



 **IA FIABLE**  
retos técnicos, éticos, legales,  
culturales y socio-económicos

# SOCIAL IMPACT OF AI

TELSEC4TAI



09.30 AM  
30-31 May, 2024

**AI and Regulation**  
**Ricard Martínez Martínez**

# European Union regulatory approach

- ❑ European Data Space Strategy
- ❑ Digital Decade
- ❑ Digital Sovereignty
  - Legal framework based on EU principles and values
  - Global territorial scope
  - Broad subjective scope.
    - ✓ Developers
    - ✓ Marketer
    - ✓ Importers
    - ✓ Users of information systems

TELECOM 54  
JAI 238  
COPEN 69  
CYBER 37  
DATAPROTECT 76  
EJUSTICE 11  
COSI 16  
IXIM 49  
ENFOPOL 63  
RELEX 180  
MI 151  
COMPET 154  
CODEC 412

**LEGISLATIVE ACTS AND OTHER INSTRUMENTS**

Subject: REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

# European Union regulatory approach

## ❑ Approach based on guaranteeing the rule of law and democratic values

- Fundamental rights impact assessment
- Systemic risk assessment for democracy and European societies.
- AI Ethical principles: ALTAI impact Assessment
- Complementary approaches: DPIA in GDPR

## ❑ Product-oriented approach:

- Law-based engineering processes
  - ✓ Sanbox and research
  - ✓ Rules for high-risk system design
  - ✓ Conformity assessment certification
  - ✓ Rules for Generative AI Conformity verification

# Definitions

- ❑ 'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments;
- ❑ 'General-purpose AI model' means an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;
- ❑ 'general-purpose AI system' means an AI system which is based on a general-purpose AI model and which has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems;

- ❑ “notifying authority” means the national authority responsible for setting up and carrying out the necessary procedures for the assessment, designation and notification of conformity assessment bodies and for their monitoring;
- ❑ ‘conformity assessment’ means the process of demonstrating whether the requirements set out in Chapter III, Section 2 relating to a high-risk AI system have been fulfilled; ‘conformity assessment body’ means a body that performs third-party conformity assessment activities, including testing, certification and inspection;
- ❑ ‘notified body’ means a conformity assessment body notified in accordance with this Regulation and other relevant Union harmonisation legislation;

# Step one: Is it forbidden?



# Prohibited AI Practices

## Preventing behaviour manipulation and free Will

- (a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm;
- (b) the placing on the market, the putting into service or the use of an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm;



## Preventing discrimination

- (c) the placing on the market, the putting into service or the use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following:
  - (i) detrimental or unfavourable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected;
  - (ii) detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity;



## ❑ Law enforcement systems based on risk prediction

- (d) the placing on the market, the putting into service for this specific purpose, or the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics; this prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity;

- Biometric predictive systems related with special categories of data

- (h) the use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement, unless and in so far as such use is strictly necessary for one of the following objectives:
  - (i) the targeted search for specific victims of abduction, trafficking in human beings or sexual exploitation of human beings, as well as the search for missing persons;
  - (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack;
  - (iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purpose of conducting a criminal investigation or prosecution or executing a criminal penalty for offences referred to in Annex II and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years.

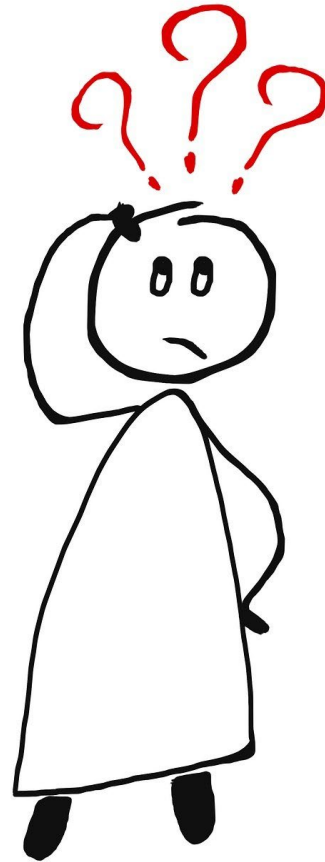
Point (h) of the first subparagraph is without prejudice to Article 9 of Regulation (EU) 2016/679 for the processing of biometric data for purposes other than law enforcement.

