Machines Learning the Rule of Law

VB verfassungsblog.de/ai-rol/

Sümeyye Elif Biber 13 July 2021

13 July 2021

EU Proposes the World's first Artificial Intelligence Act

On 21 April 2021, the European Commission (EC) proposed the world's first Artificial Intelligence Act (AIA). The proposal has received a warm welcome across the EU as well as from the US, as it includes substantial legal provisions on ethical standards. After its release, the media's main focus laid on the proposal's "Brussels Effect", which refers to the EU's global regulatory influence: EU laws exceed their "local" influence and become global standards. With the AIA, the EU has the potential to become the world's "super-regulator" on AI.

More than the Brussels Effect, however, the emphasis should lie on the EU's intention to explicitly protect the rule of law against the "rule of technology". Despite this expressed goal, the normative power of the regulation to ensure the protection of the rule of law seems inadequate and raises serious concerns from the perspective of fundamental rights protection. This shortcoming becomes most evident across three main aspects of the AIA, namely in the regulation's definition of AI systems, the AI practices it prohibits, and the preeminence of a risk-based approach.

Do We Need a Definition for AI?

In the current version of the Commission's AIA proposal, AI systems are broadly defined as "software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with" (Article 3(1)). The referenced Annex 1 covers machine learning, logic-, knowledge-based, as well as statistical approaches. In total, Article 3 presents an encyclopedic 44 definitions regarding AI systems and practices.

The large number of definitions reflects the attempt to consider a huge variety of different AI systems. This follows the standard EU approach, which includes defining the area of interest first, and focusing on necessary requirements and possible concerns second. However, due to the so-called "odd paradox", which describes the lack of a widely accepted definition of AI based on its extremely dynamic characteristics, the AIA should deviate from this strategy. Consequently, even the ambitious attempt proposed by the EC leaves some open questions. For instance, if a software does not have a human-defined objective but a computer-defined objective, would the AIA cover that software? Can we

clearly define, understand or follow the techniques listed in Annex I? The EC indirectly acknowledges this criticism when it places AI inside a “fast-evolving family of technologies”.

It becomes obvious that the AIA needs to consider both legal and technical aspects of AI and its development. To phrase the regulation in an overly-specific manner could prove counterproductive and impede legal certainty. Even worse, over-specified definitions in a dynamic field might open legal loopholes and allow providers to circumvent the obligations under the AIA, consequently putting fundamental rights at great risk. For example, “connectionist AI” has tested the limits of European patent law, as it is able to define its own technical purpose, raising questions such as whether a machine can be considered as an inventor? Such advanced applications may not fall within the conventional understanding of AI. Therefore, as highlighted by experts, it is better to focus on certain kinds of practices or uses that could raise questions in terms of fundamental rights, instead of focusing on definitions of particular technologies.

Prohibited AI Practices

Following the (rather counterproductive) efforts of defining AI systems, Article 5 of the AIA covers prohibited AI practices. They include (i) manipulation through “subliminal techniques”, (ii) exploitation of vulnerabilities of specific groups, (iii) general purpose social scoring, and (iv) mass surveillance through “real-time” remote biometric identification systems for law enforcement purposes. Exceptions to this prohibition allow the use of targeted “real-time” remote biometric identification for specific objectives such as preventing a “specific, substantial and imminent threat to the life” or searching for missing children.

The ban on social scoring is a clear reference to China, which has experienced a surge in reports concerning dystopian AI systems. One of the most recent examples concerns color-coded emotion-recognition systems, which exceed even the wildest Orwellian imagination. Evidently, some practices are profoundly incompatible with the enjoyment and protection of fundamental rights.

Unfortunately, the AIA’s prohibitions do not go far enough. For instance, Article 5 prohibits AI practices for manipulation and exploitation of vulnerabilities which can cause “physical and psychological harm”. This phrasing defines a very high threshold and makes the proof of individual harm extremely difficult. Furthermore, these practices might greatly impact society as a whole, even in the absence of any specific physical and psychological harm. The use of disinformation in elections, seen, for example, in the Facebook-Cambridge-Analytica scandal, show how the individualization of harm might offer inadequate prohibition or protection.

To select only one, the ban on mass surveillance through real-time “biometric identification systems” (BIS) invites misuse. “Identification” does not explicitly exclude “biometric categorization systems” (BCS). More than a mere semantic peculiarity, the difference between identification and categorization is substantial. While a BIS identifies

natural persons on the basis of their biometric data, a BCS is able to assign natural person to specific categories, such as sex, age, ethnic origin or sexual or political orientation. The first heading of Annex 3 of the AIA, "biometric identification and categorization", seems to acknowledge this difference. However, its content does only focus on BIS. Such vagueness leaves providers with legal uncertainty and users with serious concerns regarding fundamental rights and algorithmic bias. Indeed, on 18 June 2021, both the EDPB and the EDPS highlighted similar concerns in a joint opinion and called for ban on uses of biometric identification and categorization systems in publicly accessible spaces.

