

The European Parliament's Position on the AI Act: a human rights-oriented approach and an attempt to regulate generative AI systems

Gabriele Rugani (Post-Doctoral Researcher in EU Law and International Law – University of Pisa) – 5 July 2023

SUMMARY: 1. Introduction: the previous phases of the legislative process for the adoption of the AI Act. – 2. The EP's Position on the AI Act: the expansion of “unacceptable” and “high risk” AI systems and the increase of fines. – 3. The EP's Position on the AI Act: the provisions on foundation models and generative AI systems (follows). – 4. Concluding remarks: two considerations on the next steps of the legislative process.

1. Wednesday 14 June 2023 marked a crucial moment in the legislative process for the adoption of the so-called “Artificial Intelligence Act” (“AI Act”): the European Parliament (EP) adopted its negotiating position on such instrument, with 499 votes in favour, 28 against and 93 abstentions; interinstitutional negotiations can now start in order to agree a final text with the Council. The EP's position is extremely interesting for several reasons and this contribution will try to investigate the most relevant ones and to make some considerations on them. For a better understanding, however, we must preliminarily retrace the previous key phases of the *iter legis* at stake.

On 19 February 2020, in its White Paper “On Artificial Intelligence – A European approach to excellence and trust” (COM/2020/65 final), the European Commission (EC) highlighted the need for a “regulatory framework” on artificial intelligence, in order to “build trust among consumers and businesses in AI, and therefore speed up the uptake of the technology”.

On 21 April 2021, the EC consequently presented the Proposal for a Regulation of the EP and of the Council “laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts” (COM/2021/206 final). Such Proposal follows what can be defined as a “risk-based approach” (on such topic, see for instance: G. De GREGORIO, P. DUNN, *The European Risk-Based Approaches: Connecting Constitutional Dots in the Digital Age*, in *Common Market Law Review*, 2022, pp. 473-500). Indeed, it differentiates between uses of AI that create an “unacceptable risk”, a “high risk”, and “low or minimal risk”. AI systems whose use is considered “unacceptable” should be *prohibited*: among them, it is

possible to mention those practices that have a significant potential to manipulate persons through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities in order to materially distort their behaviour in a manner that is likely to cause them or another person psychological or physical harm; the Proposal also prohibits AI-based social scoring for general purposes done by public authorities, and even the use of “real time” remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless certain limited exceptions apply (Title II, i.e. Art. 5 of the Proposal). AI systems that create a “high risk” to the health and safety or fundamental rights of natural persons, instead, should be permitted on the European market, but subject to compliance with certain mandatory requirements and an *ex-ante* conformity assessment (Title III, i.e. Articles 6-51); in particular, there are two main categories of “high risk” AI systems: AI systems intended to be used as safety component of products that are subject to third party *ex-ante* conformity assessment and stand-alone AI systems with mainly fundamental rights implications (that are explicitly listed in Annex III of the Proposal). On the contrary, if the risk level does not fall in the first two categories, there should be transparency obligations in certain cases (Title IV, i.e. Art. 52). Moreover, the Proposal encourages national competent authorities to set up “regulatory sandboxes” (Title V, Articles 53-55), that establish a controlled environment to test innovative technologies for a limited time (on the 2021 AI Act Proposal, see for instance: M. VEALE, F. Z. BORGESIU, *Demystifying the Draft EU Artificial Intelligence Act*, in *Computer Law Review International*, 2021, pp. 97-112; W. G. VOSS, *AI Act: the European Union’s Proposed Framework Regulation for Artificial Intelligence Governance*, in *Journal of Internet Law*, 2021, pp. 7-17).

