Ethikos Volume 37, Number 1. January 02, 2023

Artificial Intelligence Act: A European approach

By Patrick Wellens, CCEP-I, CIA, CFE, CRMA, MBA

Patrick Wellens (patrickwellens@hotmail.com) is currently a Compliance Manager for a division of a multinational pharma company based in Zurich, Switzerland. He is a Board Member of Ethics and Compliance Switzerland and co-chair of the Working Group Life Sciences.

Artificial intelligence (AI) is a technology that mimics human intelligence to perform tasks and can iteratively improve itself based on the information it collects.[1] AI is used widely in many technologies and industries; some may notice, others may not. These include (but are by no means limited to): self-driving cars (automotive industry), making the diagnosis of certain diseases more accurate (healthcare industry), product search recommendations (e-commerce & marketing), chatbots (customer service), robotics process automation (manufacturing), facial recognition (defense industry), and talent acquisition (corporations).

AI optimizes operations and resource allocation and improves the prediction and analysis of large datasets. At the same time, AI can also create new risks or negative consequences for individuals and society. AI technology can be misused and provide powerful tools for manipulative, exploitative, or social control practices. Therefore, the European Union (EU) Artificial Intelligence Act defined a risk-based framework that differentiates AI systems with unacceptable risk, high risk, or low or minimal risk, and defined minimum standards that AI systems should comply with.[2]

Common principles for the use of AI

The Artificial Intelligence Act is based on several principles.

- The use of AI technology must be in line with EU values and fundamental rights.[3] The charter of fundamental rights of the EU defines the universal values of human dignity, freedom, equality, and solidarity based on the principles of democracy and the rule of law. This contains principles on nondiscrimination and gender equality.
- AI technology must comply with existing General Data Protection Regulation (GDPR), the EU Data Governance Act,[4] the EU strategy for data,[5] and AI Liability Directive.[6]
- The Artificial Intelligence Act applies a risk-based approach by defining AI services that create unacceptable risk, high risk, and low or minimal risks.
- Transparency obligation for AI systems—i.e., AI systems shall be designed and developed in such a way that their operation is sufficiently transparent to enable users to interpret the systems’ output.
- AI systems shall have human oversight—high-risk AI systems shall be designed and developed in such a way that natural persons can effectively oversee them during the period in which the AI system is in use, intending to prevent or minimize the risks to health, safety, or fundamental rights.

Scope of the AI Directive

The Directive applies to AI systems placed on, put into service, or used in the European market. Therefore, the AI Directive also applies to producers and users that are established in a third country (outside the EU) to the extent that the output produced by AI systems are used in the EU.

AI systems exclusively used developed or used for military purposes are excluded from the scope of the Directive.

Prohibited AI services

The following AI services are forbidden:

- Those which materially distort a person's behavior in a manner that causes or is likely to cause that person or another person physical and/or psychological harm.
- Those which exploit vulnerabilities of a specific group of people due to their age or physical or mental disability.
- Those which evaluate or classify the trustworthiness of real people over time based on social behavior or personality characteristics. This includes the social score leading to detrimental treatment of those people or groups in ways unrelated to the context in which the data was collected or unjustified or disproportionate to social behavior.
- Those systems that use real-time remote biometric identification in publicly accessible spaces for law enforcement, unless they are targeted to search for potential victims of crime (including missing children), are in prevention of a specific, substantial, and imminent threat to the life (such as a terrorist attack), or are for the detection, localization, identification, or prosecution of suspect of criminal offense.
- Those which use real-time remote biometric identification in publicly accessible spaces for law enforcement need prior authorization by judicial authority.

High-risk AI services

AI systems identified as high risk are those that:

- Have significant harmful impact on health, safety, and fundamental rights of persons in the EU.
- Are safety components of products or systems.
- Are intended for remote biometric identification of natural persons (technical inaccuracies could lead to biased results and discriminatory effects regarding age, ethnicity, sex, or disabilities).
- Are to be used as safety components in management and operation of critical infrastructure (operation of road traffic, supply of water/gas, electricity, etc.).
- Are used for credit scoring or credit worthiness.
- Are used in employment (recruitment and selection of persons, for making decisions on promotions).
- Are used in migration, asylum, border control, or affect vulnerable people. The accuracy, nondiscriminatory nature, and transparency are important.

Requirements for high-risk AI systems

The Artificial Intelligence Act defines the following requirements for high-risk AI systems:

Risk management

- A risk management system shall be established, implemented, documented, and maintained concerning high-risk AI systems. The risk management measures shall be such that any residual risk associated with each hazard, as well as the overall residual risk of high-risk AI systems, is judged acceptable. The reduction of risks as far as possible should be eliminated or reduced through adequate design and development, and proper mitigation and control measures should be implemented for those risks that cannot be eliminated.

