UNIVERSITÄT
ZU KÖLN

AI Act

Nils Reiter,
nils.reiter@uni-koeln.de

December 19, 2023

INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

What do we know about the AI Act?

Current State

- Trilogue negotiations completed
  - I.e.: There is an agreement between Commission, Parliament and member states
- Proposal is public: AI Act
  - ...as a gigantic HTML page or – without navigation

What is Artificial Intelligence?

> *"artificial intelligence system" (AI system) means 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*

(Title I, Article)

What is Artificial Intelligence?

"artificial intelligence system" (AI system) means 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

(Title I, Article)

ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 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.

Classification of AI Systems

- Three classes
  - Prohibited Artificial Intelligence Practices (Title II)
  - High-Risk AI Systems (Title III)
  - Other AI Systems

Classification of AI Systems

- Three classes
  - Prohibited Artificial Intelligence Practices (Title II)
  - High-Risk AI Systems (Title III)
  - Other AI Systems
- No exceptions for “foundational models”

High-Risk AI Systems

- What is considered high risk?
  - AI systems if they are components of other systems that fall under listed regulations (e.g., "relating to lifts and safety components for lifts" and many more)
  - Specifically
    - Biometric identification and categorisation of natural persons
    - Management and operation of critical infrastructure
    - Education and vocational training
    - Employment, workers management and access to self-employment
    - Access to and enjoyment of essential private services and public services and benefits
    - Law enforcement
    - Migration, asylum and border control management
    - Administration of justice and democratic processes

High-Risk AI Systems

Requirements

- Risk management system (Article 9)
- Data and data governance (Article 10)
- Technical documentation (Article 11): "drawn up before that system is placed on the market[,] shall be kept up-to date"
  - Annex IV contains the minimum documentation
- Record-keeping (Article 12)
- Transparency and provision of information to users (Article 13)
  - "level of accuracy, robustness and cybersecurity [...] against which the high-risk AI system has been tested and validated and which can be expected"
  - "specifications for the input data, [...] information [on] training, validation and testing data"
- Human oversight (Article 14)
- Accuracy, robustness and cybersecurity (Art. 15)
- Central registration (Article)

Notified Bodies (Art. 33ff. / Annex VII)

- Public agency to verify that high-risk AI systems are in accordance with the regulations
- Independent of the provider of a high-risk AI system

> 4.3. The technical documentation shall be examined by the notified body. To this purpose, the notified body shall be granted full access to the training and testing datasets used by the provider, including through application programming interfaces (API) or other appropriate means and tools enabling remote access.
>
> 4.4. In examining the technical documentation, the notified body may require that the provider supplies further evidence or carries out further tests so as to enable a proper assessment of conformity of the AI system with the requirements set out in Title III, Chapter 2. Whenever the notified body is not satisfied with the tests carried out by the provider, the notified body shall directly carry out adequate tests, as appropriate.

Figure: Annex VII

Modulprüfungen

Drei Geschmacksrichtungen

- Klassisch-computerlinguistisch: Sprachphänomen, Operationalisierung mit ML, Evaluation
  - Z.B.: Automatische Erkennung sexistischer Äußerungen in Spiele-Chats
- Schrauben am Konzept: Konzept, Annotationsrichtlinien, mehrere Annotationsrunden mit Agreement etc.
  - Z.B.: Argumentation für und gegen den Einsatz von Wahlcomputer. Was sind Argumente, Behauptungen, Begründungen etc.
- Inhaltlich: Inhaltliche Fragestellung basierend auf Textanalyse beantworten
  - Z.B.: Wie positionieren sich verschiedene Parteien im Bezug auf die Chatkontrolle, basierend auf parlamentarischen Reden, Parteiprogrammen, social-media-Äußerungen?

References I