WilhelmGilliéron
AVOCATS

Auteur: Wilhelm Gilliéron Avocats | Le : 30 April 2024

The EU AI Act – 4 : Requirements related to high-risk AI systems

In our 2nd and 3rd papers of this series, we had put under scrutiny the classification of artificial intelligence systems under the Regulation.

We shall focus this week on the requirements to be complied with by high risk AI systems.

I. Preliminary remarks

Chapter 2 of the Regulation (art. 8 – 15) lays down specific requirements to be complied with by the developers of high risk AI systems (Art. 16 lit. a). The intended purpose and the state of the art will have to be taken into account, making it a dynamic compliance exercise.

Products containing an AI system that are already subject to compliance requirements in accordance with the legislations set out in Annex II of the Regulation (see our third paper related to the classification of high risk systems as well as the Blue Guide on the implementation of EU product rules 2022) will obviously have to comply with all the requirements they are subject to.

To ensure consistency and avoid unnecessary administrative burden or costs, providers of such products will however be granted some flexibility on operational decisions as to how to ensure compliance of their product with the overall legislations. They may for instance decide to incorporate the necessary testing and reporting processes, information and documentation in their already existing documentation to make it easier.

The requirements to be met by high risk AI systems are the following ones:

II. Risk Management systems (Art. 9)

Taking into account the potential impact of these systems, it is crucial that risks are identified both at the start of the AI creation process and throughout the lifecycle. The setting up of a risk management system is therefore a key requirement that needs to take

into account the intended purpose of the AI system (the risk measurements may indeed differ if the system is meant to impact persons under the age of 18 or vulnerable groups of people for instance).

The risk management system is a continuous iterative process to be planned and run through the entire lifecycle of the system at stake. It should be reviewed and updated to ensure its continuing effectiveness, as well as justification and documentation of any significant decisions and actions taken under the Regulation.

This process, to be carried out in light of the intended purpose of the system at stake and, as a result, with due consideration to the technical knowledge, experience, education and training of users, should identify risks or adverse impacts and implement mitigation measures for the known and reasonably foreseeable risks of the concerned AI system with regards to health, safety or fundamental rights. More generally, developers should not only understand potential harms to the concerned individuals or groups of persons, but also society, the company or the ecosystem.

The implementation of these measures through the development or design of the AI system, or through the provision of adequate technical documentation as contemplated in Art. 11 (what one may call *safe AI by design*) should lead to an acceptable overall residual risk.

The risk management measures should be tested, potentially in real world conditions (see Art. 54a) throughout the development process against defined metrics and probabilistic thresholds appropriate to the intended purpose.

The risk assessment should further evaluate and estimate risks that may emerge under conditions of reasonably foreseeable misuse, including other possibly arising risks based on the analysis of data gathered from the post monitoring system. Such system will be implemented by providers and used by deployers in accordance with the post monitoring plan that will be part of the technical documentation referred to in Annex IV, based upon a template to be adopted by the Commission within six months before the entry into application of this Regulation (which we understand to be the entry into force of the Regulation with regards to high risk AI systems and its related provisions).

Several models of AI risk assessment frameworks already exist such as, in particular, NIST (January 2023), the ISO/IEC 42001:2023 and ISO/IEC 23894:2023 standards or the more narrowly tailored Microsoft AI Security Risk Assessment Framework, respectively the Microsoft AI Impact Assessment Guide and related template, both released in June 2022.

Form an operational standpoint, it will be important for risk management team to review already existing risk management programs, assess which risks those programs are meant to address and determine if the planned AI project introduces new risk that would require new processes.

III. Data and data governance (Art. 10)

Generally speaking, proper governance requires developers to understand the key steps at different AI project phases, namely (i) the planning phase (which AI model should I use to solve my business problem? Do I have the right data? Should I look for new data?), (ii) the design phase (what data should I gather and in which format? Have they been properly prepared, cleansed and labeled? Have privacy related requirements been taken into account?), (iii) the development phase (have the model and the related algorithms been properly selected to reach the desired level of accuracy and interpretability of data? Has there been proper training, testing and validating with proper datasets?) to end up with (iv) the implementation one (has there been a readiness assessment prior to deployment? Have proper metrics been defined to set a measurement baseline to enable continuous monitoring?).

