EU AI Act

Report on quality and use of data

April 2024

Mobilisation of data assets to support objectives

This presentation examines data from ESMA's 2023 report on quality and use of data to provide insights relating to how data assets may be mobilised to support the EU AI Act's strategic and thematic objectives. Salient aspects are as follows:

Ensuring High-Quality Data for AI Systems
The EU AI Act mandates that high-risk AI systems be developed based on training, validation, and testing data sets that meet specific quality criteria. This includes ensuring that data sets are relevant, representative, free of errors, and complete for their intended purpose.

Data Governance and Management Practices
The Act emphasizes the importance of appropriate data governance and management practices. These practices cover the design choices, data collection processes, data preparation operations, and the assessment of data availability and suitability.

Facilitating Access to High-Quality Data Sets
The Act acknowledges the role of European common data spaces and the facilitation of data sharing between businesses and governments in providing access to high-quality data for AI training, validation, and testing.

Data is defined through several key terms:

1. **Training Data**: Data used for training an AI system by fitting its learnable parameters;

2. **Validation Data**: Data utilized for evaluating the trained AI system and for tuning its non-learnable parameters and learning process to prevent underfitting or overfitting;

3. **Testing Data**: Data employed for providing an independent evaluation of the AI system to confirm its expected performance before it is placed on the market or put into service; and

4. **Input Data**: Data provided to or directly acquired by an AI system, based on which the system produces an output.

Data: *"Data, in the context of the EU AI Act, encompasses various forms and uses within AI systems."*

Under the EU AI Act, the articles that specifically cover aspects related to data, including its governance, handling, and requirements for high-risk AI systems, are found in Articles 10, 59, and 71.

Linking objectives with use and quality of reported data

Ensuring Safety and Performance

High quality data plays a vital role in ensuring that AI systems, particularly high-risk ones, perform as intended and safely. The EU AI Act mandates that training, validation, and testing data sets meet specific quality criteria to prevent these systems from becoming sources of discrimination prohibited by Union Law.

Preventing Discrimination

The EU AI Act requires that data sets used for training, validation, and testing of high-risk AI systems are relevant, sufficiently representative, and as free of errors and complete as possible. This is to mitigate possible biases in the data sets that are likely to affect the health and safety of persons.

Data Governance and Management

High-risk AI systems adhere to appropriate data governance and management practices. These practices should cover relevant design choices, data collection processes, data preparation operations, and the assessment of data availability, quantity, and suitability.
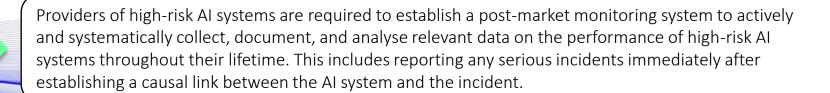
Data Protection Compliance

In facilitating compliance with Union data protection, such as Regulation (EU) 2016/679, data governance and management practices should include transparency about the original purposes of the data collection, ensuring that processing of data for AI systems respects the right to privacy.

Reporting plays a key role in risk identification and monitoring

Post-Market Monitoring and Reporting of Serious Incidents

Providers of high-risk AI systems are required to establish a post-market monitoring system to actively and systematically collect, document, and analyse relevant data on the performance of high-risk AI systems throughout their lifetime. This includes reporting any serious incidents immediately after establishing a causal link between the AI system and the incident.

Market Surveillance and Control

The Act empowers market surveillance authorities to conduct surveillance and control of AI systems within the Union market. These authorities have access to documentation, training, validation, and testing data sets used for the development of high-risk AI systems. This access facilitates the assessment of AI systems' conformity with the Act's requirements, ensuring the orderly functioning of the AI market.

Cooperation and Information Sharing

The Act promotes cooperation and information sharing among market surveillance authorities, the Commission, and other relevant bodies. This includes the provision for joint activities and investigations aimed at promoting compliance, identifying non-compliance, and raising awareness about the regulation.
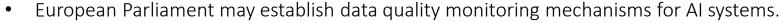
Providing a common approach to monitor data quality over time

EU AI Act Objectives Alignment

- European Parliament may establish data quality monitoring mechanisms for AI systems.
- Monitoring depth to correlate with system complexity and legislative responsibilities.

Unified Risk-Based Methodology

- Develop across various data quality frameworks within the Act.
- Offer consistency, facilitate issue identification, enable effective follow-up, and ensure continuous oversight.

Enhanced Data Quality Through Collaboration

- Industry stakeholders, national authorities, and the European Parliament to collaborate.
- Notable improvements to be observed; identify opportunities for further refinement.

Commitment to Ongoing Surveillance

- European Parliament to remain dedicated to continuous monitoring of data quality.
- Addresses emerging signs of degradation.

Overview of data quality indicators and their methodology

Name	Area	Scope	Objective
Relevance and representativeness	High-risk AI systems development	Training, validation, and testing data sets	To ensure data sets are relevant to the AI system's intended purpose and represent the diversity of scenarios and individuals the system will encounter, thereby reducing biases and enhancing the system's performance and fairness.
Error rate and completeness	Data set integrity	Training, validation, and testing data sets	To minimize errors and ensure data sets are as complete as possible, enhancing the AI system's accuracy and reliability in real-world applications.
Bias detection and mitigation	Fairness and non-discrimination	Data preparation and processing operations	To identify and mitigate biases in data sets that could lead to discrimination or negative impacts on fundamental rights, ensuring the AI system's outputs are equitable and do not perpetuate existing inequalities.
Data set documentation and transparency	Transparency and accountability	Technical documentation of AI systems	To provide comprehensive documentation of data sets, including their origin, selection process, and preparation methods, enhancing the transparency of AI system development and facilitating accountability.
Quality management system accuracy	System design and development	Quality management system for the design, development, and testing of AI systems	To ensure the quality management system is adequate and effective in overseeing the development of high-risk AI systems, maintaining the system's compliance with regulatory requirements over time.

*All data sourced from ESMA (2024)

Thank You!

Disclaimer

This Presentation may contain information, text, data, graphics, photographs, videos, sound recordings, illustrations, artwork, names, logos, trade marks, service marks, and information about us, our lines of services, and general information may be provided in the form of documents, podcasts or via an RSS feed ("the Information").

Except where it is otherwise expressly stated, the Information is not intended to, nor does it, constitute legal, accounting, business, financial, tax or other professional advice or services. The Information is provided on an information basis only and should not be relied upon. If you need advice or services on a specific matter, please contact us using the contact details for the relevant consultant or fee earner found on the Presentation.

The Presentation and Information is provided "AS IS" and on an "AS AVAILABLE" basis and we do not guarantee the accuracy, timeliness, completeness, performance or fitness for a particular purpose of the Presentation or any of the Information. We have tried to ensure that all Information provided on the Presentation is correct at the time of publication. No responsibility is accepted by or on behalf of us for any errors, omissions, or inaccurate information on the Presentation. Further, we do not warrant that the Presentation or any of the Information will be uninterrupted or error-free or that any defects will be corrected.

Although we attempt to ensure that the Information contained in this Presentation is accurate and up-to-date, we accept no liability for the results of any action taken on the basis of the Information it contains and all implied warranties, including, but not limited to, the implied warranties of satisfactory quality, fitness for a particular purpose, non-infringement, compatibility, security, and accuracy are excluded from these Terms to the extent that they may be excluded as a matter of law.

In no event will we be liable for any loss, including, without limitation, indirect or consequential loss, or any damages arising from loss of use, data or profits, whether in contract, tort or otherwise, arising out of, or in connection with the use of this Presentation or any of the Information.