Journal-Law and Economy 9[15]1-8: e2915108
Article

AI regulation in relation to training databases

La regulación de la IA en materia de Bases de Datos de Entrenamiento

Orozco-Orozco, José Zócimo * [a]

a ROR Universidad de Guadalajara • S-5977-2018 • 0000-0001-5888-0627 • 247561

Classification:

Area: Social Sciences
Field: Social Sciences
Discipline: Legal Sciences
Subdiscipline: Other Legal Sciences

https://doi.org/10.35429/JLE.2025.9.15.2.1.8
History of the article:
Received: September 30, 2025
Accepted: November 30, 2025

* ✉ [zocimo.orozco@academicos.udg.mx

Abstract

AI systems are increasingly advancing technologically, while their regulation lags behind in terms of the risks AI can pose to people´s fundamental rights. We will start with regulation [EU] 2024/1689 of the European Parliament and of the Council of june 13,2024. Which is a leading legislation on AI systems. The regulation includes the implementation, introduction to market and public introduction, monitoring, reporting documentation and other technical requirements specific to this systems.

AI Regulation Concerning Training Databases		
OBJECTIVES	**METHODOLOGY**	**CONTRIBUTION**
To investigate the regulation of artificial intelligence in relation to training databases. To analyze the legal protection of data used for AI training. To study the proposals that would allow AI models to use training data in Mexico.	The methodology is qualitative and documentary. It includes the problem statement, defining the research objectives, the justification, and the context. Data collection was documentary. The conclusions interpret the current context of AI training databases.	The legal protection of AI training databases was investigated. The protection of fundamental rights in the regulatory treatment of AI training databases was promoted

Databases, AI providers, High-Risk AI Systems

Resumen

Los sistemas de IA cada vez presentan más avances tecnológicos, en tanto, su regulación sufre un retraso en cuanto a los riesgos que la IA puede representar a los derechos fundamentales de las personas. Se empezará por comentar el Reglamento UE 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024, el cual es una legislación pionera en la regulación de los sistemas de IA para la puesta en marcha, introducción al mercado y al público, la vigilancia, la elaboración de reportes, llevar a cabo la documentación y otros requerimientos técnicos propios de estos sistemas.

La regulación de la IA en materia de Bases de Datos de Entrenamiento		
OBJETIVOS	**METODOLOGÍA**	**CONTRIBUCIÓN**
Investigar la regulación de la IA de Bases de Datos deEntrenamiento. Analizar la protección legal de los datos para el entrenamiento de la IA. Estudiar las propuestas para que los modelos de IA usen datos de entrenamiento en México.	La metodología es de tipo cualitativo y documental. Incluye el planteamiento del problema, definir los objetivos de investigación, la justificación y el contexto. La recolección de datos documental.	Se investigó la protección legal de las bases de dato de entrenamiento en materia de IA. Se promovió la protección de los derechos fundamentales para la regulación de las bases de datos de entrenamiento de la IA.

Bases de Datos, Proveedores de IA, Sistemas de IA de Alto Riesgo.

Area: Promotion of frontier research and basic science in all fields of knowledge

Citation: Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.

Introduction

This article deals with the regulation of AI, more precisely in terms of training databases, obligations of AI system providers, and finally, we will look at the draft bill to regulate AI in Mexico.

EU Regulation 2024/1689 of the European Parliament and of the Council of 13 June 2024.The regulation of AI in terms of training databases.

Regulation [EU] 2024/1689 of the European Parliament and of the Council [2024].

The regulation aims to promote the development and adoption of safe and reliable Artificial Intelligence [AI] systems in the European Union, in both the private and public sectors, to ensure the health and safety of Union citizens.

67. High-quality data and access to high-quality data are necessary to provide structure and ensure the functioning of many AI systems, especially when model training techniques are used to ensure that high-risk AI systems function as intended and safely and do not become a source of any form of discrimination prohibited by Union law.

Data management and governance practices should be put in place to ensure that the data sets for training, validation and testing are of high quality.

The data sets should be relevant and sufficiently representative and, as far as possible, free from errors and complete in accordance with the intended purpose of the system.
In order to comply with European Union law on data protection, including Regulation [EU] 2016/679, data management and governance practices must include transparency regarding the original purpose of data collection.