The use of ‘real-time’ remote biometric identification systems in publicly accessible spaces for the purposes of law enforcement for any of the objectives referred to in paragraph 1, first subparagraph, point (h), shall be deployed for the purposes set out in that point only to confirm the identity of the specifically targeted individual, and it shall take into account the following elements:

- (a) the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm that would be caused if the system were not used;
- (b) the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences.

- ❑ Internet scrapping: AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;
- ❑ AI systems to infer emotions of a natural person in the areas of workplace and education institutions, except where the use of the AI system is intended to be put in place or into the market for medical or safety reasons.

## Step two: If not banned, what kind of AI system is this?



## High-risk AI Systems conditions:

- ☐ (a) the AI system is intended to be used as a safety component of a product, or the AI system is itself a product, covered by the Union harmonisation legislation listed in Annex I;
- ☐ the product whose safety component pursuant to point (a) is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment, with a view to the placing on the market or the putting into service of that product pursuant to the Union harmonisation legislation listed in Annex I.

❑ An AI system referred to in Annex III. Except:

- It does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.
  - ✓ (a) the AI system is intended to perform a narrow procedural task;
  - ✓ (b) the AI system is intended to improve the result of a previously completed human activity;
  - ✓ (c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
  - ✓ (d) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.

❑ An AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons



# Step three: Fundamental rights impact assessment for high-risk AI systems

1. Prior to deploying a high-risk AI system referred to in Article 6(2), with the exception of high-risk AI systems intended to be used in the area listed in point 2 of Annex III, deployers that are bodies governed by public law, or are private entities providing public services, and deployers of high-risk AI systems referred to in points 5 (b) and (c) of Annex III, shall perform an assessment of the impact on fundamental rights that the use of such system may produce. For that purpose, deployers shall perform an assessment consisting of:
  - (a) a description of the deployer's processes in which the high-risk AI system will be used in line with its intended purpose;
  - (b) a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used;
  - (c) the categories of natural persons and groups likely to be affected by its use in the specific context;
  - (d) the specific risks of harm likely to have an impact on the categories of natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13;
  - (e) a description of the implementation of human oversight measures, according to the instructions for use;
  - (f) the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms.
2. The obligation laid down in paragraph 1 applies to the first use of the high-risk AI system. The deployer may, in similar cases, rely on previously conducted fundamental rights

# Step four: designing Requirements

## *Article 9*

### *Risk management system*

1. A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI systems.

## *Article 10*

### *Data and data governance*

1. High-risk AI systems which make use of techniques involving the training of AI models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria referred to in paragraphs 2 to 5 whenever such data sets are used.

## *Article 11*

### *Technical documentation*

1. The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.

## *Article 12*

### *Record-keeping*

1. High-risk AI systems shall technically allow for the automatic recording of events (logs) over the lifetime of the system.

## *Article 13*

### *Transparency and provision of information to deployers*

1. High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret a system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured with a view to achieving compliance with the relevant obligations of the provider and deployer set out in Section 3

## *Article 15*

### *Accuracy, robustness and cybersecurity*

1. High-risk AI systems shall be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness, and cybersecurity, and that they perform consistently in those respects throughout their lifecycle.

## *Article 14*

### *Human oversight*

1. High-risk AI systems shall be designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons during the period in which they are in use.

# Release:

## Article 4 AI literacy

Providers and deployers of AI systems shall take measures to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operation and use of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context the AI systems are to be used in, and considering the persons or groups of persons on whom the AI systems are to be used.

