Ensure meaningful transparency of AI systems for affected people

This paper outlines amendments to the European's Commission proposal for the Artificial Intelligence Act (AIA) with respect to transparency obligations for AI systems vis-a-vis people affected by them.

What AI systems are currently subject to transparency obligations?

Currently, Article 52 of the AIA is the only provision which introduces basic transparency of AI systems vis-a-vis people affected by them, stipulating that in certain cases people should be informed about the fact that an AI system is in use. In the Commission's proposal this requirement applies only to a few systems, unless they are used to detect, prevent, investigate and prosecute criminal offences, namely:
- AI systems intended to interact with natural persons;
- emotion recognition systems;
- biometric categorisation systems;
- AI systems that generate 'deep fakes' (images, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful).

It should be noted that currently there are **no transparency obligations vis-a-vis persons subject to high-risk AI systems or systems that haven't been categorized as high-risk but nevertheless affect individuals and pose a risk for them,** such as AI systems used to assess or evaluate people by businesses or public authorities, unless they also belong to the limited catalogue above. Another shortcoming of the proposed transparency obligation is that it is limited to the duty to inform the individual **about the mere fact that an AI system is in use**, without the obligation to provide any other information, e.g. on the purpose or functioning of the system, or even the identity of the user.

While Article 52 in its current form may lay the ground for a right to a non-individualised notification about the use of some AI systems, **the AIA does not currently envision the possibility for people subject to AI-supported decisions to obtain substantive information tailored to their specific situation**. In particular, under the Commission's proposal, people significantly impacted by such decisions will not be able to ask for an explanation of how and why the AI system produced a specific outcome in relation to them and – if a human was involved in the decision-making process - to what extent the algorithmic outcome influenced the decision. Such an explanation is a precondition for individuals' ability to meaningfully challenge the outcome of the system and effectively defend themselves from violations of their rights[1].

Why do we need meaningful transparency of AI systems for individuals and how to achieve it?

1. Notification of the use of an AI system

---

[1]   See a separate paper with recommendations related to ensuring rights and redress for people impacted by AI systems.

To ensure a high level of protection of fundamental rights, the law should make it easy for people to find out if they are subject to, or impacted by, an AI system. Knowing that an AI system is used and for what purpose is a precondition to identify and report a violation of fundamental rights, such as discriminatory or unfair treatment, and to exercise one's rights. When companies or institutions using AI are transparent about it and do not operate in the dark, it contributes not only to increasing their accountability for the AI systems they deploy, but also to creating an ecosystem of trust and excellence around AI, which is one of the key goals of the proposed regulation.

Article 52 currently only covers situations where the nature of an AI system poses a risk of manipulation or deception, but notification about the use of AI systems is even more crucial when the use of an AI system may significantly impact someone's life, legal situation, or social status, which is the case for all high-risk AI systems included in Annex III. The opportunity to find out that an AI system is in operation and what it does is also essential from the perspective of consumer protection, where consumers' individual traits are assessed or evaluated for the purposes of offering them services or goods or determining the conditions of access to such goods or services.

In this context, the list of systems subject to transparency obligations under Article 52 is unjustifiably limited. All the more so when we consider that:
- the AIA in its current form does not envision a parallel obligation relating to high-risk AI systems. In the context of protection of fundamental rights it is not clear why people should be informed about the use of a chatbot, even when it poses a relatively lower risk, but not about an AI system which participates in assessing their CV or their benefits request[2];
- GDPR information requirements related to profiling are insufficient in this context because, first of all, they apply only to situations where personal data as defined by the GDPR is processed (while an AI system might rely on big data or non-personal variables) and second of all, specific obligations related to providing meaningful information about the functioning, logic and consequences of an AI system are limited to situations which fall under the narrow scope of Article 22 GDPR (**solely automated decisions** which produce legal or otherwise significant effects for the data subject), thus excluding situations where impact on the individual's life is still significant but the decision is taken with the assistance of - and not solely by - an AI system;
- certain AI systems which assess or evaluate people's individual traits and influence their access to goods and services, even when such systems are not designated as high-risk under the AIA (e.g. price determination systems, systems which rank offers in online shops or recommend personalised diets based on an individual's health condition), pose an inherent risk of manipulation, consumer harm, discrimination or even a threat to health and safety, and as such people should be aware that such systems are used on them.

