

The Role of Synthetic Data within the European Artificial Intelligence Act

Simona Mazzarino



Who I am



Master's degree in Digital
Humanities - Language
Technologies at University of Pisa



Data Scientist at Clearbox AI



Outside of work, I enjoy traveling and
diving into mystery, thriller, and true
crime books.



[Simona Mazzarino](#)



[Simona Mazzarino](#)

Introduction to the AI Act

What is the AI Act?

The Artificial Intelligence Act (AI Act) is a proposed regulation by the **European Union** aimed at establishing a comprehensive legal framework to ensure the safe and ethical development, deployment, and use of AI technologies.



**PRESS
EN**

PRESS RELEASE
409/24
21/05/2024

Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI

Today the Council approved a ground-breaking law aiming to harmonise rules on artificial intelligence, the so-called **artificial intelligence act**. The flagship legislation follows a **'risk-based' approach**, which means the **higher the risk** to cause harm to society, **the stricter the rules**. It is the first of its kind in the world and can set a global standard for AI regulation.

The new law aims to foster the development and uptake of **safe and trustworthy** AI systems across the EU's single market by both private and public actors. At the same time, it aims to ensure respect of **fundamental rights** of EU citizens and stimulate investment and innovation on artificial intelligence in Europe. The AI act applies only to areas within EU law and provides exemptions such as for systems used exclusively for **military and defence** as well as for **research** purposes.

”

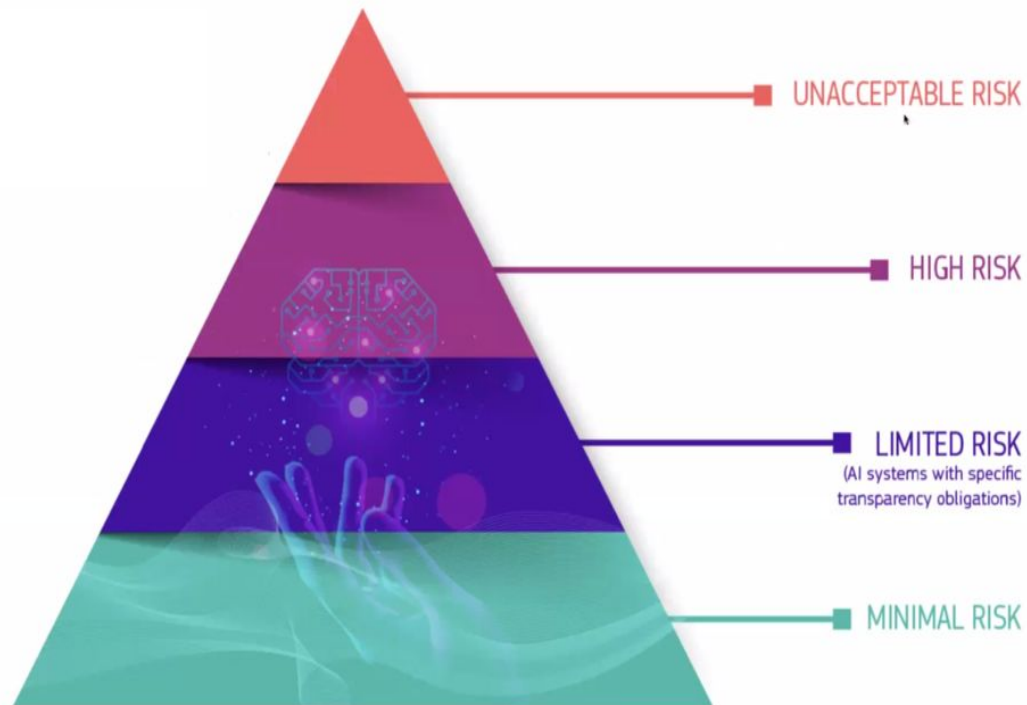
The adoption of the AI act is a significant milestone for the European Union. This landmark law, the first of its kind in the world, addresses a global technological challenge that also creates opportunities for our societies and economies. With the AI act, Europe emphasizes the importance of trust, transparency and accountability when dealing with new technologies while at the same time ensuring this fast-changing technology can flourish and boost European innovation.

Mathieu Michel, Belgian secretary of state for digitisation, administrative simplification, privacy protection, and the building regulation

Definition of AI system within the AI Act

‘AI system’ means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments (Art. 3).

A risk based approach



Requirements for High-Risk AI Systems



The Role of Synthetic Data

Synthetic Data within the AI Act

Synthetic Data is mentioned several times within the AI Act:

Art. 10 (5)(a)

(...) the bias detection and correction cannot be effectively fulfilled by processing other data, including **synthetic** or anonymised data;