Data and data governance

- High-quality data is essential for the performance of many AI systems. Training, validation, and testing datasets should be sufficiently relevant, representative, and free of errors, complete in view of the intended purpose of the AI system.
- Training, validation, and testing datasets shall be subject to appropriate governance.

Technical documentation

- Before putting any AI system on the market, technical documentation shall be drawn up, and authorities should be allowed to evaluate the system.

Record keeping

- AI shall be designed and developed with logging capabilities to ensure a level of traceability of AI functioning throughout the lifecycle that is appropriate to the intended purpose of the AI system. The following things are required:
    - Recording of the period of each use of a system (start/end date and time)
    - Reference database against which input data has been checked
    - Input data for which search led to a match
    - Identification of natural persons involved in the verification of results

Transparency

- AI systems shall be designed and developed to ensure that their operation is sufficiently transparent to interpret system output and use it appropriately.
- AI systems shall be accompanied by instructions for use that include concise, complete, correct, and clear information that is relevant, accessible, and comprehensible to users.
- Identity and contact details of the provider or authorized representative must be accessible.
- Characteristics, capabilities, and limitations of performance of AI system, including intended purpose; the level of accuracy, robustness, and cybersecurity; any foreseeable circumstance related to use of AI system which may lead to risk to health and safety of fundamental rights; and its performance as regards to persons/group of persons on which the system is intended to be used.
- Expected lifetime and any necessary maintenance to ensure proper functioning of AI.

Human oversight

- AI systems shall be designed and developed so they can be overseen by natural persons when in use.
- Human oversight shall aim to prevent or at least minimize risks to health, safety, or fundamental rights.
- The measures shall enable individuals to which human oversight is assigned:
  - Fully understand the capacities and limitations of AI systems and monitor operation so that anomalies and unexpected performances can be detected and addressed as soon as possible.
  - Remain aware of automatically (over) relying on the output produced by high-risk AI systems (automation bias).
  - Be able to correctly interpret the AI system output, considering characteristics of the system and interpretation tools/methods available.
  - Be able to decide not to use an AI system or disregard, override, or reverse the AI system output.
  - Be able to intervene and or stop the AI system.

Accuracy, robustness, cybersecurity

- AI systems should be designed in such a way that throughout their lifecycle, they perform consistently on the above dimensions.
- Levels of accuracy shall be declared in the accompanying instructions of use.
- AI systems should be resilient as regards to attempts by unauthorized third parties.
- Technical solutions to ensure cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and risks. Technical solutions shall include measures to prevent and control for attacks to manipulate the training dataset and inputs designed to cause the model to make a mistake or model flaws.

Conclusion

AI can provide huge benefits for companies: it can significantly reduce the costs of operations and can improve the prediction and accuracy of analysis of large datasets. As such, AI is used in many applications and industry sectors.

However, in case the AI systems are not properly designed and tested to evaluate whether they are working as planned (e.g., the datasets used to validate the AI system is not representative or contains errors), then this could lead to discrimination of certain groups and/or could cause harmful impact on health, safety, or fundamental rights of persons in the EU.

Therefore, the Artificial Intelligence Act has forbidden AI services and has defined high-risk AI services and minimum expectations regarding human oversight, transparency, quality management system, technical documentation, risk management system, and data governance.

Takeaways

- AI systems placed on, put into service, or used in the European market should align with EU values and charter of EU fundamental rights.

- Certain AI systems are prohibited, given the potentially harmful implications and impact on natural persons, whereas others are defined as high risk.

[1] Oracle, " What is AI? Learn about Artificial Intelligence," last accessed December 1, 2022, https://www.oracle.com/artificial-intelligence/what-is-ai/.
[2] Future of Life Institute, "The Artificial Intelligence Act," last accessed December 1, 2022, https://artificialintelligenceact.eu.
[3] European Union, *Charter of Fundamental Rights of the European Union*, Oct. 26, 2012, 2012/C 326/02, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN.
[4] European Union, *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*, June 3, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868.
[5] European Union, *Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: A European strategy for data*, COM (2020) 66 final (Feb. 19, 2022), https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1582551099377&uri=CELEX:52020DC0066.
[6] European Commission, *Proposal for A Directive Of The European Parliament And Of The Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)*, COM (2022) 496 final (Sept. 28, 2022), https://ec.europa.eu/info/sites/default/files/1_1_197605_prop_dir_ai_en.pdf.