On 6 December 2022, the Council adopted its common position (“General approach”) on the AI Act. The Council’s text, *inter alia*, narrows the definition of “AI systems” (Art. 3, No. 1, of the “General approach” only includes systems developed through machine learning approaches and logic- and knowledge-based approaches). With regard to the classification of AI systems as “high risk”, it contains new provisions according to which the significance of the output of the AI system in respect of the relevant action or a decision to be taken should also be taken into account when classifying AI systems as “high risk”, in order to ensure that AI systems that are not likely to cause serious fundamental rights violations or other significant risks are not captured (Art. 6, para 3). Many of the requirements for “high risk” AI systems have been clarified and adjusted in such a way that they are more technically feasible and less burdensome for stakeholders to comply with. An explicit reference has been made to the exclusion of national security, defence and military purposes from the scope of the AI Act (Art. 2). AI regulatory sandboxes should also allow for testing of innovative AI systems in “real world conditions” (Art. 53), while new provisions would enable unsupervised “real-world testing” of AI systems (Art. 54a), under specific conditions and safeguards. And so on. It can be easily noticed that many of the amendments of the Council are mainly aimed at promoting the development of AI systems and at making the text more attentive to the needs of the enterprises.

2. After retracing the previous key phases of the legislative process for the adoption of the AI Act, that mainly involved the EC and the Council, it is now possible to consider the latest steps which took place in the framework of the EP. Within such institution, the discussions were led by the Committee on Internal Market and Consumer Protection (IMCO; rapporteur: Brando Benifei, S&D, Italy) and by the Committee on Civil Liberties, Justice and Home Affairs (LIBE; rapporteur: Dragos Tudorache, Renew, Romania) under a joint committee procedure.

On 11 May 2023, IMCO and LIBE adopted a draft report on the AI Act, that deeply amends the EC's Proposal, with 84 votes in favour, 7 against and 12 abstentions.

Finally, as anticipated, on 14 June 2023 the IMCO/LIBE report was adopted in plenary, with an overwhelming majority (499 votes in favour, 28 against and 93 abstentions) and without amendments outside the text adopted at the Committee level. Now that the EP's position is set, interinstitutional negotiations with the Council can start in order to reach an agreed text. As already mentioned, however, the above cited position is extremely interesting for several reasons: it is now possible to examine some of the most relevant ones.

First of all, the EP's position significantly expands the list of Art. 5, which includes AI systems with an "unacceptable" level of risk that must consequently be prohibited. In other words, the EP tries to add bans on further intrusive and discriminatory uses of AI. Let's make some relevant examples. As partially anticipated, the EC's Proposal prohibits "the use of 'real-time' remote biometric identification systems in publicly accessible spaces", but only "for the purpose of law enforcement", and "unless and in as far as such use is strictly necessary" for certain objectives; the EP's position, instead, prohibits "the use of 'real-time' remote biometric identification systems in publicly accessible spaces" *tout court*, without any further condition or specification. Moreover, the Parliament adds *ex novo* a ban on "'post' remote biometric identification systems", with the only exception of law enforcement for the prosecution of serious crimes and only after judicial authorisation, and on "biometric categorisation systems that categorise natural persons according to sensitive or protected attributes or characteristics", such as gender, race, ethnicity, citizenship status, religion, political orientation.

The Parliament also introduces prohibitions concerning AI systems aimed at "making risk assessments of natural persons [...] in order to assess the risk of a natural person for offending or reoffending", based for example on "the person's location" or on "past criminal behaviour"; AI systems that "infer emotions of a natural person in the areas of law enforcement, border management, in workplace and education institutions"; and also "AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet".

The EP's position is seeking to expand even the classification of systems that create a "high risk", provided by Art. 6 and by Annex III (devoted to "High-risk AI systems referred to in Article 6(2)"). In particular, the Parliament included systems that pose a significant risk of harm to the health, safety or fundamental rights of natural persons, as well as to the environment. There are also other

additions: for instance, “AI systems intended to be used for influencing the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote in elections or referenda”.

Moreover, the Parliament adds to the “high risk” list the recommender systems, to recommend to the recipient of the service user-generated content, of “social media platforms that have been designated as very large online platforms within the meaning of Article 33 of Regulation EU 2022/2065”; the latter provision, in turn, refers to “online platforms and online search engines which have a number of average monthly active recipients of the service in the Union equal to or higher than 45 million”. This means, for example, that also as Meta and Twitter will fall under such classification (on the point, see for instance: F. Y. CHEE, S. MUKHERJEE, *EU lawmakers vote for tougher AI rules as draft moves to final stage*, in *reuters.com*, 14 June 2023).