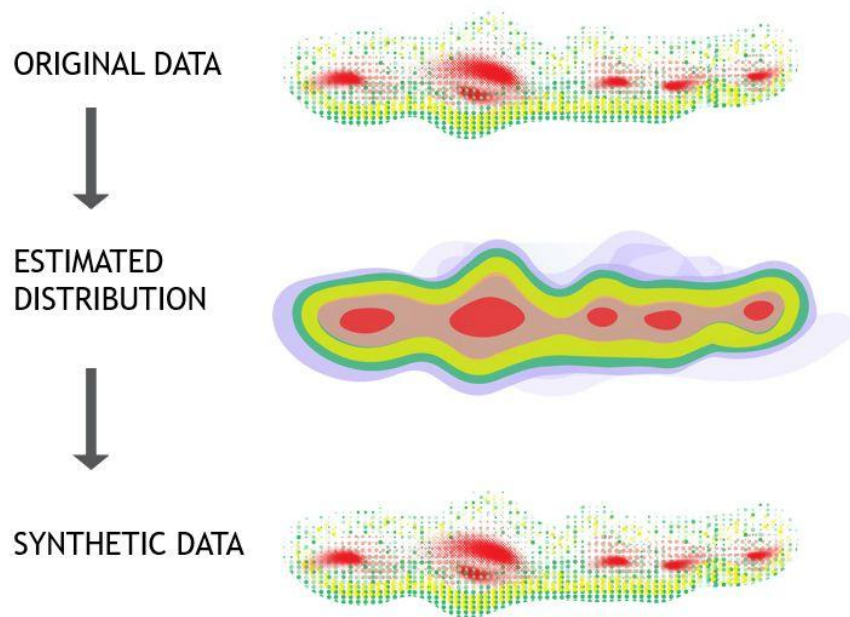
Art. 50 (2)

Providers of AI systems, (...) generating **synthetic** audio, image, video or text content, shall ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated.

Art. 59 (1)(b)

(...) the data processed are necessary for complying with one or more of the requirements referred to in Chapter III, Section 2 where those requirements cannot effectively be fulfilled by processing anonymised, **synthetic** or other non-personal data;

What is Synthetic Data?



- **Synthetic data** is artificially generated using artificial intelligence algorithms on real data samples.
- They possess the **same statistical properties** and **predictive capabilities** as the real data from which they were generated.

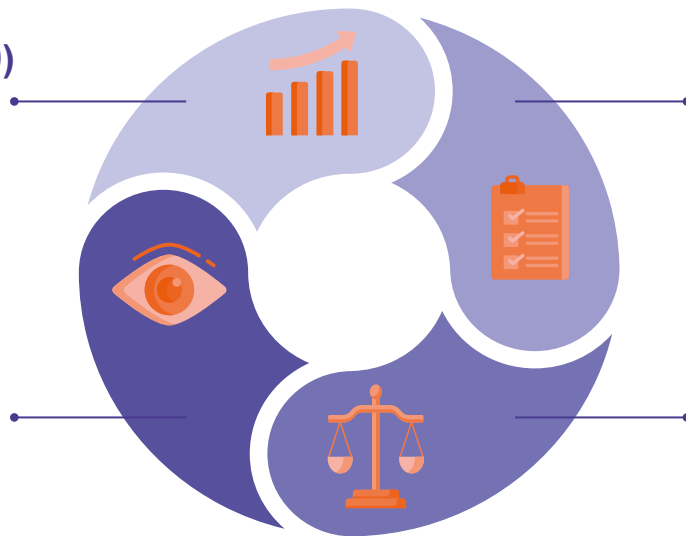
Advantages of Synthetic Data within the AI Act

**ENHANCING DATA QUALITY
AND GOVERNANCE (ART. 10)**

**MITIGATING BIAS AND
PROMOTING FAIRNESS
(ART. 10)**

**ENSURING PRIVACY AND
DATA PROTECTION (ART. 10)**

**ENHANCING MODEL ACCURACY
AND ROBUSTNESS (ART. 15)**



Real Use Cases

Financial Sector: Credit Scoring

Scenario:

A bank uses an AI-based credit scoring system to assess loan applications. Traditional datasets can be incomplete and can include biases, such as underrepresentation of certain socioeconomic or demographic groups.



Image Source: <https://www.tdecu.org/blog/what-no-credit-score-means>

Scenario Source: Namperumal, G., Selvaraj, A., & Surampudi, Y. (2022). Synthetic Data Generation for Credit Scoring Models: Leveraging AI and Machine Learning to Improve Predictive Accuracy and Reduce Bias in Financial Services. *Journal of Artificial Intelligence Research*, 2(1), 168-204. (<https://thesciencebrigade.com/JAIR/article/view/375>)

Healthcare Sector: Patient Diagnostics

Scenario:

AI systems in healthcare use patient data to develop diagnostic tools, such as predicting diseases or identifying risk factors. However, using real patient data can raise significant privacy concerns and compliance challenges, especially in light of GDPR and HIPAA regulations.



Image Source: <https://online.stanford.edu/programs/artificial-intelligence-healthcare>

Scenario Sources:

Giuffrè, M., & Shung, D. L. (2023). Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *NPJ digital medicine*, 6(1), 186.

(<https://www.nature.com/articles/s41746-023-00927-3>)

Hernandez, M., Epelde, G., Alberdi, A., Cilla, R., & Rankin, D. (2022). Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 493, 28-45.

(<https://www.sciencedirect.com/science/article/abs/pii/S0925231222004349>)

Conclusion

The Role of Synthetic Data within AI Act

Synthetic data plays a crucial role in achieving compliance with the AI Act, and its explicit mention in the AI Act reflects the growing recognition of this technology's potential to shape a future where AI innovation aligns with ethical principles and fosters societal trust.





Thanks for listening

Feel free to contact us:



www.clearbox.ai



simona@clearbox.ai
info@clearbox.ai



[Simona Mazzarino](#)
[Clearbox AI](#)