Therefore, we propose to expand the list of AI systems subject to transparency obligations to also **include all high-risk AI systems as well as AI systems which assess and evaluate people, predict their behaviour, interests or personal**

---

[2]   Please note that one of the supporters of these recommendations, the European Disability Forum, advocated for a prohibition on these systems.

This paper was drafted by and with the support of Panoptykon Foundation, European Digital Rights (EDRi), Access Now, Algorithm Watch, European Disability Forum (EDF), and European Not for Profit Law Center. It follows the Joint Civil Society Statement 'An EU Artificial Intelligence Act for Fundamental Rights' signed by 123 organisations in November 2021.

traits, recommend information, goods or services to them based on their activity or personal traits, regardless of whether these systems have been classified as high-risk. The AIA should also envision an update mechanism for Article 52 in order to ensure a swift response to technological developments and emerging challenges for fundamental rights[3].

To be useful for individuals affected by an AI system, the notice provided to them should include not just the mere information that an AI system is in use (which does not offer much insight), but also:

- concise, easily understandable, accessible for persons with disabilities information about the purpose or task of the system (what it does);
- indication where more information about the system can be found (e.g. a link to EU public database, other publicly available and accessible for persons with disabilities resource, or contact details of the user);
- the information about the right to request an explanation if the decision taken with the use of an AI system significantly impacted them (see point 2 below).

It should also be noted that due to the severe risks that their use poses to fundamental rights, emotion recognition systems and some uses of biometric categorisation systems should be elevated to 'prohibited AI' under Article 5[4].

Finally, any exceptions to the obligation to inform natural persons about the use of an AI system for the purposes of detection, prevention, investigation and prosecution of criminal offences should not happen by default and should be limited to specific cases where it is strictly necessary in order to avoid obstructing or prejudicing proceedings. Please also note that in a separate paper a number of civil society organisations is advocating for the prohibition under Article 5 of uses of AI systems in the context of predictive policing.

2. The right to request an explanation of significant decisions

As mentioned above, the AIA does not give individuals the right to inquire why a certain decision was made about them with the assistance of an AI system. While some might say that such a right exists under Article 22 of the GDPR, this is questioned by distinguished legal scholars[5] who argue that, firstly, Article 22 provides other remedies, such as the right to human intervention, but not a right to explanation, and secondly, even if this right existed it would be limited to situations where decisions are **solely** automated. This seriously limits the practical application of this right as many decisions which significantly impact individuals are not made "solely" by an AI system, but rather by AI systems that assist the human in some stages of the decision-making process[6].

As a result, individuals have no legal tools to obtain an explanation as to why a certain outcome was produced by the AI system and how it influenced the final decision in

---

[3]   See the ancillary amendment proposing a new Article 52a below.
[4]   See a separate paper with recommendations on biometric categorisation and emotion recognition for more information about this.
[5]   See: S. Wachter et al., *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.
[6]   See: R. Binns, M. Veale, *Is This Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR*, https://academic.oup.com/idpl/article/11/4/319/6403925.

their case. **Lack of access to a meaningful explanation of reasons behind AI-assisted decisions makes it difficult, if not impossible, to contest false, inaccurate or discriminatory outcomes[7].** This is especially important if the decision is made by a public authority, because under EU law the administration has the duty to give reasons for its decisions. The AIA should explicitly adapt this duty to the AI context[8].

For these reasons, we strongly recommend that **the AIA fill this gap and create a possibility to request an explanation of outcomes of an AI system generated in relation to them in cases where the AI-assisted decision produced legal effects or otherwise significantly impacted the individual**. The interpretation of "significant impact" could be inspired by existing jurisprudence related to Article 22 of the GDPR and the EDPB guidelines WP251 on automated decision-making and profiling, which mention, for instance, decisions which affect an individual's financial circumstances (incl. differential pricing), access to education or employment opportunities, or even some forms of micro-targeting in online advertising[9].

---

[7]   In a separate paper we recommend the introduction of individual and collective redress mechanisms in the AIA, including the right to lodge a complaint with a supervisory authority.

[8]   See: M. Fink, *The EU Artificial Intelligence Act and Access to Justice*, https://eulawlive.com/op-ed-the-eu-artificial-intelligence-act-and-access-to-justice-by-melanie-fink/.

[9]   https://ec.europa.eu/newsroom/article29/items/612053