




# **AI ACT DATA AND DATA GOVERNANCE OF HIGH RISK SYSTEMS**



- Title I: General Provisions
- Title II: Prohibited AI Practices
- Title III: **High-risk systems**
- Title IV: Transparency obligations for certain systems
- Title V: Measures in support of innovation 
- Title VI: Governance
- Title VII: EU database for stand-alone high-risk AI systems
- Title VIII: Post-market monitoring, information sharing, market surveillance
- Title IX: Codes of Conduct
- Title X: Confidentiality and penalties

# Chapter 2

## Requirements for high risk systems



# Chapter 2

## Art. 10

### Data and data governance

1. 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 *quality criteria referred to in paragraphs 2 to 5.*

# Chapter 2

## Art. 10

# Data and data governance

2. Training, validation and testing data sets shall be subject to appropriate data governance and management practices. Those practices shall concern in particular,
- a) the relevant **design choices**;
  - b) data **collection**;
  - c) relevant data **preparation processing operations**, such as annotation, labelling, cleaning, enrichment and aggregation;
  - d) the **formulation of relevant assumptions**, notably with respect to the information that the data are supposed to measure and represent;
  - e) a **prior assessment of the availability, quantity and suitability** of the data sets that are needed;
  - f) examination in view of **possible biases**;
  - g) the **identification of any possible data gaps or shortcomings**, and how those gaps and shortcomings can be addressed.



## Chapter 2

### Art. 10

# Data and data governance

3. Training, validation and testing data sets shall be
  - *relevant, representative, free of errors and complete.*
  - They shall have the *appropriate statistical properties*,
    - *including, where applicable,*
    - *as regards the persons or groups of persons on which the high-risk AI system is intended to be used.*
  - These characteristics of the data sets may be met
    - *at the level of individual data sets*
    - *or a combination thereof.*

## Chapter 2 Art. 10


# Data and data governance

4. Training, validation and testing data sets shall take into account,
  - *to the extent required by the intended purpose,*
  - *the characteristics or elements that are particular to*
  - *the **specific geographical, behavioural or functional setting***
  - *within which the high-risk AI system is **intended to be used.***

## Chapter 2

### Art. 10

# Data and data governance

5. To the extent that it is strictly necessary
- for the purposes of **ensuring bias monitoring, detection and correction**
  - in relation to the high-risk AI systems, 
  - the providers of such systems **may process special categories of personal data** referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725,
  - subject to appropriate safeguards for the fundamental rights and freedoms of natural persons,
  - including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving measures,
  - such as pseudonymisation, or encryption where anonymisation may significantly affect the purpose pursued.



## Chapter 2 Art. 10

# Data and data governance

6. Appropriate data governance and management practices shall apply
  - for the development of high-risk AI systems
  - *other than those which make use of techniques involving the training of models*
  - in order to ensure that those high-risk AI systems comply with paragraph 2.



- **Article 11: Technical documentation**
- **Article 12: Record-keeping**



- HumanE-AI-NET

- AI Act: enables, prohibits, restricts
  - *Reasonable?*
  - *Pertinent?*