- ❑ 'AI literacy' means skills, knowledge and understanding that allow providers, deployers and affected persons, taking into account their respective rights and obligations in the context of this Regulation, to make an informed deployment of AI systems, as well as to gain awareness about the opportunities and risks of AI and possible harm it can cause;
- ❑ The operation of the system is subject to information sharing and post-market surveillance processes.
  - Providers of high-risk AI systems placed on the Union market shall report any serious incident to the market surveillance authorities of the Member States where that incident occurred

## **Chapter VIII**

### **EU database for high-risk AI systems**

#### *Article 71*

#### *EU database for high-risk AI systems listed in Annex III*

1. The Commission shall, in collaboration with the Member States, set up and maintain an EU database containing information referred to in paragraphs 2 and 3 of this Article concerning high-risk AI systems referred to in Article 6(2) which are registered in accordance with Articles 49 and 60 and AI systems that are not considered as high-risk pursuant to Article 6(3) and which are registered in accordance with Article 6(4) and Article 49. When setting the functional specifications of such database, the Commission shall consult the relevant experts, and when updating the functional specifications of such database, the Commission shall consult the Board.

# Classification of general-purpose AI models as general-purpose AI models with systemic risk

- ❑ A general-purpose AI model shall be classified as a general-purpose AI model with systemic risk if it meets any of the following conditions:
  - it has high impact capabilities evaluated on the basis of appropriate technical tools and methodologies, including indicators and benchmarks;
    - ✓ AI model shall be presumed to have high impact capabilities pursuant, when the cumulative amount of computation used for its training measured in floating point operations is greater than 10
    - ✓ The relevant provider shall notify the Commission without delay and in any event within two weeks after that requirement is met or it becomes known that it will be met
      - its notification, sufficiently substantiated arguments to demonstrate that, exceptionally, although it meets that requirement, the general-purpose AI model does not present, due to its specific characteristics, systemic risks and therefore should not be classified as a general-purpose AI model with systemic risk.
  - Based on a decision of the Commission, ex officio or following a qualified alert from the scientific panel

# Criteria for the designation of general-purpose AI models with systemic risk

- ☐ (a) the number of parameters of the model;
- ☐ (b) the quality or size of the data set, for example measured through tokens;
- ☐ (c) the amount of computation used for training the model, measured in floating point operations or indicated by a combination of other variables such as estimated cost of training, estimated time required for the training, or estimated energy consumption for the training;
- ☐ (d) the input and output modalities of the model, such as text to text (large language models), text to image, multi-modality, and the state of the art thresholds for determining high-impact capabilities for each modality, and the specific type of inputs and outputs (e.g. biological sequences);
- ☐ (e) the benchmarks and evaluations of capabilities of the model, including considering the number of tasks without additional training, adaptability to learn new, distinct tasks, its level of autonomy and scalability, the tools it has access to;
- ☐ (g) whether it has a high impact on the internal market due to its reach, which shall be presumed when it has been made available to at least 10 000 registered business users established in the Union;
- ☐ (g) the number of registered end-users.

# Obligations for providers of general-purpose AI models

- ❑ (a) Draw up and keep up-to-date the technical documentation of the model
- ❑ (b) Draw up, keep up-to-date and make available information and documentation to providers of AI systems who intend to integrate the general-purpose AI model into their AI systems:
  - i) enable providers of AI systems to have a good understanding of the capabilities and limitations of the general-purpose AI model and to comply with their obligations pursuant to this Regulation; and
  - (ii) contain, at a minimum, the elements set out in Annex XII put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;

The obligations set out in paragraph 1, points (a) and (b), shall not apply to providers of AI models that are released under a free and open-source licence under certain accessibility, information and transparency requirements

- ❑ (c) put in place a policy to comply with Union law on copyright and related rights, and in particular to identify and comply with, including through state-of-the-art technologies, a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790;
- ❑ (d) draw up and make publicly available a sufficiently detailed summary about the content used for training of the general-purpose AI model, according to a template provided by the AI Office.



# Obligations of providers of general-purpose AI models with systemic risk

## ☐ Risk management

- (a) perform model evaluation in accordance with standardised protocols and tools reflecting the state of the art, including conducting and documenting adversarial testing of the model with a view to identifying and mitigating systemic risks;
- (b) assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk;

## ☐ Serious incidents management (evidence)

- (c) keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, relevant information about serious incidents and possible corrective measures to address them;

## ☐ Security

- (d) ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risk and the physical infrastructure of the model.

