MASTT2040

THHINKBV 2024

AI Act

The new EU AI Act introduces a risk-based classification scheme for AI applications considering the level of risk posed by the AI application to individuals or society as a whole. The classification ranges from minimal risk to applications which are banned entirely. The term AI covers a wide variety of data analysis techniques which are already being exploited by companies and includes things like deep learning and machine learning.

The OECD definition of an Artificial Intelligence system' (AI system) is "a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions that influence physical or virtual environments".

The Act applies to organisations within Europe but also to organisations outside of Europe. There are, however, a number of exemptions with respect to the use of AI for national security, military and defence. There are also exemptions in the use of AI for research.

Different types of AI are defined according to risk:

- Prohibited AI - In some cases the risk of harm from using AI, such as in social scoring systems or manipulative systems is considered unacceptable and thus its use is banned.
- High-Risk AI - A number of high-risk applications are identified which could adversely affect citizens' lives, e.g. evaluating creditworthiness, educational opportunities, and critical infrastructure applications. In these cases strict assessment regimes will need to be followed.
- Limited Risk AI - Applications with a limited risk, e.g. image processing, recommender systems and chatbots, will need to follow best practices with respect to data quality and fairness.
- Minimal Risk AI - "minimal risk applications" are also defined to cover things like applications of AI such as spam filtering and video games. In these cases there is no need for regulation.

The most requirements are placed on providers and users of High-Risk AI. The Act also provides requirements for transparency and disclosure for generative AI.

Prohibited AI

The use of AI is prohibited in a number of areas:

- Social credit scoring systems
- Emotion recognition systems at work and in education

- AI used to exploit peoples' vulnerabilities, e.g. age, disability
- Behavioural manipulation and circumvention of free will
- Untargeted scraping of facial images for facial recognition
- Biometric categorisation systems using sensitive characteristics
- Specific predictive policing applications
- Law enforcement use of real-time biometric identification in public apart from limited, pre-authorised situations

High-Risk AI

- Medical devices
- Vehicles
- Recruitment, HR and worker management
- Education and vocational training
- Influencing elections and voters
- Access to services (e.g. insurance, banking, credit, benefits, etc.)
- Critical infrastructure management (e.g. water, gas, electricity, etc.)
- Emotion recognition systems
- Biometric identification
- Law enforcement, border control, mitigation and asylum
- Administration of justice
- Specific products and/or safety components of specific products.

Compliance Requirements for High-Risk AI

Key to the act will be that companies must demonstrate that the design, implementation and post-market entry phases are compliant with a range of activities. These include:

- Risk Management System
- Data and Data Governance
- Technical Documentation
- Record Keeping
- Transparency and provision of information to user
- Human Oversight
- Accuracy, Robustness and Cybersecurity
- Quality Management System
- Fundamental Rights Impact Assessment

High Risk AI Systems will have to undergo a Conformity Assessment (Article 19) producing the appropriate documentation to demonstrate adherence to the AI Act before going to market in the EU.

- Fundamental rights impact assessment and conformity assessment
- Registration in public EU database for high-risk AI systems
- Implementation risk management and quality management system
- Data governance (e.g. bias mitigation, representative training data etc.)

- Transparency (e.g. instructions for Use, technical documentation, etc.
- Human oversight (e.g. explainability, auditable logs, human-in-the-loop, etc.)
- Accuracy, robustness and cyber security (e.g. testing and monitoring)

Companies will need to comply with the Act but will be given a compliance grace period of between 6-24 months.

General Purpose AI

The Act also provide specific requirements for General Purpose AI (GPAI) and Foundational Models:

- Transparency for all GPAI (e.g. technical documentation, training data summaries, copyright and IP safeguards, etc.)
- Additional requirements of high-impact models with systemic risk: model evaluations, risk assessments, adversarial testing, incident reporting, etc.
- Generative AI individuals must be informed when interacting with AI (e.g. chatbots), AI content must be labelled and detectable (e.g. deepfakes).

Limited risk systems will also need to follow these requirements, but with less compliance scrutiny than High-Risk AI applications considering conformity assessments and product safety reviews.

Penalties and Enforcement

The fines for violations of the AI Act have been set as a percentage of the offending company's global annual turnover in the previous financial year or a predetermined amount, whichever is higher. It should be noted, however, that the provisional agreement provides for more proportionate caps on administrative fines for SMEs and start-ups in case of infringements.

- Up to 7% of global annual turnover or €35m for prohibited AI violations
- Up to 3% of global annual turnover or €15m for most other violations
- Up to 1.5% of global annual turnover or €7.5m for supplying incorrect information
- Caps on fines for SMEs and startups

Compliance will be enforced by a European "AI Office" and "AI Board" established centrally at the EU level. Market surveillance authorities will be set up in EU countries to enforce the AI Act. Any individual will be able to make complaints about non-compliance.