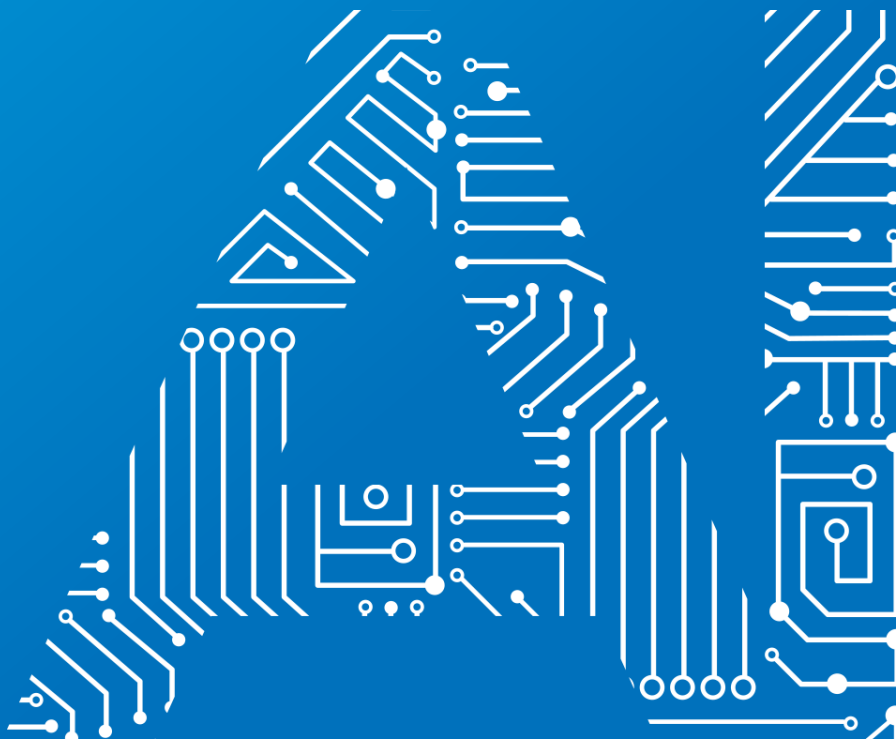
Ethics and regulations of AI – Guide for responsible AI

BASQUE ARTIFICIAL INTELLIGENCE CENTER



Contents

Illustrations

BASQUE ARTIFICIAL INTELLIGENCE CENTER

Context

Within the AI ecosystem in the Basque Country, BAIC is a benchmark, a meeting and collaboration point for both public and private agents for the promotion of AI in the Basque Country. BAIC's mission focuses on increasing Basque competitiveness by accelerating the development and implementation of AI in an ethical, collaborative and effective manner.

In terms of BAIC's values, these relate to its contributions and actions being aligned with the sustainable, equitable and ethically responsible development and implementation of AI. In the same way, BAIC acts as the lighthouse that guides and supports the various agents of the Basque Country's AI ecosystem towards the development and application of trustworthy and responsible AI, ensuring that AI innovations are not only technically advanced and safe, but also legally sound and ethically flawless, thus fostering an environment of trust and collaboration that drives technological progress at the service of society.

BAIC fulfils its mission through the deployment of different strategic axes, which together allow it to build comprehensive solutions to complex AI challenges. As part of the Observatory strategic axis, BAIC aims to strengthen and develop the AI ecosystem in the Basque Country with a focus on vigilance, knowledge, connection and positioning between the different agents in the territory. This axis includes initiatives aimed at guiding and supporting the agents of the ecosystem towards alignment with the pillars of robustness, legality and ethics, as well as with the requirements for trustworthy and responsible AI in its development and implementation.

In this context, in a first definition of an ethical framework for the use, implementation and development of AI in the Basque Country ecosystem, in early 2024 BAIC presented the Code of Ethics for the development, use and implementation of AI in the Basque Country[1], with an approach based on the voluntary and unilateral support of this code of ethics by the organisations in the Basque Country's AI ecosystem.

For its part, through this document, **BAIC presents an Ethics and Regulatory Guide with the aim of helping agents in the Basque AI ecosystem to understand and align themselves with the new AI regulations, facilitating the adaptation of their strategies and processes to comply with both legislative and ethical aspects**. This new guide provides detailed guidelines and practical examples to ensure regulatory compliance and promote responsible and transparent practices in the development and use of AI.

