Universidad del País Vasco Euskal Herriko Unibertsitatea

PANELFIT
PARTICIPATORY APPROACHES TO A NEW ETHICAL AND LEGAL FRAMEWORK FOR ICT

Feedback

on

Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence

José Antonio Castillo Parrilla

Investigador Posdoctoral UPV – Cátedra Derecho y Genoma Humano UPV

Member of the PANELFIT PROJECT, funded by the EC, GA number: No: 788039

Context of the comment

As highlighted in the proposal, AI is a rapidly evolving technology whose use can contribute to economic development in several fields such as health, education or transport. Equally, AI evolution cannot be separated from the legal situation of data processing[1] since AI tools use big amounts of data to obtain useful knowledge.

Problems to be tackled and options

The starting point of the proposal seems correct: the development of AI (as has happened with other innovative technologies) needs a clear legal and ethical framework in order to guarantee legal certainty[2]. Legal certainty must be achieved, at least, at the level of EU law since this is a global legal challenge (EU law principle of subsidiarity).

Protection against damages is highlighted as one of the aims of future regulatory development: (1) material harm (such as damages to health or individuals´ properties because of autonomous vehicles or AI-driven tools); (2) and immaterial harm (such as loss of privacy, limitations to the right of freedom of expression, and discriminatory situations).

Most of these damages, as stated correctly in the text, will not necessarily be linked to AI, but can be caused by poor quality data (either unreliable or poorly collected data), which confirms the importance of data processing techniques for the development of AI as well as their applicable law.

The proposal presents four levels of intervention: 0 (no intervention), 1 (EU soft law), 2 (EU law on voluntary labelling scheme), 3 (compulsory law, either limited to specific areas – 3.1 and 3.2 – or general – 3.3-), 4 (a combination of options 1, 2 and 3).

---

[1] Vid. European Parliamentary Research Service, The impact of the General Data Protection Regulation (GDPR) on artificial intelligence, Junio de 2020. Disponible en: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.

[2] Vid. "Personal autonomy and the digital revolution", in DE FRANCESCHI, A., SCHULZE, R., Digital Revolution – New Challenges for Law, C.H. Beck & Nomos, München, 2019, p. 17: "A robust

Feedback on the different levels of intervention

- **General comment**

Option 4 seems the most advisable one: compulsory law on specific sectors (option 3.2) combined with rules on a voluntary labelling scheme and soft law for low-risk areas or areas where political agreements have not been reached.

Political agreements are key to develop rules to provide with legal certainty on AI. To get them, industry and civil society (and the academia) must collaborate as soon as possible to develop drafts to be submitted for consideration by the EU institutions.

- **On option 0, explanatory notes and authentic interpretations**

No legal intervention (option 0) would not solve legal uncertainty problems, but it would even increase them according to the proposal. Nonetheless, some problems related to AI do not need a complete regulation nor even legal modifications: some of the problems related to the development or use of AI only need to adapt the interpretation of the current law to the new situation. This goal can be achieved through Guidelines or explanatory notes as, for instance, those issued by the EDPB on data protection to favour a common and updated interpretation of the GDPR as well as solve specific problems.

For instance, an explanatory note by the European Commission on the black box phenomenon and EDPB Guidelines on accountability applied to AI tools would contribute to diminish the problem of algorithms opacity regardless of the needed future legal development to get major legal certainty (see below).

These documents can be doubly useful: they can help to mitigate a problem while a new legal development is being discussed or eliminate the need for a legal development.

- **On soft law (option 1) and voluntary labelling scheme (option 2)**

Options 1 and 2 are a great complement for the necessary binding law on certain areas (option 3.2). What areas should use soft law or voluntary labelling schemes? The criteria, from our point of view, should be: (1) secondary issues of areas regulated by binding rules, and (2) areas where political agreements have not been possible or sufficiently strong to develop binding rules.

- **On the possibilities for binding rules (option 3 and sub-options)**

Option 3 offers three sub-options. It does not seem reasonable reducing legal development to a single area of AI, nor trying to develop a generally comprehensive legal development or single legal developments for each area of AI (sub-options 1 and 3). The best choice is to identify those areas that represent higher risk for rights damages (option 2).

