EU's Artificial Intelligence Act: A Summary and How ZeroTrusted.ai Can Help Organizations Comply

Waylon Krush
CEO of ZeroTrusted.ai
October 17, 2024

ZER0TRUSTED.AI

INTRODUCTION

The EU AI Act is a regulation that intends to improve the functioning of the internal market by creating a uniform legal framework for AI system development, placement on the market, putting into service, and use in the Union. The purpose is to promote the use of human-centric and trustworthy AI while maintaining a high level of protection of public interests like health and safety, and protection of fundamental rights. The regulation aims to facilitate international convergence by aligning the notion of "AI system" with the work of international organizations working on AI. It also aims to provide flexibility to accommodate the rapid technological advancements in the field of AI.

Risk-Based Approach

The EU AI Act employs a risk-based approach to regulate AI systems. This means that the requirements of the law depend on the level of risk the AI system poses. The AI Act uses the following categories:

- **Unacceptable Risk:** Certain AI practices are banned entirely due to their unacceptable risk to fundamental rights. These AI systems include those designed or used in a way that manipulates human behavior to circumvent users' free will (for example, by exploiting vulnerabilities of specific groups) or that allows social scoring.
- **High Risk:** AI systems that pose a high risk to health, safety, or fundamental rights are subject to strict requirements. These include risk management systems, data quality governance, technical documentation, automatic logging, transparency requirements, and conformity assessments.
- **Limited Risk:** AI systems with limited risk are subject to specific transparency requirements, like informing individuals that they are interacting with an AI system.
- **Minimal Risk:** AI systems with minimal risk are not subject to additional legal obligations.

High-Risk AI Systems: Specific Requirements

For providers placing high-risk AI systems on the market or putting them into service, the following requirements apply:

- Establish a **risk management system**, a continuous process running through the entire lifecycle of a high-risk AI system. It includes identifying and analyzing known and foreseeable risks, estimating and evaluating those risks, and implementing risk control measures to mitigate and minimize those risks. The risk management system should also be regularly updated and reviewed.

- The development of high-risk AI systems must be based on training, validation, and testing datasets that meet quality criteria. This includes data relevance, representativeness, accuracy, completeness, and the absence of bias. **Data governance** measures must be implemented to ensure this.
- Providers must create and maintain **technical documentation** that outlines the specifications and characteristics of their AI systems. This documentation should allow authorities to assess whether the AI system complies with legal requirements. It should include information like general information about the AI system, design specifications, and information related to the data and data governance used to train the model.
- Keep **automatic logs** generated by the high-risk AI systems. These logs help to monitor the operation of the AI system, identify potential issues, and ensure accountability. The logs should be kept for a period appropriate to the intended purpose of the system or at least 6 months.
- Providers must fulfill **transparency and information** requirements. This includes providing clear information to users about how the AI system works and its limitations. High-risk AI systems should also be accompanied by clear instructions for use.
- Undergo **conformity assessments** before the AI system can be placed on the market. This process aims to ensure that the AI system meets the requirements laid out in the AI Act.
- **Register** the high-risk AI system in the EU database before it can be placed on the market or put into service. This database will contain key information about the AI system, allowing authorities to track high-risk AI systems.

For deployers using high-risk AI systems, the following requirements apply:

- Use the high-risk AI systems in accordance with the provider's **instructions for use**. This ensures that the system is used as intended and helps to minimize the risks associated with its use.
- Establish and maintain a **monitoring system** to detect incidents or anomalies that might indicate problems with the AI system. The monitoring system should be designed to capture and analyze relevant data that could reveal issues with the AI system’s operation.

- Keep **automatic logs** of events generated by the AI system. This helps track the AI system's behavior, identify potential problems, and ensure accountability.
- If using a high-risk AI system in the workplace, deployers must inform **workers and their representatives** about its use. This obligation ensures transparency and allows workers to understand how the AI system might affect their work.
- Conduct a **fundamental rights impact assessment** before deploying certain high-risk systems. This assessment aims to identify and mitigate potential negative impacts of the AI system on fundamental rights.

