CENTER FOR DATA INNOVATION

datainnovation.org

Benjamin Mueller, Senior Analyst
bmueller@datainnovation.org

A Quick Explainer of the Artificial Intelligence Act

On April 21, 2021, the European Commission published a draft law to regulate artificial intelligence (AI) in the European Union. The Artificial Intelligence Act (AIA) is notable for its expansive definition of AI systems, and the imposition of extensive documentation, training, and monitoring requirements on AI tools that fall under its purview. Any company with EU market exposure that develops or wants to adopt machine-learning-based software will be affected by the AIA.

Who Would Be Subject to the AIA?

The AIA will apply extraterritorially to any provider or distributor of AI whose services or products reach the EU market. This includes providers and users of AI systems outside the EU if the output of the AI system is used in the EU. The AIA's impact will be widely felt across the economy. In particular, the AIA creates new regulatory obligations for AI tools used in financial services, education, employment and human resources, law enforcement, industrial AI, medical devices, the car industry, machinery, and toys.

What Would Count As AI?

The AIA defines AI broadly as a suite of software development frameworks that encompass machine learning, expert and logic systems, and Bayesian or statistical approaches. A software product featuring these approaches whose outputs "influence the environments they interact with" will be covered. The AIA distinguishes three categories of AI uses: prohibited AI uses, high-risk AI uses, and systems with limited risk.

Which Uses of AI Would Be Prohibited?

The Act explicitly bans AI systems that do any of the following:

1. Uses subliminal techniques to manipulate a person's behavior in a manner that may cause psychological or physical harm;
2. Exploits vulnerabilities of any group of people due to their age, physical, or mental disability in a manner that may cause psychological or physical harm;
3. Enables governments to use general-purpose "social credit scoring;"
4. Provides real-time remote biometric identification in publicly accessible spaces by law enforcement except in certain time-limited public safety scenarios.

Which Uses of AI Would Be Considered "High-Risk"?

The AIA considers an AI system high-risk if it is used as a safety component of a product, or if it is covered by one of 19 specified pieces of EU single market harmonization legislation (e.g., aviation, cars, medical devices). If the AI system is a component of a product covered by existing single market harmonization legislation, the product is already required to undergo a third-party conformity assessment. These mandatory third-party conformity checks will incorporate the AIA's requirements after the legislation is passed.

In addition, AI systems deployed in the following sectors are deemed to be high-risk to safety or fundamental rights:

- Critical infrastructure where the AI system could put people's life and health at risk;
- Educational and vocational settings where the AI system could determine access to education or professional training;
- Employment, worker management and self-employment;
- Essential private and public services, including access to financial services such as credit scoring systems;
- Law enforcement;
- Migration, asylum and border control, including verifying the authenticity of travel documents;
- The administration of justice.

Importantly, the Commission can expand this list through an administrative process without new legislation. The Commission is able to deem future AI products as high-risk to health, safety, and fundamental rights, as well as having the potential to affect a "plurality of persons" and the inability of end-users to opt-out of an adverse outcome.

What Requirements Would the AIA Impose on High-Risk Uses of AI?

To develop or use a high-risk AI system, an organization must meet a range of technical and regulatory requirements before the system can be brought to market. This includes establishing safeguards against various types of biases in data sets, using prescribed data governance and management practices, ensuring the ability to verify and trace back outputs throughout the system's life cycle, incorporating provisions for acceptable levels of transparency and understandability for users of the systems, and appropriate human oversight over the system generally. There are further ongoing compliance obligations once the system is in the market.

Conformity Assessments for High-Risk Uses

The AIA mandates an ex-ante conformity assessment for high-risk AI applications. In other words, AI systems—regardless of being products or services—in high-risk sectors need to be compliant with the AIA's obligations before they are placed on the EU market.

For AI products and services governed by existing product safety legislation—such as cars, aviation, machinery, medical devices and toys—the Act's requirements will fall under the existing third-party conformity assessment structures and regulatory frameworks that already apply. In general, whichever supervisory body or legislation is responsible for the business that provides a regulated AI service will oversee the AIA. For instance, a financial services company wishing to use AI tools for credit risk assessment will continue to be overseen by the competent financial supervisory authorities as per the existing setup of the EU single market rulebook.