Option 3.2 needs, however, some considerations. Previously it would be advisable to develop a double analysis on the potential of option 0: first analysis, to identify what AI problems could be solved or diminished by explanatory notes or guidelines issued by

EU institutions as authorized interpretations of the current law; and second analysis to identify minor legal changes where needed.

This analysis would allow us to:

- (1) focus efforts on genuinely necessary regulatory developments,
- (2) increase the efficiency of these efforts, and
- (3)avoid normative hypertrophy, which would, paradoxically, increase legal certainty.

So, after discarding subjects related to AI that would only need clarifying notes, guidelines or minor legal changes, a list of AI-related subjects that need legal development can be elaborated. The Commission White Paper on artificial intelligence of 19 February 2020 (pp. 21 et seq.) provides us with criteria on the risk level.

- **On industry intervention**

The involvement of industry is essential in the regulatory development of AI, as in the case of any other technology, since industry is directly involved in research and development of AI. No other actor could explain and highlight better problems, challenges, limitations, legal needs or convenient options for regulatory frameworks.

Nonetheless, this is not a statement in favour of industry-led intervention, nor in favour of co-leadership between public institutions and the industry. As explained before, soft law and voluntary labelling schemes should be complementary. Moreover, it does not seem reasonable that one of the parties most affected by this future regulation would be the one that develops and/or monitors its compliance.

Nevertheless, for those areas where political agreements are not possible by soft law, voluntary labelling schemes or even self-regulatory codes developed and monitored by industry are better than nothing.

Feedback on specific proposals

- **Black box as a problem? Objective liability as an answer**

The black box phenomenon refers to the difficulty (or even impossibility) of knowing entirely the reasoning process of the AI tool, that is, the path between entering big amounts of raw data and obtaining useful knowledge. This phenomenon is argued either to (1) elude liability or to (2) dissuade of trespassing certain limits when developing rules to boost algorithm transparency.

The black box phenomenon as an argument to elude liability (e.g., liability for a defective product) works as follows:

- (1) it is impossible to know the reasoning process by which the algorithm takes a decision or makes a suggestion (e.g., focus on certain people to investigate if there is a social security system fraud – SyRI case – select overwhelmingly white men for certain jobs – Amazon recruitment algorithm, already retired -);
- (2) this makes it difficult (or impossible) to detect who beaks a rule, and when a rule is broken;

- (3) this situation renders it impossible to claim for liability due to bad development of the algorithm when it is not possible to know with certainty (a) that this is the problem and/or (b) who introduced the wrong instructions or what instructions were revealed as wrong in certain situations.

This argument may have been successful because we are faced with a relatively new challenge. It is also impossible to know the human reasoning process, that is, the path between observing big amounts of information and making certain decisions based on our thoughts (e.g., a driver who overtakes recklessly). In certain situations, it can also be very difficult to know exactly what went wrong in a car accident. Despite all this, liability rules have mechanisms to designate a party responsible for compensating the damage, even if it has been impossible to know exactly who or what caused the damage: that is, the objective liability paradigm. Objective liability rules for vehicles are one of many examples.

Objective liability is the current paradigm: subjective liability (or liability based on culpability) was demonstrated to be slightly useful in the XIX century, giving way to objective liability, that is, liability not linked to culpability but based on the risk generated by certain activities[3] (such as driving a car). This (not so) new paradigm is based on two assumptions: since (1) certain activities are beneficial or even necessary for society, (2) society then must accept the risk generated by those activities and develop systems (objective or risk-based liability) to compensate unavoidable future damages related to them. Most of the times, liability rules combine both culpability and risk/benefit criteria.

There is no reason to exclude AI tools and their depending and related activities from this liability paradigm. First, it would be necessary to accept that AI is a risky activity (as explained in the Commission White paper on artificial intelligence 2020) and, so, that risk-based liability is the appropriate inspiring paradigm to develop liability rules on the topic. Hence, the black box phenomenon is not a serious impediment to claim compensation for damages or develop a system in the future that guarantees it.