How ZeroTrusted.ai can help organizations meet the requirements of the EU AI Act:

- **Risk Management:** ZeroTrusted.ai can aid in developing and implementing a risk management system, as required by Article 9.
- **Data Governance:** The platform helps with data quality control and governance, ensuring compliance with Article 10.
- **Documentation and Logging:** ZeroTrusted.ai can streamline the process of maintaining technical documentation and automatic logs required by Articles 11 and 12.
- **Transparency and Explainability:** The platform enhances transparency by providing tools for model explainability and documentation, helping to meet the requirements of Article 13 and Article 50.
- **Conformity Assessments:** ZeroTrusted.ai assists in preparing for and managing conformity assessments via our real-time AI monitoring and AI HealthCheck, facilitating compliance with Article 43.
- **Registration and Reporting**: The platform supports the registration of high-risk AI systems in the EU database and aids in reporting serious incidents as outlined in Articles 49 and 73.
- **Copyright Compliance:** ZeroTrusted.ai helps organizations to implement a copyright compliance policy and provides plagiarism detection.
- **Cybersecurity:** The platform employs robust cybersecurity measures, such as encryption, to protect AI systems from unauthorized access and data breaches, anonymizes sensitive data and tokens.

General-Purpose AI Models

General-Purpose AI (GPAI) models are AI systems with broad applicability across various domains. Examples are Large Language Models or image recognition models that can be used for multiple purposes. They are distinguished from AI systems that have a specific intended purpose.

Providers of general-purpose AI models have the following obligations:

- Prepare and maintain **technical documentation** for the model, including the training and testing processes and evaluation results. This documentation provides information about the model's capabilities, limitations, and potential risks.
- Provide information and documentation to **downstream providers** who intend to integrate the general-purpose AI model into their own AI systems. This enables downstream providers to comply with their obligations under the AI Act.
- Implement a **copyright compliance policy**, especially to identify and comply with a reservation of rights expressed pursuant to Article 4(3) of Directive (EU) 2019/790. This involves taking steps to ensure that the training data used for the model does not infringe copyright law.
- Draw up and make publicly available a detailed **summary** of the content used for training the general-purpose AI model. This summary allows stakeholders, including copyright holders, to assess the potential risks associated with the model's training data.

General-purpose AI models with systemic risks:

- In addition to the general requirements for GPAI models, **providers of general-purpose AI models with systemic risks** are subject to additional obligations, like **model evaluation** and **mitigation of systemic risks**. Systemic risks refer to the potential for the model to have a significant negative impact on society, the economy, or fundamental rights.
- These models are presumed to be high-impact and the provider must notify the AI Office at the latest two weeks after the requirements for high-impact are met.

Enforcement and Governance

- **AI Office**: The AI Office plays a central role in the enforcement and governance of the AI Act. It supports the implementation and enforcement of the AI Act as regards general-purpose AI models and systems, and provides advice on the classification of general-purpose AI models. It will have expertise in AI to monitor and supervise compliance with the regulation.

- **National competent authorities:** Each member state will designate national competent authorities to supervise and enforce the AI Act at the national level. They will be responsible for investigating potential violations and imposing penalties.

Conclusion

The EU AI Act introduces a comprehensive framework for regulating AI systems within the EU. The regulation aims to ensure that AI systems are developed and used safely, ethically, and in a way that respects fundamental rights. The risk-based approach and the specific requirements for high-risk AI systems and GPAI models aim to address the potential risks posed by different types of AI.

ZeroTrusted.ai can assist organizations in meeting the requirements of the EU AI Act by providing tools and features that support risk management, data governance, documentation, transparency, conformity assessments, registration, and reporting. The platform's capabilities can help organizations to develop, deploy, and use AI systems in a responsible and compliant manner, while leveraging the benefits of this transformative technology.

Written by

Waylon Krush, CISSP, CISA, CGRC, CEO of ZeroTrusted.AI

Waylon Krush is the CEO of ZeroTrusted.ai, a pioneering tech firm focused on fortifying the security, privacy, and reliability of AI, ML, and related systems. With over 27 years of experience across industry and government, Mr. Krush has established himself as a visionary leader in cybersecurity, particularly within the rapidly evolving fields of AI risk management and ML security. Under his leadership, ZeroTrusted.ai is not only enhancing the safety of AI-driven technologies but also developing innovative solutions to counter the growing lack of AI security and accountability. Mr. Krush's expertise and forward-thinking approach make him a trusted authority on the technical and ethical complexities that shape the future of AI and cybersecurity.