Javier Miranzo Díaz

Universidad de Castilla-La Mancha

The Artificial Intelligence Act as a risk regulation tool

Universitat de Barcelona

Universidad de Castilla-La Mancha

The need for regulation

- High potential of IA vs identified risks

- Growing concern among public bodies and employees
  + EU, National Governments, and subcentral authorities

- Increasing attention from the scientific literature
  + Concerns for the autonomy in the decision

- Uncertainty over solutions lead to diverting regulatory strategies
  + Initial attempts to rely on soft law
  + Critical voices demanded binding nature regulations
  + The risk of “adanic solutions”

Te EU approach

- First text approved in April 2021 (amendments proposed by the Council of the EU and the European Parliament June 2023)
  + Horizontal and risk-based regulatory approach.
  + Depending on the kind of AI, the context of implementation, etc.
  + Different potential damages.

- Prohibition of certain activities
- Authorizations or licenses
- Oversight control mechanisms
- Risk regulation: no specific obligations, but individual self-assessment and control.

The classification system

- Unacceptable risk: AI systems deemed to be a clear threat to security, livelihoods and human rights will be banned.
- High risk: AI technologies used in critical infrastructure, education and vocational training, product safety components, employment, essential public and private services, law enforcement, migration management and the administration of justice and democratic processes will be considered high risk, among others.
- Limited risk: Specific transparency requirements will be required for systems such as conversational robots, among others.
- Minimal or no risk: Free use of applications such as AI-based video games or spam filters will be allowed. The vast majority of AI systems fall into this category.

Unacceptable risk

- Those who "manipulate" citizens
- Those who take advantage of any of the vulnerabilities of a specific group of people
- Those whose purpose is to evaluate or classify with “unjustified, disproportionate, or unjustified” unfavourable consequences.
- Indiscriminate and general surveillance systems ("Big Brother")

High risk applications

- Intended to be used as a safety component of one of the devices covered by the Union harmonisation legislation
- It is included in Annex II: New Legislative Framework
- Or in Annex III:
  + (1) biometric identification and categorization of natural persons,
  + (2) management and operation of essential infrastructures (operation of road traffic and the supply of water, gas, heating and electricity);
  + (3) student assessment);
  + (4) employment, worker management and access to self-employment;
  + (5) access to and enjoyment of essential public and private services and their benefits;
  + (6) law enforcement issues (risk of criminal offences or reoffending, or risk to potential victims of crime, evidence in criminal proceedings, etc.);
  + (7) migration, asylum and border control management;
  + (8) administration of justice and democratic processes

Risk Management Stystem (article 9)

- No definition of "risk management system”
- It must be “implemented, documented and maintained”
- Integrated in other risk management system (European Parliament)
- four phases or stages:
  + (1) the identification and analysis of known and foreseeable risks;
  + (2) the estimation and assessment of the risks identified, both when the system is used in accordance with its intended purpose and when it is reasonably foreseeably misused;
  + (3) evaluation through the post-market monitoring system; and
  + (4) timely risk management measures
- However, **indetermination remains**

Infringement procedure

- Very serious infringements:
  + €30 million or up to 6% of total annual turnover

- Serious offences:
  + up to €20 million euros or 4% of annual turnover

- Minor infringements:
  + Up to 10 million euros or 2% of turnover

- Legal uncertainty + potential sanctions => disincentive

The answer is in the standards

- The need to standardize AI guarantees:
  + Committee on Artificial Intelligence
  + Canada prequalification system

- ISO:
  + ISO/IEC JTC 1/SC 42, on Artificial Intelligence,
  + ISO/IEC AWI TR 24368, on ethics in AI,
  + ISO/IEC PDTR 24028, on the confidence or robustness of AI software, or
  + ISO/IEC 23894, on the risk management of AI systems

- IEEE P7000™ issues at the intersection of technological and ethical considerations

The conformity assessment

- AI-systems will be agreed to comply with the requirements of the regulation if they are in conformity with
  + harmonised standards published in the Official Journal of the European Union (Article 40)
  + or common specifications adopted by the European Commission (Article 41).
- Agencies in charge:
  + European Committee for Standardisation (CEN),
  + the European Committee for Electrotechnical Standardisation (CENELEC),
  + and the European Telecommunications Standards Institute (ETSI)
- No subject to legislative procedure, scrutiny, lack o legitimacy etc.

The standards

- Controversy over its nature as delegated acts of the EU, BUT:
- In any case, constitutes a necessary and strictly directed implementing measure of the requirements
- Only EU and national standardization bodies
- Directive 2015/1535/EU of 9 September establishing a procedure for the provision of information in the field of technical regulations and rules relating to the services of the Information society
  + The commission starts and supervises the procedure (reinforced in 2018)
  + Publication in the Official Journal of the EU (transparency)