Data sets must have the appropriate statistical properties in relation to individuals or groups of individuals in relation to the intended use of the high-risk AI system, seeking to mitigate possible biases in data sets that could affect the health and safety of individuals, possible negative impacts on fundamental rights or give rise to any form of discrimination prohibited by Union law...

The requirement that data sets be as complete and error-free as possible should not affect the use of techniques for protecting sensitive data to the extent required by their intended purpose, features, characteristics or particular elements of the geographical, contextual, behavioural or functional environment in which AI systems are intended to be used.

Governance requirements may be met by using third parties that offer compliance services, including verification of data governance, data set integrity and data training, validation and testing practices, to the extent that compliance with the data requirements of this Regulation is ensured.

68. In order to develop and evaluate high-risk AI systems, providers and other relevant bodies and entities such as European digital innovation hubs, testing and experimentation facilities and researchers must have access to high-quality data sets in their areas of activity related to this Regulation and must be able to use them.

The European common data spaces established by the Commission and the facilitation of data exchange between businesses and with governments in the public interest are essential to provide reliability and accountability in access to AI systems, and in a non-discriminatory manner, to high-quality data with which to train, validate and test AI systems.

The relevant competent authorities, including sectoral authorities that provide or facilitate access to data, can also support the provision of high-quality data with which to train, validate and test AI systems.

69. Measures taken by providers to ensure the right to privacy and the protection of personal data may include not only anonymisation and encryption, but also the use of technology that allows algorithms to access data and train AI systems...

76. Cyberattacks against AI systems may target specific AI assets such as training datasets. Providers of high-risk AI systems should take appropriate measures such as security controls, taking into account the underlying ICT infrastructure where appropriate.

ISSN: 2524-2113
RENIECYT: 1702902

Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.
https://doi.org/10.35429/JLE.2025.9.15.2.1.8

104. Providers of licensed and open-source AI models whose parameters, including weights, information on the model architecture and information on the use of the model, are made publicly available should be subject to exemptions from transparency requirements, unless they are considered a systemic risk. in which case the fact that the model is transparent and accompanied by an open-source licence should not be considered sufficient reason for it to be exempt from compliance with the obligations of this Regulation.

Where the disclosure of general-purpose AI models under a free and open-source licence does not necessarily reveal information about the dataset used to train the model or make adjustments, nor about how compliance with copyright law was ensured, the above exception should not exempt from the obligation to provide a summary of the content used for training the model or the obligation to adopt guidelines for compliance with Union copyright law, in particular to identify and respect the reservation of rights provided for in Article 4[3] of Directive [EU] 2019/790 of the European Parliament and of the Council.

109. AI model providers who develop or use models for professional or scientific research purposes should be exempted from compliance with the applicable obligations; compliance with these obligations should be voluntary for these providers.

Compliance with these obligations should not entail excessive costs or discourage the use of such models.

Section 2

Requirements for high-risk AI systems

Article 8

2. Providers shall be responsible for ensuring that their product complies with all applicable requirements of Union legislative acts.

Article 9

A risk management system shall be implemented, documented and maintained in relation to high-risk AI systems.

3. The risks referred to in this Article are only those that can be reasonably mitigated or eliminated through the development or design of the high-risk AI system.

Article 10

Data and data governance

1. High-risk AI systems developed using AI model training techniques with data shall be developed from a data set, validation and testing that meet the quality criteria referred to in paragraphs 2 to 5, provided that such data sets are used.

2. Governance practices for the training, validation and testing data set shall focus on the following:

a] Design decisions, b] Data collection and sourcing processes, c] Data preparation operations, including cleansing, updating, enrichment and aggregation, d] Formulation of assumptions for measuring and representing data, e] Assessment of the availability, quantity and adequacy of datasets, f] Biases that may affect the health and safety of individuals, that may adversely affect fundamental rights or generate any type of discrimination prohibited by Union law, g] Effective measures to detect, prevent and mitigate biases detected as established in the previous paragraph, h] Detection of deficiencies that prevent compliance with this regulation.

3. The training, validation and testing data sets shall be relevant, as representative as possible and free from errors, and shall be complete for their intended purpose, either as individual data sets or in combination.

Article 11. Technical documentation

The technical documentation for a high-risk AI system shall be drawn up before it is placed on the market or put into service and shall be kept up to date.