[1] BAIC – Code of Ethics for the development, use and implementation of AI in the Basque Country

Ethical and legal AI

BAIC

Ethical and legal AI

What is an AI System?

In this Guide we take the definition provided by the European Artificial Intelligence Regulation[2] (AI ACT):

> *"A system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts".*

What are the requirements for TRUSTWORTHY AI?

Trustworthy AI rests on three pillars: it must be legal (comply with applicable law), ethical (respect ethical principles), and robust (operate safely, so as not to cause accidental damage).

Each of these pillars is necessary, but not sufficient.

Thus, for trustworthy AI to exist, it is not only necessary to comply with the law. We must bear in mind that technological development usually comes before the law, so it is necessary for AI systems to respect ethical principles.

The use of AI systems in our society raises some ethical challenges, related, for example, to their effects on individuals and society, decision-making capabilities, and safety.

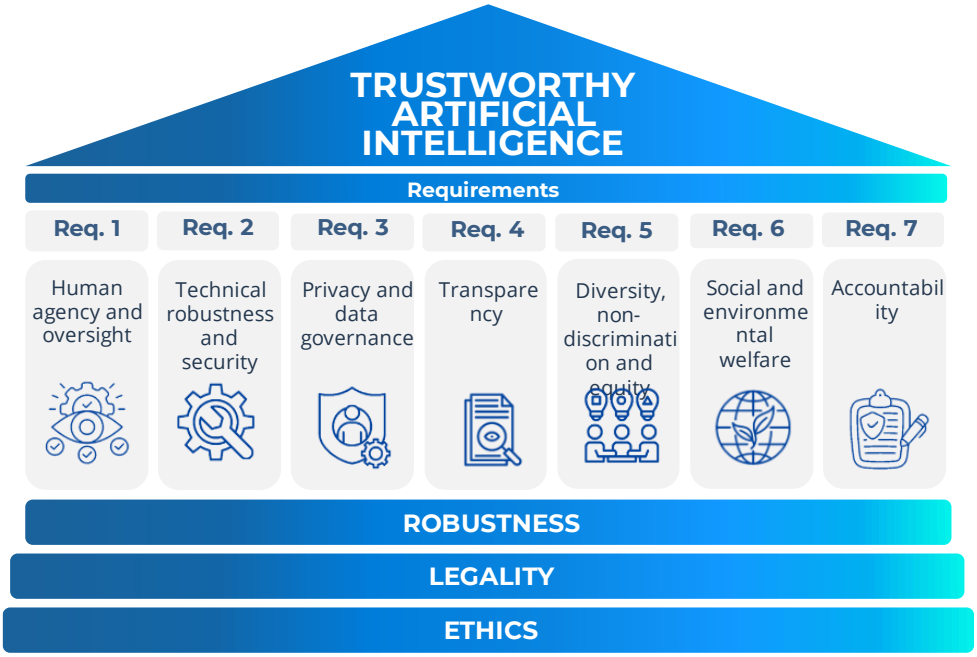
The ethical principles that must inspire the development and use of AI are based on fundamental rights (respect for human dignity is the common basis), and are: (i) respect for human autonomy, (ii) prevention of harm, (iii) fairness, (iv) explainability.

The ethical guidelines for trustworthy AI developed by the group of high-level experts set up by the European Commission[3] establish seven key requirements for trustworthy AI.

[2] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

[3] https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

Illustration 1: European Commission - Requirements for trustworthy AI



BAIC aims to guide the different agents of the Basque AI ecosystem towards alignment with the pillars of robustness, legality and ethics, as well as with the requirements for the development and use of trustworthy and responsible AI through the *Code of Ethics for the development, use and implementation of Artificial Intelligence in the Basque Country*[4]. In it, an ethical framework is proposed with the aim of promoting ethical and responsible practices that maximise the benefits of AI, ensure its reliability and minimise potential risks, fostering the trust of society and agents in the AI ecosystem and expressing our shared values.