# Accountability of the producers

## *Article 56*

### *Codes of practice*

1. The AI Office shall encourage and facilitate the drawing up of codes of practice at Union level in order to contribute to the proper application of this Regulation, taking into account international approaches.

# National & regional regulatory sandboxes

## □ Goals:

- improving legal certainty to achieve regulatory compliance with this Regulation or, where relevant, other applicable Union and national law;
- supporting the sharing of best practices through cooperation with the authorities involved in the AI regulatory sandbox;
- fostering innovation and competitiveness and facilitating the development of an AI ecosystem;
- contributing to evidence-based regulatory learning;
- facilitating and accelerating access to the Union market for AI systems, in particular when provided by SMEs, including start-ups.
- Warning:
  - ✓ The AI regulatory sandboxes shall not affect the supervisory or corrective powers of the competent authorities supervising the sandboxes, including at regional or local level. Any significant risks to health and safety and fundamental rights identified during the development and testing of such AI systems shall result in an adequate mitigation. National competent authorities shall have the power to temporarily or permanently suspend the testing process, or the participation in the sandbox if no effective mitigation is possible, and shall inform the AI Office of such decision. National competent authorities shall exercise their supervisory powers within the limits of the relevant law, using their discretionary powers when implementing legal provisions in respect of a specific AI regulatory sandbox project, with the objective of supporting innovation in AI in the Union

## ❑ Secure environment

- AI regulatory sandboxes shall provide for a controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems for a limited time before their being placed on the market or put into service pursuant to a specific sandbox plan agreed between the providers or prospective providers and the competent authority. Such sandboxes may include testing in real world conditions supervised therein.

## ❑ Oriented risk support, supervision and guidance

- Competent authorities shall provide, as appropriate, guidance, supervision and support within the AI regulatory sandbox with a view to identifying risks, in particular to fundamental rights, health and safety, testing, mitigation measures, and their effectiveness in relation to the obligations and requirements of this Regulation and, where relevant, other Union and national law supervised within the sandbox.

## ❑ Regulatory supervision

- Competent authorities shall provide providers and prospective providers participating in the AI regulatory sandbox with guidance on regulatory expectations and how to fulfil the requirements and obligations set out in this Regulation.

# Enforcement

- ☐ Market surveillance authorities (document request, evaluation, supervisión, request of measures)
- ☐ Authorities protecting fundamental rights shall have the power to request and access any documentation created or maintained under this Regulation in accessible language and format when access to that documentation is necessary for effectively fulfilling their mandates within the limits of their jurisdiction
- ☐ The Commission shall have exclusive powers to supervise and enforce general-purpose AI models

# Penalties

## ❑ General

- administrative fines of up to 15 000 000 EUR or, if the offender is an undertaking, up to 3 % of its total worldwide annual turnover for the preceding financial year, whichever is higher

## ❑ Supply of incorrect, incomplete or misleading information to notified bodies or national competent authorities in reply to a request

- administrative fines of up to 7 500 000 EUR or, if the offender is an undertaking, up to 1 % of its total worldwide annual turnover for the preceding financial year, whichever is higher

## ❑ Forbidden AI systems:

- administrative fines of up to 35 000 000 EUR or, if the offender is an undertaking, up to 7 % of its total worldwide annual turnover for the preceding financial year, whichever is higher.

## ❑ Proportionality in case of SMEs

# Final remarks

- ☐ Lessons learnt from GDPR
- ☐ Accountability is the key
- ☐ Human centric and fundamentals rights based approach
- ☐ AI Risk Impact Assessment
- ☐ Compliance by design



# Thanks

- ❑ [ricard.martinez@uv.es](mailto:ricard.martinez@uv.es)
- ❑ [https://www.linkedin.com/in/ricardmartinezmartinez/?locale=en\\_US](https://www.linkedin.com/in/ricardmartinezmartinez/?locale=en_US)