DEMOCRACY AND FUNDAMENTAL RIGHTS

FOUNDATIONS OF ARTIFICIAL INTELLIGENCE REGULATION

Prof. Dr. Mireille Hildebrandt

Legal protection by design & fundamental rights impact assessmen

- This concerns technologies that are fundamentally different
  - They automate error, bias and safety & health hazards
  - They redistribute power due to speed, remote control, network effects
- Some of them anticipate and pre-empt us – whether or not correct
  - Risks to fundamental rights and freedom of natural persons
  - Due to invisible visibility, undoing checks and balances rule of law
- Legal protection at individual and societal level will only work if:
  - Providers and deployers must conduct FRIAs
  - Providers and deployers must build checks and balances into the architectures

How to define AI?

1. in terms of systems (software, hardware, softwared infrastructure)
2. learning or not (the excel sheet)
3. the issue is impact: interaction with the environment (decisions or behaviour)

How to define AI?

- Focus should be on:
  - Impact (precautionary approach)
  - of a specific type of systems (broadly defined)

How to define AI?

OECD 2019:

- An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions
  - influencing real or virtual environments.
- AI systems are designed to operate with varying levels of autonomy

How to define AI?

- OECD 2019:

–AI system lifecycle: AI system lifecycle phases involve: i) 'design, data and models'; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; ii) 'verification and validation'; iii) 'deployment'; and iv) 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

–AI knowledge: AI knowledge refers to the skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes and best practices, required to understand and participate in the AI system lifecycle.

–AI actors: AI actors are those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.

–Stakeholders: Stakeholders encompass all organisations and individuals involved in, or affected by, AI systems, directly or indirectly. AI actors are a subset of stakeholders.

How to define AI?

OECD:

1. ‘Inclusive growth, sustainable development and well-being’, requiring a proactive approach of all stakeholders.
2. ‘Human-centred values and fairness’, demanding respect for the rule of law, human rights and democratic values, to be implemented by appropriate mechanisms and safeguards.
3. ‘Transparency and explainability’, calling for ‘a general understanding of AI systems’, the need to make stakeholders aware of their interactions with AI systems, asking that those affected by AI systems are enabled to understand the outcome, and to challenge it insofar as it adversely affects them, based on ‘easy-to-understand information on the factors, and the logic that served as the basis for the prediction, recommendation or decision’.
4. ‘Robustness, security and safety’, insisting that the actors ensure appropriate functioning in the case of ‘normal use, foreseeable use or misuse, or other adverse conditions’, preventing unreasonable safety risk.

The devil is in the details

EU AI Act, art. 14

- From ‘human in the loop’ to effective ‘human oversight’:
  - By natural persons in service of the deployers
  - Preventing or minimising the risks to health, safety or fundamental rights
  - Fully understanding the capacity and limitations
  - Being aware of automation bias
  - Being able to correctly interpret the output
  - Being able to overrule or even halt the system

Who is protected?

- Natural persons
- Corporations
- Public administration
- Judiciary
- Legislature

Who is liable for what?

- EU: mainly providers (legal protection by design)
  - Not the developers
  - Not those who deploy (or hardly)
  - Burden of proof, wrongfulness, damage?
  - Role of insurance?

Requirements for high risk systems?

- Sufficiently detailed and properly effective
  - Risk management also with regard to reasonably foreseeable other use
  - Data governance: appropriateness, reliability, representativeness, bias
  - Accuracy, robustness, cybersecurity
  - Record keeping and technical documentation
  - Post market monitoring
  - Automated logging
  - Quality management

Risk approach

- Requiring that providers foresee risks:
  - When used for its intended purpose
  - And when used for reasonably foreseeable other purposes
- Risks to health and safety
- Risks to fundamental rights and freedoms
  - Risk that these will be infringed or violated
  - Precautionary approach
  - Requiring legal protection by design