Providers of AI tools not governed by explicit regulatory frameworks will conduct their own conformity assessment and have to file their system in a newly established EU-wide database for high-risk AI systems.

Technical and Auditing Requirements for High-Risk AI

The requirements of the Act for high-risk AI systems are:

- Creating and maintaining a risk management system for the entire lifecycle of the system;
- Testing the system to identify risks and determine appropriate mitigation measures, and to validate that the system runs consistently for the intended purpose, with tests made against prior metrics and validated against probabilistic thresholds;
- Establishing appropriate data governance controls, including the requirement that all training, validation, and testing datasets be complete, error-free, and representative;
- Detailed technical documentation, including around system architecture, algorithmic design, and model specifications;
- Automatic logging of events while the system is running, with the recording conforming to recognized standards;
- Designed with sufficient transparency to allow users to interpret the system's output;
- Designed to maintain human oversight at all times and prevent or minimize risks to health and safety or fundamental rights, including an override or off-switch capability.

Most of the Act's regulatory obligations fall on the party that places the system on the market (the "provider"), which can be a third-party provider or a company developing the AI itself. Distributors, importers, users and other third parties are subject to provider obligations if they place a high-risk AI system on the market under their name or make a substantial modification to it. This relieves the original provider of responsibility. Distributors and importers have various verification obligations before making a high-risk AI system available on the market. The Act further mandates that "users" (the entity employing the high-risk AI system) deploy the system correctly, ensure the input data is of high quality, and monitor the system's performance on an ongoing basis with specific logging and audit requirements. Users need to put in place a risk management system to ensure that all risks associated with the AI system are documented and mitigated. Furthermore, if an AI system is used to assist the company with interacting with its customers, then certain transparency duties apply.

Post-Market Monitoring for High-Risk AI

The Act creates mandatory post-market monitoring obligations for high-risk systems. Serious incidents or faults of the AI system which breach safety laws or fundamental rights must be reported to the national supervisory body. In case of a violation of the Act, regulators can mandate access to the source code of a high-risk AI system. High-risk systems that violate the Act can be forcibly withdrawn from the market by the regulator.

What About Limited-Risk AI Systems?

Certain limited-risk systems are covered by the Act under transparency requirements. AI systems that interact with people face similar obligations to GDPR— notifying users they are interacting with an AI system, what personal data it is collecting and for what purpose, and if users are classified into specific categories like gender, age, ethnic origin, or sexual orientation. This does not apply if it is "obvious from the circumstances and the context of use" that someone is interacting with an AI system. The Act further imposes a disclosure obligation for deep fakes, except when used for artistic or satirical purposes.

All non-high-risk AI systems have to comply with existing product safety legislation and preserve the fundamental rights of EU citizens.

Who Will Oversee the Act's Implementation and Enforcement?

The AIA creates a European AI Board, composed of representatives of member states and the Commission. The Act relies on member state regulators for enforcement and sanctions. This structure mimics that of the GDPR, except that the Board is chaired by the Commission. The Board can issue opinions, recommendations, and written contributions on "matters related to the implementation of this regulation." The Board may invite external experts and observers to attend meetings and can hold exchanges with interested third parties.

What Are the Penalties For Violations?

As with GDPR, these rules apply extraterritorially to providers and users outside of the EU if the output of the system is used in the EU. Non-compliance with prohibited uses and data governance obligations is punishable with a fine of up to €30M or 6 percent of worldwide annual turnover (whichever is higher); for high-risk AI applications, the ceiling is €20M or 4 percent of turnover. The supply of incorrect, incomplete, or misleading information to national competent bodies is subject to a fine of up to €10M or 2 percent of turnover.

What Happens Next?

The Act is now working its way through the European Parliament, it has not yet been assigned to a specific committee. In addition, it will be subject to scrutiny by the Council of Ministers. After it is passed, it is subject to a two-year implementation period; the AIA exempts AI systems existing at implementation from meeting these requirements unless they subsequently experience a significant change in purpose or design.