The seven requirements for trustworthy AI have been incorporated, to a greater or lesser extent, in the articles of the Artificial Intelligence Regulation[5]:

Requirement	Description	Article of AI ACT
Human action and oversight	AI systems are developed and used as a tool at the service of people, respecting human dignity and personal autonomy, and operated in a way that can be adequately controlled and monitored by human beings.	Art. 14. Human oversight

[4] BAIC – Code of Ethics for the development, use and implementation of AI in the Basque Country

[5] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

Technical soundness and security	AI systems are developed and operated in such a way as to be robust in the event of problems and resilient to attempts to alter the use or operation of the AI system to allow its unlawful use by third parties and to minimise unintended harm.	Art. 15. Accuracy, robustness and cybersecurity
Privacy and data management	AI systems are developed and operated in accordance with privacy and data protection standards, and the data they process meet high standards in terms of quality and integrity.	Art. 10. Data and data governance
Transparency	AI systems are developed and used in a way that allows for adequate traceability and explainability, while at the same time making individuals aware that they are communicating or interacting with an AI system and adequately informing those responsible for deployment about the capabilities and limitations of that AI system and affected individuals about their rights.	Art. 11. Technical documentation (transparency of the provider vis-à-vis the authorities) Art. 13. Transparency and communication of information to the deployers (transparency of the provider vis-à-vis deployers) Art. 50. Transparency obligations of providers and deployers of certain AI systems (transparency vis-à-vis individuals).
Diversity, non-discrimination and equity	AI systems are developed and used in a way that is inclusive of diverse agents and promotes equal access, gender equality and cultural diversity, while avoiding discriminatory effects and unfair biases prohibited by national or Union law.	Art. 5. Prohibited AI practices Art. 9. Risk management system (context to be taken into account) Art. 14. Human oversight Art. 27. Fundamental rights impact assessment for high-risk AI systems

		Appendix III. High-risk AI systems included in the list of use cases in Appendix III.
Social and environmental welfare	AI systems are developed and used in a sustainable and environmentally sound manner and for the benefit of all human beings, while monitoring and assessing the long-term effects on people, society and democracy.	Art. 1, Art. 26 (obligation to report serious incidents, including environmental damage).
Accountability	AI systems must be auditable, minimise negative effects, and if negative effects exist they must be reported and offset.	Art. 17. Quality management system (including accountability framework defining responsibilities of staff, management and non-management). Chapter IX, Section 4. Remedies.

In short, the adoption of anthropocentric AI, i.e., human-centred AI, is promoted.

In any case, the EU encourages providers and those responsible for the deployment of all AI systems, whether high-risk or not, and AI models, to apply, on a voluntary basis, additional requirements set out in the EU Ethical Guidelines for Trustworthy AI[6].

What is the relationship between TRUSTWORTHY AI and RESPONSIBLE AI?

Responsible AI that, when designed, developed and implemented, respects ethical principles and applicable regulations, and is auditable, will result in AI that is trustworthy towards society and towards the people exposed to that AI. In other words, an AI without risk to health, safety and fundamental rights and respectful of the environment and democracy.

[6] https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1

What is the difference between ETHICAL and REGULATORY?

Ethical principles are not legally binding, and therefore non-compliance does not give rise to legal consequences, notwithstanding the reputational consequences that may be suffered by an organisation engaging in unethical behaviour.

However, the obligations set out in the regulations are legal obligations and, therefore, non-compliance constitutes a legal violation from which legal consequences may arise for the non-compliant organisation in the form of claims, fines, etc.

In practice, we find ethical principles that are embedded in the applicable regulations, which gives them legal force (as we have seen in the table above).

Regulatory framework for AI

BAIC

Regulatory framework for AI

Artificial intelligence, to the extent that it involves the use of data, must respect applicable data regulations, such as data protection and intellectual property law.

Moreover, the recently published European Regulation on Artificial Intelligence (AI ACT)[7] must also be taken into account, the key elements of which are discussed in the next section of this Guide.

General Data Protection Regulation

When training an AI model with datasets that include personal data, when the input information to the AI system incorporates personal data, or when the output results of the AI system affect individuals in a significant way (for example, by having legal effects on them) and, in general, if personal data are processed, we have to apply data protection regulations (mainly the General Data Protection Regulation or GDPR).

To summarise, the data protection principles must be complied with, which oblige us to:

1. Process data only when we have a legitimate basis for doing so; and there are only six bases: consent, contract, legal compliance, vital interest, legitimate interest and public interest (principle of lawfulness).
2. Inform individuals of what we are going to do with their data (principle of transparency).
3. Use data only for the purposes for which we have informed them (principle of purpose limitation).
4. Use only those data that are necessary for those purposes (principle of data minimisation).
5. Use up-to-date data (principle of accuracy).
6. Erase data when they are no longer necessary for those purposes (principle of limitation of the retention period).
7. Use data while ensuring adequate security (principle of integrity and confidentiality).