Data governance is particularly key in the first two phases, namely the planning and design phases.

“Garbage in, garbage out” as one would say. High quality data and related datasets plays a vital role in the training, validation and testing of AI systems to ensure their performance and avoid their becoming a source of discrimination. Datasets should have the appropriate statistical properties in terms of targets (*i.e.* persons or groups of persons in relation to whom the AI system is intended to be used, including as to the specific geographical, contextual, behavioral or functional setting), and address proper remedial action to mitigate potential biases (be they implicit biases, sampling or temporal ones) likely to affect fundamental rights, notably with regards to feedback loops (*i.e.* when data outputs influence inputs for future operations).

The EU AI Act expressly puts certain safeguards in Art. 10 to ensure that developers have the proper data governance and management practices in place, by providing the following information: (i) design choices and underlying assumptions, (ii) assessment of the availability, quantity and suitability of the datasets, (iii) data collection processes as to their origin and original purpose when personal data are involved, (iv) data preparation processing operations (annotation, labelling, cleaning, updating, enrichment and aggregation), (v) assessment of potential biases and appropriate measures to detect, prevent and mitigate them as well as the shortcomings that prevent compliance with the EU AI Act and how they can be addressed.

The EU AI Act allows the use of special categories of personal data within the meaning of Art. 9 GDPR, 10 of Directive 2016/680 or 10 of Regulation 2018/1785 if this is the only way to detect and correct biases; provided, however, that certain safeguards are satisfied, namely: (i) there is no alternative at disposal, such as the use of synthetic or anonymized data (a growing area with an increasing number of players, such as Mostly, an Austrian company which specializes in privacy-preserving synthetic data generation, leveraging

advanced AI algorithms to generate realistic and privacy-compliant datasets for various industries, including healthcare, finance, and marketing); (ii) technical measures are implemented to prevent the re-use of data and state of the art security and privacy-preserving measures, including strict controls and documentation as to the access; (iii) the data should not be shared; (iv) their use and the reason for such use (*i.e.* absence of any alternative at disposal) should be duly documented and (v) the data should be deleted once the biases have been corrected.

The preamble highlights the importance for the EU to provide high quality datasets through European common data spaces, notably with regards to health related data. It further stresses the importance that competent sectorial authorities may also play in that space.

One may refer to ISO 8000:51-2023 and ISO/IEC 38805 series related to data governance.

IV. Technical documentation (Art. 11)

To ensure the traceability of high-risk AI systems throughout their lifetime, technical documentation should be drawn so as to demonstrate compliance with the EU AI Act. This documentation should contain the elements described in Annex IV; provided, however, that (i) a single technical documentation may be drawn for products already submitted to regulatory requirements as set forth in Annex II ensuring compliance with both the EU AI Act and the applicable regulations to such products, and that (ii) SMEs and start-ups (terms whose constructions remain to be defined) may use a simplified technical form to be established by the Commission and accepted by the notified bodies.

This technical documentation should contain (i) a general description of the AI system (intended purpose, interaction with hardware, software or others AI systems, versions of relevant software or firmware, forms in which the AI system is place on the market, description of the hardware on which the system is intended to run, basic description of the user interface, instructions of use); (ii) a detailed description of the elements of the AI system and of the process for its development (methods and steps such as the use of pre-trained systems and third parties' tools, design specifications [logic, rationale and assumptions, classification choices, relevance of parameters, expected output], description of the system architecture and computational resources used, data requirements in terms of datasets as set out in Art. 10, assessment of human oversight measures as set out in Art. 14, description of pre-determined changes, validation and testing procedures used), cybersecurity measures); (iii) information about the monitoring, functioning and control of the AI system [capabilities and limitations, expected level of accuracy, foreseeable unintended outcomes and risks]; (iv) a description of the appropriateness of the performance metrics; (v) a detailed description of the risk management system as set out in Art. 9; (vi) a description of relevant changes made by the provider; (vii) a list of harmonized standards applied (such as the ISO standards mentioned above); (viii) a copy of the EU declaration of conformity and (ix) a description of the post-monitoring plan as set out in Art. 61(3).

