Regulating Artificial Intelligence

Dan Shefet

May 2021

4 CATEGORIES

- Prohibited AI (Unacceptable Risk)
- High Risk
- Limited Risk
- Minimal Risk

- European Commission Proposal for a Regulation laying down Harmonised rules on Artificial Intelligence (21 April 2021: https://ec.europa.eu/newsroom/dae/items/709090)

DATA GOVERNANCE

- https://www.youtube.com/watch?v=Hz1fyhVOjr4

DAN SHEFET

PROHIBITED ARTIFICIAL INTELLIGENCE PRACTICES

- The prohibitions cover 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.
    - Toys encouraging dangerous behaviour
    - Social scoring

- The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the following objectives.
- The targeted search for specific potential victims of crime, including missing children; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or of a terrorist attack; (iii) the detection, localization, identification or prosecution of a perpetrator or suspect of a criminal offence.

HIGH-RISK

- Management and operation of critical infrastructure:
- AI systems intended to be used as safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.
- Education and vocational training:
- AI systems intended to be used for the purpose of determining access or assigning natural persons to educational and vocational training institutions.
- AI systems intended to be used for recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
- AI intended to be used for making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.
- Access to and enjoyment of essential private services and public services and benefits:
- AI systems intended to be used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;
- AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;
- AI systems intended to be used to dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.

HIGH-RISK

- Law enforcement:
- Systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;
- Systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;
- Systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences;
- AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups;
- AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.

HIGH-RISK

- Systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;
- Systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;
- Systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features;
- Systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.
- Administration of justice and democratic processes:
- Systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.
- Presumption of innocence

HIGH-RISK OBLIGATIONS

- Accuracy, robustness and cybersecurity
- High-risk AI systems shall be designed and developed in such a way that they achieve, in the light of their intended purpose, an appropriate level of accuracy,
- Before placing on the market or putting into service a high-risk AI system the authorised representative shall register that system in a specific EU database.
- High quality data sets
- Logging to ensure traceability
- Detailed documentation and transparency
- Appropriate human oversight

DATA GOVERNANCE

- High-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the following quality criteria:
- Training, validation and testing data sets shall be subject to appropriate data governance and management practices. Those practices shall concern in particular,
- The relevant design choices;
- Data collection;
- Relevant data preparation processing operations, such as annotation, labelling, cleaning, enrichment and aggregation;
- The formulation of relevant assumptions, notably with respect to the information that the data are supposed to measure and represent;
- A prior assessment of the availability, quantity and suitability of the data sets that are needed;
- Examination in view of possible biases;
- The identification of any possible data gaps or shortcomings, and how those gaps and shortcomings can be addressed.

LIMITED RISK

- Specific transparency obligations

MINIMAL RISK

- Video games, spam filters

TERRITORIALITY

- To prevent the circumvention of this Regulation and to ensure an effective protection of natural persons located in the Union, the Regulation applies to providers and users of AI systems that are established in a third country, to the extent the output produced by those systems is used in the Union.

LIABILITY

- A specific natural or legal person, defined as the provider, takes the responsibility for the placing on the market or putting into service of a high-risk AI system, regardless of whether that natural or legal person is the person who designed or developed the system.

- Under certain conditions, rapid availability of innovative technologies may be crucial for health and safety of persons and for society as a whole. It is thus appropriate that under exceptional reasons of public security or protection of life and health of natural persons and the protection of industrial and commercial property, Member States could authorise the placing on the market or putting into service of AI systems which have not undergone a conformity assessment.

PENALTIES

- The following infringements shall be subject to administrative fines of up to 30 000 000 EUR or, if the offender is a company, up to 6 % of its total worldwide annual turnover for the preceding financial year, whichever is higher:

- (a) non-compliance with the prohibition of the artificial intelligence practices.
- (b) non-compliance of the AI system with the requirements of Data Governance, accuracy, transparency.
- The non-compliance of the AI system with any requirements or obligations under this Regulation, other than those laid down in Articles 5 and 10, shall be subject to administrative fines of up to 20 000 000 EUR or, if the offender is a company, up to 4 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.
- The supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request shall be subject to administrative fines of up to 10 000 000 EUR or, if the offender is a company, up to 2 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.