For all that has been said, we must warn of the basic error that could be made if a future development of an AI civil liability system had, as its main reference, the subjective civil liability paradigm, as it seems when reading the following statement: "if safety risks materialise, the characteristics of AI technologies mentioned above may make it difficult for persons having suffered harm to obtain compensation under the current EU product liability legislation because those characteristics make it difficult to obtain documentation that would allow *to identify a person responsible for the damage and to trace back the damaging outcome to a particular human action or omission*".

Black box phenomenon is also argued when trying to dissuade the development of transparency binding rules for algorithms in order to protect fundamental rights.

[3] Ulrich BECK, Risiko-Gesellschaft. Auf dem Weg in eine andere Moderne, Suhrkamp, Frankfurt am Mein, 1986, Ulrich BECK, Risk Society, Towards a New Modernity, Sage Publication, London, 1992, (trad. por Mark RITTER), Jonathan SIMON, "The Emergence of a Risk Society. Insurance, Law and the State", Socialist Law Review, núm. 95, 1987, pp.61-89.

Nonetheless, what matters is developing rules to obtain a level of algorithm transparency that prevents discriminatory situations or violations of fundamental rights. This level of transparency does not refer to "technical transparency" but "legally sufficient transparency": that is, as much transparency as necessary to avoid discrimination or violation of fundamental rights (or presumptions that make up for the lack of transparency), together with sanctioning criteria. Sanctioning criteria should be enough to dissuade AI developers and AI beneficiaries from violating fundamental rights or discrimination. None of these (legal) aims are contradictory to the existence of the (technical) black box phenomenon.

- **Accountability as an immediate tool to get algorithm transparency**

EDPB Guidelines and EDPS opinions have been revealed as very useful tools to get a more uniform interpretation and application of GDPR as well as to increase its fulfilment. The last of these Guidelines clarifies the concepts of data controller and data processor and is open to feedback until 19th October 2020[4].

AI tools are fed by massive data processing. In other words, whenever obvious, an AI tool implies multiple data processing situations.

Art. 5.2 GDPR (accountability principle) states that a data controller will be responsible for, and able to demonstrate compliance with paragraph 1 (principles relating to processing of personal data).

If an AI tool process data, EDPB would be fully entitled to issue Guidelines on data processing by AI tools in order to solve, among other problems, who is the data controller and data processor or, to summarise, specify criteria for the fulfilment of GDPR when developing and using AI tools (including, for instance, (1) if/when and (2) how the data controller shall carry out the assessment of the impact of the envisaged processing operations on the protection of personal data ex art. 35 GDPR)[5].

Among the principles relating to the processing of personal data, we would like to highlight the following: data shall be processed lawfully, fairly and in a transparent manner (art. 5.1.a); data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes (purpose limitation principle, art. 5.1.b); data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (data minimisation principle, art. 5.1.c); or, data shall be processed in a manner that ensures appropriate security of the personal data (art. 5.1.e).

The data controller is, according to art. 5.2 GDPR, not only responsible for the fulfilment of these principles but also shall also be able to demonstrate it. There is no

[4] https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2020/guidelines-072020-concepts-controller-and-processor_en.

[5] Article 35.1 GDPR states that "where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operation on the protection of personal data".

other way for the AI-tool data controller to comply with this rule than guaranteeing transparency of the algorithm and the AI-tool. Also, according to art. 5.2, the AI-tool data controller will be responsible for this transparency.

To summarise:

- (1) using AI tools implies data processing; hence, this activity shall be compliant with GDPR;
- (2) there is an obligation to designate an AI-tool data controller, who will be responsible for, and able to demonstrate, compliance with art. 5.1 GDPR (due to art. 5.2 GDPR);
- (3) to comply with the accountability principle, the AI-tool data controller will also be responsible to guarantee (and show) an adequate level of AI-tool transparency, in order to be able to demonstrate compliance with art. 5.1 GDPR as required by art. 5.2 GDPR.

This reasoning or a similar one can be independently issued or be part of a broader document such as, for instance, the future EDPB Guidelines for using AI tools.