V. Record keeping (Art. 12)

For the same reasons of traceability, high risk AI systems should allow the for the automatic recording of logs for the duration of the lifetime of the system.

For systems enabling remote biometric identification, these logs should provide the period of use (start and end date times of each use), the reference database against which input data has been checked, the input data as well as identify the natural persons involved in the verification of the results.

VI. Transparency and provision of information to deployers (Art. 13)

To address concerns related to opacity of AI systems, transparency should be required before such systems are placed on the market and designed in a manner to enable deployers to understand how such systems work, evaluate their functionality, comprehend their strengths and limitations and enable them to interpret the system's ouput and use it appropriately.

This transparency should take the form of instructions of use to be provided in a digital format and contain the following information (already to be contained in part in the technical documentation): (i) name of the provider and its authorized representative; (ii) characteristics, capabilities and limitations of the AI system (purpose; level of accuracy and cybersecurity measures required in Art. 15; known or foreseeable circumstances that may have an impact upon such accuracy, robustness or that may lead to risks to health, safety or fundamental rights; information relevant to explain the output and interpret it; performance regarding specific persons or groups of persons on which the system is intended to be used; relevant information on input data, training, validation and testing datasets used); (iii) changes to the system and its performance predetermined at the moment of the initial conformity assessment; (iv) the human oversight measures set out in Art. 4 and the ones put in place to facilitate the interpretation of the outputs; (vi) the computational and hardware resources needed, the expected lifetime of the system and the required maintenance and care measures, including as to software updates, and (vii) description of the mechanisms including in the AI system to record the logs.

VII. Human oversight

High risk AI systems have to be designed in such a way that natural persons can oversee their functioning to prevent or minimize the risks that may emerge when the system is used in accordance with its intended purpose or reasonably foreseeable misuse.

Depending upon the risks at stake and context of the use, such measures have to be implemented either prior to putting the system on

the market or enabling their implementation by users once the system is put into service.

Such measures should guarantee that built-in operational constraints cannot be overridden by the system itself and is responsive to human operator, who should have the necessary competence, training (to understand capabilities and limitations and be able to detect anomalies and dysfunctions) and authority to carry out that role.

The operator should have the required information to make informed decisions as to whether not to use the system, disregard the output or even stop it.

When such high risk AI system is used for remote biometric identification, the human oversight should be carried out by two natural persons unless such system is used for law enforcement, border control migration or asylum and Union Law considers such requirement to be disproportionate.

VIII. Accuracy, robustness and cybersecurity

Finally, high risk AI systems should perform consistently and have an appropriate level of accuracy, robustness and cybersecurity in light of their intended purpose and the generally acknowledged state of the art.

The expected level of performance metrics should be disclosed in the instructions of use, in a clear and understandable way.

The technical robustness requires the system to be resilient as regards attempts by third parties to alter their use, outputs or performance and address vulnerabilities such as data poisoning, model poisoning or adversarial examples to name a few. If need be, these mechanisms should include fail-safe plans so as to ensure the interruption of the system.

Products with digital elements falling under the Cyber Resilience Act will be considered as meeting the cybersecurity requirements of the EU AI Act if they comply with that piece of legislation.

In our next paper, we shall focus on the duties put upon each stakeholder by the EU AI Act, namely the developer, the importer, the distributor and the deployer.

Source :
https://www.wg-avocats.ch/en/actualites/data-and-privacy/the-eu-ai-act-4-requirements-related-to-high-risk-ai-systems/