For example, under the principle of data minimisation, it is necessary to analyse whether the AI model can be trained with pseudonymised or even anonymised data without affecting the intended purpose. Anonymisation requires the application of specific techniques by specialised staff to minimise the risk of re-identification.

Moreover, for AI-based projects processing personal data, a data protection impact assessment will most likely be necessary to analyse whether the processing is necessary and proportionate, and its impact on the rights and freedoms of the individuals concerned.

[7] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

If the results of the AI output involve automated individual decisions on individuals, it must be verified whether any of the permitted assumptions are present, and they must be informed (i) that such automated decisions will exist, (ii) of the logic applied, and (iii) of the significance and expected consequences for the data subject.

Along the same lines, the Spanish Workers' Statute establishes the right of the Works Council to be informed of the rules on which algorithms (whether AI or non-AI) affecting decision-making on working conditions are based.

Intellectual property regulations

Pre-existing data used for model training or as input to an AI system may be subject to intellectual property rights. Occasionally, and in the case of non-profit research organisations, it may be possible to rely on the text and data mining exception. For all other cases, that is, works other than text and data and for purposes other than research, the question of authorisation to use protected works arises.

Furthermore, questions are raised concerning the protection of both the artificial intelligence itself and the output results obtained.

European AI Regulation

BAIC

European AI Regulation

General information

The Artificial Intelligence Regulation (AI ACT)[8] was published on 12/07/2024 in the Official Journal of the European Union (OJEU) and entered into force on 01/08/2024.

However, it will not apply until 02/08/2026, with some different deadlines for certain provisions, to which we refer below.

What is the approach of the AI ACT?

The Artificial Intelligence Regulation (AI ACT) is a product safety regulation, like regulations governing toys, lifts, vehicles or medical devices.

Product safety regulations set regulatory requirements to ensure that products with the potential to cause harm to human health or safety have been tested and validated before being placed on the market, so that only safe products reach the market.

Thus, this type of regulation requires the manufacturer to conduct conformity assessments of the product[9] (including by independent third parties), to include the CE marking on the product, to withdraw the product from the market if it poses a risk, to provide instructions for use to prevent misuse of the product leading to damage, etc.

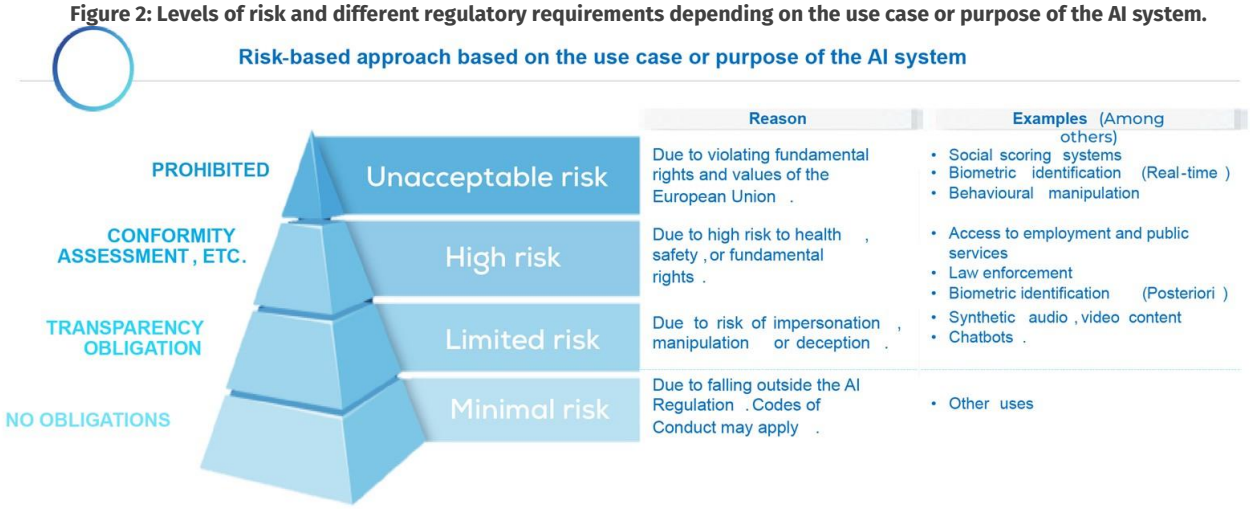
These requirements (conformity assessment, CE marking, instructions for use, post-market surveillance) will also be found in the AI ACT, which aims to ensure that AI systems in the European Union respect the values of the Union and are safe, and do not cause harm to the health and safety of individuals and their fundamental rights.