The documentation shall be drafted in such a way as to demonstrate that the high-risk AI system complies with the requirements of this section of this Regulation and shall provide the authorities with the information necessary to assess the high-risk AI system.

ISSN: 2524-2113
RENIECYT: 1702902

Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.
https://doi.org/10.35429/JLE.2025.9.15.2.1.8

Article 12. Record keeping

High-risk AI systems shall technically enable the automatic recording of events throughout the entire life cycle of the system.

Article 13

1. High-risk AI systems shall be designed and developed with a sufficient level of transparency to enable the correct use of their output results.

2. High-risk AI systems shall be accompanied by instructions for use in digital format.

The instructions shall contain at least the following information:

a]. The identity and contact details of the supplier and its designated representative b] the characteristics, capabilities and limitations of the functioning of the high-risk AI system, including i] intended purpose ii] the level of accuracy, robustness and cybersecurity iii] any known or foreseeable circumstances associated with the high-risk AI system in accordance with its intended purpose that could give rise to risks to health, safety or fundamental rights. iv] the capabilities and technical characteristics of the high-risk AI system to provide information explaining its output results] changes to the high-risk AI system and its operation predetermined by the supplier at the time of the initial assessment. d] human oversight measures, including technical measures to facilitate the interpretation of output results from high-risk AI systems. e] the necessary IT and hardware resources, the expected lifespan of the high-risk AI system, and the necessary maintenance and care measures [including their frequency to ensure the proper functioning of the system, including software updates]. f] where possible, a description of the mechanisms that enable the correct deployment, storage, and interpretation of log files.

Article 14 Human supervision

High-risk AI systems shall be developed in such a way that they can be supervised by natural persons during the period they are in use.

The purpose of human supervision is to prevent or reduce risks to fundamental rights arising when a high-risk AI system is used or when it is misused in a reasonably foreseeable manner.

3. Supervisory measures shall be proportionate to the risks, level of autonomy and context of use of the high-risk AI system. The measures may be:

a. Measures defined by the supplier prior to placing on the market or putting into service

b. Measures defined by the supplier prior to the introduction of the high-risk AI system and suitable for implementation by the deployment manager.

4. For the purposes of implementing paragraphs 1, 2, and 3, the high-risk AI system shall be offered to the person responsible for deployment in such a way that the natural persons entrusted with human supervision:

a] understand the capabilities and limitations of AI systems and adequately monitor their functioning.

b] raise awareness of the potential tendency to rely automatically or excessively on the output of a high-risk AI system, in particular in high-risk AI systems that are used to provide information or recommendations for the purpose of human decision-making. c] correctly interpret the output results of the high-risk AI system d] decide in any specific situation not to use the high-risk AI system or to discard, invalidate or reverse the output results generated by this system. E] intervene in the operation of the high-risk AI system or interrupt its operation so that the AI systems are safely shut down.

5. The requirement for verification by at least two natural persons shall not apply to high-risk systems used for the purposes of law enforcement, migration, border control or asylum where national or Union law considers this requirement to be disproportionate.

ISSN: 2524-2113
RENIECYT: 1702902

Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.
https://doi.org/10.35429/JLE.2025.9.15.2.1.8

Article 15

High-risk AI systems shall be designed to achieve an adequate level of accuracy, robustness and cybersecurity and to function uniformly throughout their life cycle.

Technical solutions to address vulnerabilities shall include measures to prevent, detect, combat, resolve and control attacks that seek to manipulate the training data set or components previously used for training, input information to cause the AI model to make an error, confidentiality attacks or model defects.

Article 16

Providers of high-risk AI systems shall shall ensure that their high-risk AI systems comply with the requirements in Section 2.

a] Ensure that high-risk AI systems comply with the requirements in Section 2 b] indicate on the high-risk AI system, on the packaging or in the accompanying documentation, its trade name or brand name and contact details c] have a quality management system d] keep documentation 3] retain the records f] ensure that AI systems undergo the assessment referred to in Article 43 before being placed on the market or put into service g] draw up an EU declaration of conformity in accordance with Article 47 h] affix the CE marking to the high-risk system or the accompanying documentation in accordance with Article 48 i] comply with the registration obligations j] take the necessary corrective measures l] ensure that the high-risk AI system complies with the accessibility requirements of the Union guidelines.