Next to the few AI practices mentioned in this short evaluation, many more should be scrutinized. On 12 January 2021, a group of civil society organizations sent an open letter to the EC, demanding the prohibition of several additional AI practices, including predictive policing, and the use of AI in border and migration control. Demonstrably, a number of predictive policing AI systems have racialized people and undermined the presumption of innocence. Here, even "minor" technical errors pose life-changing threats to individuals. Such a risk is undoubtedly unacceptable. Therefore, it is of utmost importance to consider and include these uses in the AIA within the prohibited practices category, to strengthen both the principle of legal certainty and the protection of fundamental rights in the digital realm. Further, in terms of fundamental rights protection, this short evaluation shows that the focus on practices is more effective than the focus on definitions. Identifying problematic practices should be the main focus of regulators.

The "Risk" of the Risk-Based Approach

Finally, the so-called risk-based approach needs to be scrutinized. Within the risk-based framework, the proposal presents a hierarchy between *unacceptable risk, high-risk, limited risk,* and *minimal risk* AI practices. The EC focuses on high-risk AI systems, which include the safety components of products which are already covered by existing product safety legislation, such as machinery, medical devices, and toys. Annex 3 lists additional stand-alone systems which directly threaten fundamental rights, such as those used for job recruitment and law enforcement processes.

As acknowledged in the Explanatory Memorandum, high-risk AI systems pose "significant risks" to fundamental rights. Thus, the AIA defines special requirements for high risk AI systems to protect fundamental rights and ensure non-discrimination. These include "appropriate data governance and management practices", technical documentation demonstrating that the system is in line with the rules, ensuring traceability of functioning, "sufficient" transparency of operation, effective human oversight, and "an appropriate level of accuracy, robustness, and cybersecurity". Despite such promising requirements, however, their review process is problematic.

The AIA provides for seven AI oversight institutions, namely, the "market surveillance authority", "national supervisory authority", "notified authority", "notified body", "conformity assessment body", "post-market monitoring", and the "European Artificial Intelligence

Board". The first three are defined as "national competent authorities" under Article 3(43). Without going into these institutions' roles and responsibilities or, fundamentally, why a need for seven institutions exists, encompassing regulation appears promising.

However, the regulatory system is heavily undermined by Article 43(2) and Annex VI AIA, which cover the review of the requirements. According to those provisions, most reviews are primarily subject to *internal* "conformity assessment procedures" conducted by providers; a procedure adapted from EU product safety law. In effect, providers will self-assess high-risk AI systems' compliance with the essential requirements. Only high-risk AI systems used for biometric identification fall within the conformity assessment by a "notified body".

Such a laissez-faire approach counters the "significant risks" high-risk AI systems pose "to fundamental rights". As also stated by the FRA, it is crucial to have robust mechanisms for the protection of fundamental rights rather than heavily relying on self-assessed checks conducted by providers. Therefore, conformity assessment procedure should be performed by external third parties for all high-risk AI systems, rather than for AI systems used for remote biometric identification only.

In their Explanatory Memorandum and Recitals, the proposal makes clear references to several fundamental rights protected by the European Charter of Fundamental Rights, including the right to human dignity, respect for private life and protection of personal data, and non-discrimination. Despite the obvious awareness, the AIA's risk-based approach endangers fundamental rights and should thus be clarified and scrutinized closely to avoid abuses.

Concluding Remarks

The public authorities, policymakers, and private entities in the EU are increasingly aware of the possible dangers advanced AI systems might pose. Reports, communications, statements and ethical guidelines attempt to secure fundamental European values and principles throughout the production process of AI, from the design phase to its deployment. It is of utmost importance that these initiatives take into consideration Article 2 of the Treaty of the European Union, the European Charter of Fundamental Rights as well as the relevant articles of the GDPR. AI should not impinge upon concepts such as human dignity, equality, justice, non-discrimination and non-stigmatization, autonomy and individual responsibility, informed consent, privacy and social responsibility.

The AIA presents a great step towards a regulated future of AI. By presenting the world's first legal framework, the EU has underlined its intention of protecting European values against new challenges. However, there are still some serious concerns from the perspective of fundamental rights protection and the rule of law. By reassuring these concerns, the AIA will be ready to be on par with the technologies it seeks to master to achieve the rule of law.

SUGGESTED CITATION Biber, Sümeyye Elif: *Machines Learning the Rule of Law: EU Proposes the World's first Artificial Intelligence Act , VerfBlog,* 2021/7/13, https://verfassungsblog.de/ai-rol/, DOI: 10.17176/20210714-015912-0.

2 Comments

WRITE A COMMENT

1. We welcome your comments but you do so as our guest. Please note that we will exercise our property rights to make sure that Verfassungsblog remains a safe and attractive place for everyone. Your comment will not appear immediately but will be moderated by us. Just as with posts, we make a choice. That means not all submitted comments will be published.

2. We expect comments to be matter-of-fact, on-topic and free of sarcasm, innuendo and ad personam arguments.

3. Racist, sexist and otherwise discriminatory comments will not be published.

4. Comments under pseudonym are allowed but a valid email address is obligatory. The use of more than one pseudonym is not allowed.

Comment

Explore posts related to this:
AI, AI Act, Digital Policy, EU

Other posts about this region:
Europa