It is worth considering the fact that the EP’s intention is also to go towards higher fines in case of non-compliance with the AI Act and, more specifically, with the prohibition of the artificial intelligence practices referred to in Art. 5. The EC’s Proposal, indeed, in Art. 71 refers to “administrative fines of up to 30 000 000 EUR or, if the offender is company, up to 6 % of its total worldwide annual turnover for the preceding financial year, whichever is higher”. In Art. 71 of the Parliament’s Position, instead, the two thresholds are increased to “40 000 000 EUR” and to “7 %” respectively (on the point see: K. J. NAHRA AND OTHERS, *European Parliament Adopts Negotiating Position on the AI Act*, in *WilmerHale Privacy and Cybersecurity Law Blog*, 15 June 2023).

3. In the EP’s Position there is also another crucial innovation, which must be examined on its own. The Parliament, indeed, takes into consideration two relevant notions that were not even mentioned in the EC’s Proposal: we refer to “foundation models” and “generative AI systems”. Foundation models are AI system models that are trained on broad data at scale, are designed for generality of output, and can be adapted to a wide range of distinctive tasks (Art. 3, para 1, No. 1c of the EP’s Position). Generative AI systems, that are based on foundation models, are AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video (Art. 28b, para 4, of the EP’s Position). The perfect example of generative AI systems is represented by the well-known ChatGPT: a powerful language model, trained on unprecedented amounts of data and able to engage in astonishingly diverse conversations, from writing movie reviews and poems to grading school essays, judging resumes or writing software code. Across a range of use cases and contexts, it is enough to ask in natural language in order to get a smooth-sounding answer. Millions of people are already using it (on the point see: N. HELBERGER, N. DIAKOPOULOS, *ChatGPT and the AI Act*, in *Internet Policy Review*, 16 February 2023, p. 2).

In order to better understand the issue, it is necessary to move back to the EC’s Proposal. The fact that it does not take into consideration the abovementioned AI systems, in truth, is not surprising: in 2021, when the Proposal was presented, foundation models and generative AI were not under the spotlight as they are two years later. However, such omission has been

strongly criticised by relevant authors. Some of them, moreover, highlight that the AI systems at stake seriously challenge the whole risk-based approach of the AI Act Proposal. According to the logic of such instrument, indeed, the classification in risk categories of an AI system depends on the purpose of use that the provider envisages. But systems such as ChatGPT have no pre-defined purpose: it is not the provider, but rather the user who determines how the system will be used; end users, however, are largely excluded from the scope of the Proposal. In order to solve the above-described problem, several possibilities have been suggested. One is to include all generative AI systems in the “high risk” category, since it cannot be excluded that they may be used also in a high-risk area; in that case, however, there may be a serious danger of over-regulation. Therefore, another suggestion is to avoid just fitting generative AI into the current provisions and to consider it more broadly as a general risk category in its own right (on the point see: N. HELBERGER, N. DIAKOPOULOS, *op. cit.*, p. 6).

Influenced by the suggestions to regulate foundation models and generative AI, the European Parliament, as anticipated, takes into consideration such systems. With reference to foundation models, Art. 28b is devoted precisely to “Obligations of the provider of a foundation model” and obliges providers of such models to assess and mitigate possible risks concerning, in particular, “health, safety, fundamental rights, the environment and democracy and the rule of law” (Art. 28b, para 2, let. a, of the EP’s Position). Moreover, providers must also register foundation models in the EU database referred to in Art. 60, i.e. the EU database for “high risk” AI systems (Art. 28b, para 2, let. g).

With regard to generative AI systems based on foundation models, the Parliament regulates them in Art. 28b, para. 4, of its Position. Such systems, first of all (Art. 28b, para 4, let. a), are required to comply with the transparency obligations outlined in Art. 52, par. 1, of the AI Act, according to which the natural persons exposed to the AI systems must be informed, in a timely, clear and intelligible manner, that they are interacting with an AI system; disclosing that the content was AI-generated would also help distinguishing so-called “deep-fake images” from real ones. Secondly (Art. 28b, para 4, let. b), the system must be trained, designed and developed in such a way as to ensure adequate safeguards against the generation of illegal content. Finally (Art. 28b, para 4, let. c), providers of such systems are required to “document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law” (on the provisions in the EP’s Position concerning foundation models and generative AI systems, see for instance: J. BRACY, *European Parliament vote pushes AI Act significant step forward*, in *iapp.org*, 14 June 2023; K. J. NAHRA AND OTHERS, *op. cit.*).