Chapter V General-purpose models Section 1 Article 51 Classification rules for general-purpose AI models with risk.

2. A general-purpose AI model is presumed to have high-impact capabilities when the cumulative amount of computation used for its training, measured in floating-point operations, exceeds .

Article 53

Providers of general-purpose AI models shall

[a] develop and maintain technical documentation of the model, including information on the training and testing process and the results of its evaluation.

[d] make available to the public a detailed summary of the content used for training the general-purpose AI model.

Article 56

Codes of good practice

2. The AI Office and the AI Council shall ensure that codes of good practice include:

[b] th appropriate level of detail regarding the summary of the content used for training.

Article 57

Member States shall ensure that their competent authorities establish a controlled space for AI testing at national level.

5 The controlled spaces for AI testing established shall provide a controlled environment to foster innovation and facilitate the development, training, testing and validation of AI systems for a limited period in accordance with the controlled testing space plan.

Article 59

In the controlled testing environment, personal data lawfully collected for other purposes may be processed solely for the purpose of developing, training and testing specific AI systems where the following conditions are met: [i] they keep a complete and detailed description of the process and logic underlying the training, testing and validation of the AI system together with the results of the testing process as part of the technical documentation.

Article 74

12. Where applicable, suppliers shall grant market surveillance authorities full access to the documentation, as well as to the training, validation and testing data sets used for the development of high-risk AI systems.

Annex III

High-risk systems referred to in Article 6

ISSN: 2524-2113
RENIECYT: 1702902

Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.
https://doi.org/10.35429/JLE.2025.9.15.2.1.8

High-risk AI systems pursuant to Article 6[2] are AI systems that form part of the following areas:

1. Biometrics

2. Critical infrastructure. AI systems intended as security and management components in the operation of digital infrastructures, traffic or the supply of water, gas, electricity or power.

3. Education and vocational training

4. Employment, worker management and access to self-employment

5. Access to essential private services and essential public services and benefits, and the enjoyment of these services and benefits.

6. Ensuring compliance with the law, to the extent that its use is permitted by applicable Union or national law.

7. Migration, asylum and border control management, to the extent that its use is permitted by applicable Union or national law:

8. Administration of justice and democratic processes.

Annex IV

The technical documentation referred to in Article 11[1] shall include at least the information applicable to the relevant AI system: 1.d] data requirements, in the form of technical specifications describing the training methodologies and techniques, the training data sets, a general description of those data sets and information about their origin, scope and main characteristics, the manner in which the data were obtained and selected, the labelling procedures and data cleansing methodologies.

Directive [EU] 2019/70 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market, amending Directives 96/9/EC and 2001/29/EC

Code of Practice on Security of General Purpose AI Models

According to the European Commission [2025], the Security chapter of the Code of Good Practices establishes obligations for individuals, companies or entities that develop, place on the market or use AI models that may have broad and significant impacts on security, fundamental rights or critical infrastructure.

It does not apply to all general AI systems, only those that may generate systemic risks.

Main obligations of suppliers

Security and Risk Framework

Create, implement and update security frameworks [structures for developing software] throughout the model's life cycle.

Carry out an assessment of the model and its trigger points, e.g. training time, development stages, user access, inference computation and/or functionalities.

2 Identify and analyse systemic risks

- Identify potential risks of the model

- Develop risk scenarios and analyse them with technical and scientific rigour

- Estimate the probability and severity of damage.

3. Risk assessment

- Establish criteria for deciding whether risks are acceptable

- Refrain from launching a model on the market if the risks are not considered acceptable.

4 Determination of risk acceptance.

- Proceed or not proceed in the event of systemic risks.

- Implement security measures [against data theft, unauthorised access].

ISSN: 2524-2113
RENIECYT: 1702902

Orozco-Orozco, José Zócimo. [2025]. AI regulation in relation to training databases. Journal-Industrial Organization. 9[15]1-8: e2915108.
https://doi.org/10.35429/JLE.2025.9.15.2.1.8

5. Implement functional security measures [training data filtering, monitoring input and output, as well as gradual access].

6 Mitigation measures

- Establish security goals and mitigation measures.
- These security measures will be implemented until the parameters are made available to the public or the model is securely deleted.

