AI ACT
OVERVIEW OF
PROPOSAL 21 APRIL 21

What's up?

- Objectives of the Tutorial
- Architecture of the proposed Act
- Connection with GDPR
- Enforcement, individual rights, oversight

Objectives of the Tutorial

[notably including the study of the audio-slides]

- An in-depth first analysis of the proposed AI Act
  - *Addressing the kinds of issues developers within the HAI NET face*
  - *Showing the complexity of the legislative ecosystem*
  - *Highlighting the objectives of the Act*
  - *Giving a taste of the salience of its obligations*
  - *Commenting on enforcement issues*

Architecture

Impressive:

- Twofold aim:
  - *protection against threats to safety*
  - *protection against threats to fundamental rights*
- Considering the landscape of existing and upcoming legislation
  - *dedicated legislation for potentially unsafe products (machinery, toys, aircraft etc)*
  - *Charter of Fundamental Rights of the EU*
  - *GDPR, ePrivacy Regulation (upcoming)*
- Part of a major legislative program
  - *DSA, DMA, DGA (proposals 'released' in 2020)*
  - *Data Act and Liability regime still upcoming in 2021*

Architecture

The architecture of the AIA is **as simple as possible, but not simpler:**

- It deploys a broad definition of AI systems to offer broad protection
- It distinguishes between high risk systems, medium risk systems and other systems
- On top of that it defines four prohibited AI practices
- Not applicable to the military
- No new individual rights are attributed
- Obligations are imposed mainly on providers
- High risk systems are defined as such due to threats to safety or to fundamental rights

Architecture

- The focus is on
  - *high risk systems and*
  - *the requirements they must meet*
- to become available on the EU market and/or
- to be put into service and/or
- to be used.

Architecture

- The focus of these requirements seems to be on 4R* AI systems:
  - *part of the requirements see to it that the claimed functionality of these systems is verified, validated and tested before becoming available, while*
  - *other requirements see to it that providers anticipate the use for other purposes and prevent or mitigate ensuing threats to safety and fundamental rights.*

*resilient, robust, reliable, and responsible

Architecture

- Resilient – fit for its ‘intended purpose’ (= claimed functionality)
- Robust – dependable over the course of time (e.g. post market monitoring)

Architecture

- Reliable – trustworthy as to design and use, based on:
- *Risk management both when used*
- for ‘intended purpose’ and
- for ‘reasonably foreseeable misuse’
- *Data and data governance (e.g. high standards for training, validation and test data)*
- *Human oversight (e.g. high standards for natural persons tasked with oversight)*
- *Performance metrics, robustness and cyber security (e.g. high quality standards)*
- *Quality management (e.g. documented strategy for conformity assessment)*
- *Post market monitoring (e.g. sustained accountability)*
- *Proper documentation (e.g. including automatically generated logs)*

Architecture

- Responsible – preventing or mitigating potential fundamental rights interferences
  - *When used for its intended purpose*
  - *In case of reasonably foreseeable misuse (= other use than intended)*
  - *Monitoring duties with regard to discriminatory bias*
  - *Prohibition of practices that are unacceptable in a constitutional democracy*

Architecture

- Definition of AI system in art. 3(1):
- software that
- is developed with one or more of the techniques and approaches listed in Annex I
- and can for a given set of human-defined objectives,
- generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with;

Architecture

- Annex 1:

a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

c) Statistical approaches, Bayesian estimation, search and optimization methods.

Architecture

- Definition has a broad scope and is meant to provide broad protection
- It is not about what AI truly is (no metaphysical discussions)
- Meant to provide ‘effective and practical protection’
- The discussion should be about:
  - *whether in concrete AI systems*
  - *the right level of protection has been implemented*
- depending on the qualification as prohibited, high risk or other

Architecture

Roles (those addressed by the Act):

- Provider: entity that develops or has others develop with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge
- User: using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity

– *Also: importer, distributor, etc.*

Architecture

- Prohibition AI practices:
  - *manipulation, exploitation of vulnerable groups or individuals, social credit scoring by governments, remote biometric identification (with exceptions)*
- High risk AI systems
  - *Products or **safety** components of products regulated in EU legislative framework Annex II*
  - *AI systems as defined in Annex III (focused on **fundamental rights interferences**)*
- Medium risk AI systems
  - *Systems interacting with natural persons*
  - *Emotion recognition systems*
  - *Biometric categorisation systems*

Connections with GDPR

- Often, ‘users’ of the AIA will be the ‘controllers’ of the GDPR
- Qualification of high risk systems under the AIA is predefined in the AIA
- Qualification as high risk to fundamental rights under the GDPR depends on an impact assessment (DPIA): more granular and flexible
  - *Systems qualified as high risk in Annex III should be considered high risk in a DPIA?*
- Qualification as high risk in AIA does not imply lawfulness, this will also depend on compliance with other legislation such as GDPR (recital 41 AIA)
  - *AI systems that process personal data will have to comply with both the AIA and the GDPR, especially relevant in the case of high risk AI systems*

Connections with GDPR

- The GDPR provides for a 'the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision' in the case of 'a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'

- The AIA requires that 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 the AI system is in use'.

Enforcement, individual rights, oversight

- Steep fines when violating the requirements:
  - *Up to 30.000.000 Euro or 6% of global turnover*
- in case of violation of the prohibition of certain AI practices in art. 5
- In case of violation of the requirements of data and data governance in art. 10
  - *Up to 20.000.000 euro or 4% of global turnover*
- In case of all other violations of the AIA
  - *Up to 10.000.000 euro or 2% of global turnover*
- In case of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request

Enforcement, individual rights, oversight

AIA = administrative law, focused on oversight bodies and administrative fines

- Upcoming legislation will settle private law liability issues
- No new individual rights are attributed to natural persons

Enforcement, individual rights, oversight

I think it would help if a small set of rights were to be attributed to natural persons, while also including some collective rights:

- The right not to be subject to prohibited AI practices
- The right to object to decisions made by high-risk AI systems
- The right to file an injunction in a court of law, and to mandate that right to an NGO in case one is subjected to prohibited AI practices or to decisions made by high-risk AI systems
- The right of dedicated NGOs to file an injunction in their own name with respect to the rights under A and B

- Assuming that the upcoming AI liability framework will provide some forms of strict liability, in alignment with the product liability directive.

- HumanE-AI-NET

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

Questions?

- Please check the audio-slides on webdav
- Do not hesitate to raise whatever questions these audio-slides generate
- Or any use cases or missing links