4. After analysing some of the most relevant aspects of the EP’s Position on the AI Act, in the opinion of the writer it is now possible to make two concluding remarks.

First of all, it can be said that the Parliament undoubtedly adopts a *human rights-oriented approach* to artificial intelligence. Such attitude is extremely evident and emerges from a significant number of elements. The first one is the expansion of *prohibited AI systems*, among which stand out real-time remote

biometric identification systems (real-time RBI). If the intention of the Commission is to forbid the use of such systems in publicly accessible spaces, but only for the purpose of law enforcement and unless certain conditions are met, as already seen the EP's Position tries to introduce a *comprehensive ban on real-time RBI*. Moreover, the Parliament intends to allow post-RBI, but only on condition of very strong safeguards. Obviously, the aim of the EP is to effectively protect our public spaces against intrusive and permanent forms of mass surveillance. The Parliament's design is further complemented by bans on discriminatory biometric categorisation; on systems used for emotion recognition in the areas of law enforcement, education, border control, and the workplace; and on predictive policing based on the profiling of people (on the point see: N. ASZÓDI, *EU Parliament vote on AI Act: Lawmakers chose to protect people against harms of AI systems*, in *algorithmwatch.org*, 15 June 2023).

The human rights-oriented approach of the EP can be also noticed by the remarkable expansion of the "high risk" list, that *inter alia* would include systems that pose significant harm to people's health, safety, fundamental rights or the environment; systems used to influence voters and the outcome of elections; and recommender systems used by very large social media platforms such as Meta and Twitter. Finally, another hint is represented by the increase of fines in case of non-compliance with the AI Act.

Ultimately, it can be stated that the Parliament is very ambitious in protecting the rights of citizens and consumers and that such approach could be strategic in view of the upcoming negotiations. As already mentioned, the Council already appeared to be more attentive to the needs of the enterprises, and in all likelihood will try to make the AI Act more technically feasible and less burdensome for economic operators. Therefore, starting from a position which is more shifted towards human rights is the only way for the EP to achieve, in the end, a satisfying agreement.

The second consideration that can be made, instead, concerns foundation models and generative AI. Such systems, that were not taken into consideration by the 2021 EC's Proposal, are now under the spotlight (the case of ChatGPT is iconic), and many authors suggested amendments to adapt the AI Act to their specificities. Therefore, the European Parliament introduced in its Position relevant provisions concerning such systems, incorporating many of the suggestions. In particular, as deemed appropriate by important scholars, the EP introduced *ad hoc* provisions (such as transparency obligations, adequate safeguards against the generation of illegal content, etc.) that take into account the peculiar features of the AI systems at stake, starting from the lack of pre-defined purpose.

However, unavoidably, other suggestions were not fully implemented. For example, given the fact that it is not the provider, but rather the user who determines how the systems at stake will be used, it was also advised to give more attention to the (contractual) relationship between providers and their users (on the point see: N. HELBERGER, N. DIAKOPOULOS, *op. cit.*). In the EP's Position, instead, even if users can fall in the scope of some obligations (for instance, when AI systems qualify as "high risk"), they are excluded from the

scope of Art. 28b, which is only devoted to “Obligations of the provider of a foundation model” (on the point see: W. LONG AND OTHERS, *European Parliament Adopts AI Act Compromise Text Covering Foundation and Generative AI*, in *sidley.com*, 23 June 2023).

In conclusion, in the near future it will be necessary to closely follow, on the one hand, the next steps of the legislative process for the adoption of the AI Act and, on the other hand, the unceasing development of these AI systems, that every day display new uses and implications. Only doing so it will be possible to assess if the steps taken by the European Parliament in regulating foundation models and generative AI are satisfying or if, on the contrary, there is still more to do.