7. Security and reporting models.

- Develop a reporting model that includes
- A detailed justification of why the systemic risk for the model is acceptable, including details of the safety margins
- The conditions under which the justification in the previous point would no longer be valid.
- A description of how the decision to proceed with development, market release and/or use was made and how external evaluators influenced that decision.

8. Definition of responsibilities

Establish mechanisms to monitor, support and follow up on systemic risk, provide assurances about the adequacy of processes and the mitigation of this risk.

9. Reports of serious incidents

Review other sources of information
Facilitate the reporting of relevant information about incidents to the signatory or the AI office or the competent authorities.

10 Additional documentation and competence

If required by the AI office: a detailed description of the architecture model, a description of how the models are integrated into AI systems, a description of the model evaluation, and a detailed description of the mitigation systems implemented.

The assessment teams will be provided with Information including the model specification [including the instruction system], relevant training data, data sets, and past evaluation models appropriate for a] systemic risk, and b] the model evaluation method.

Guidelines on the Scope of Obligations for General-Purpose AI Models Established by the AI Act

European Commission. [2025, July 18]. Article 3[63] of the AI Act defines a general-purpose AI model as an AI model, including when such an AI model is trained on large amounts of data using large-scale self-supervision, that demonstrates significant generality and is capable of competently performing a wide range of distinct tasks, regardless of how the model is marketed.

This definition generally lists the factors that determine whether a model is considered general-purpose but does not establish specific criteria that potential providers can use to assess whether their model qualifies as a general-purpose AI model.

The case of Mexico

Initiative with draft decree issuing the the Federal Law regulating Artificial Intelligence

Presented by Senator Ricardo Monreal Ávila, Senator of the Republic and member of the MORENA party parliamentary group in the LXV legislatura, Mexico, Senate of the Republic. [2024, April 2].

Among its points, the following stand out:

Article 1

This law is of public order and general observance throughout the national territory in federal matters and has the following objectives:

To regulate the development and commercialisation of artificial intelligence systems.

To guarantee respect for the human rights of users.

ISSN: 2524-2113
RENIECYT: 1702902

To protect intellectual property rights and facilitate the national development of artificial intelligence systems.

Article 21.- Developers and suppliers of artificial intelligence systems based on Large Language Models, who use databases of information generated or created by third parties for the training of these systems, may only use this information or content with the prior agreement of the intellectual property rights holders of that information or content.

Article 22.- If 90 calendar days have elapsed since the developer or supplier of the artificial intelligence system referred to in Article 21 formally requested an agreement with the owner of the intellectual property rights of the information or content intended to be used for the purposes of training the artificial intelligence system, they may request the Institute to resolve the terms and conditions that could not be agreed upon.

The Institute shall resolve the matter within a period not exceeding 120 calendar days from the date on which either party has notified it of the disagreement.

Conclusions

European legislation continues to lead the world in the field of AI systems
.

AI system providers must have greater control and oversight of AI systems in order to eliminate or mitigate the risks associated with these systems as far as possible.

AI systems must have continuous human supervision in case they fail under reasonable use to avoid or mitigate risks to fundamental rights.

The fact that an AI system is open source does not exempt providers from having controls, documentation, monitoring, sending periodic reports, having mitigation plans and other controls.

Declarations

Conflict of interest

The author declares that there is no conflict of interest.

Author contribution

The author contributed to the entire research process, including study design, data analysis, writing, and results expressed in conclusions and proposals.

Funding

The research was supported by a grant to improve researcher conditions, as a member of the National System of Researchers Level I.

From the 2025 Research Support Program [PROSNI] of the University of Guadalajara.

Abbreviations

AI, Artificial Intelligence.

References

Background

Comisión Europea. [2025, 10 de julio]. Código de buenas prácticas para modelos de IA de propósito general [Capítulo de Seguridad]. Comisión Europea.

Basics

Unión Europea. [2024]. *Reglamento [UE] 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024*. *Diario Oficial de la Unión Europea*, L 2024/1689.

Support

México, Senado de la República. [2024, 2 de abril]. *Iniciativa con proyecto de decreto por el que se expide la Ley Federal que regula la Inteligencia Artificial*.

Discusions

Comisión Europea. [2025, 18 de julio]. *Directrices sobre el alcance de las obligaciones para los proveedores de modelos de IA de propósito general conforme a la Ley de IA*.

ISSN: 2524-2113
RENIECYT: 1702902