of the impact assessment are **publicly available** and accessible for everyone, in order to facilitate public scrutiny and contribute to trustworthy use of AI in the EU;
- confer individual transparency and redress rights to people affected by AI systems, including the right to receive information about the system, the right to explanation of its outcome (even when a human is involved) and the possibility to file a complaint;
- ensure the participation of public interest organisations in the investigation and enforcement process by making it possible for them **to trigger an investigation** into the system in the case of identified violations of fundamental rights.