The AI ACT is often referred to as a risk-based approach because four risks can be deduced from its articles.

[8] Regulation (EU) 2024/1689, of the European Parliament and of the Council, dated 13 June 2024, establishing harmonised rules in the field of artificial intelligence and amending Regulations (EC) No. 300/2008, (EU) No. 167/2013, (EU) No. 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828.

[9] Procedure to demonstrate that the product complies with all mandatory product requirements before placing it on the EU market. The conformity assessment is sometimes required to be conducted by an independent body.

Figure 2: Levels of risk and different regulatory requirements depending on the use case or purpose of the AI system.



What are the relevant concepts?

For the purposes of the Artificial Intelligence Regulation, it is necessary to understand the concepts of "AI system", "general-purpose AI model", "general-purpose AI system" and "systemic risk general-purpose AI model":

Concept	Characteristics
AI system	1. machine-based 2. autonomy 3. adaptive capacity 4. inference of output results (predictions, contents, recommendations or decisions) from input information, 5. influence of output results on the environment
General-purpose AI model (MIAUG)	1. generality and the ability to perform a wide variety of different tasks competently, 2. can be integrated into various downstream systems or applications. For example, large natural language models are general-purpose AI models. They are typically trained using large volumes of data and through a variety of methods, such as self-supervised, unsupervised or reinforcement learning. They can be brought to market in a variety of ways, for example through libraries, application programming interfaces (APIs), as a direct download or as a hard copy. These models can be modified or refined and transformed into new models. Although AI models are essential components of AI systems, they do not

	in themselves constitute AI systems. AI models require the addition of other components, such as a user interface, to become AI systems. AI models are often embedded in and form part of AI systems.
General-purpose AI system	An AI system based on a MIAUG, which can serve a variety of purposes, both for direct use and for integration into other AI systems. For example, Chat GPT.
Systemic risk MIAUG	A general-purpose AI model that has high-impact capabilities. Assumption: when the cumulative amount of computation used for training, measured in floating point operations, is greater than 10^{25}.

Which areas are outside the AI ACT?

There are some areas where the Artificial Intelligence Regulation does not apply:

a) AI systems for military, defence or national security purposes.

b) AI systems or models developed and put into service for the sole specific purpose of scientific research and development.

c) Research, testing or development activities relating to AI systems or AI models prior to their introduction into the market or commissioning. Tests under real conditions are not covered by this exclusion.

d) Purely personal (non-professional) activities carried out by individuals.

Which organisations must comply with the AI ACT?

The AI ACT establishes obligations both for providers that design and develop AI systems within its scope, and for the organisations that use such systems, which are referred to in the AI ACT as deployers.

> In an earlier version they were called users, but this term has been dropped, presumably to avoid confusion with the terminology used in consumer and user regulation, which is aimed at individuals who purchase products or services for domestic or non-professional purposes.

It also establishes obligations for providers of general-purpose and systemic risk AI models.

What are the competent authorities for AI in the Member States and the EU?

The AI ACT aims to harmoniously regulate the use, development and oversight of AI in the European Union. To this end, it establishes a governance framework at European level and at the level of each Member State:

1) European Union:

- AI Office (attached to the Commission): will advise on general-purpose AI models, and facilitate the development of codes of good practice for AI.
- AI Council[10]: will advise the Commission and Member States to facilitate the consistent and effective implementation of the AI ACT.

2) Competent national authorities:

Each Member State must designate at least one market surveillance authority, responsible for ensuring compliance with the AI ACT, and applying sanctions in case of violations.

The Spanish Agency for the Supervision of Artificial Intelligence (abbreviated as AESIA in Spanish) will be the market surveillance authority in Spain.

Are there plans to publish standards to accredit compliance with the AI ACT?

Standardisation bodies are currently developing, at the request of the European Commission, a set of standards that will allow accreditation (by means of presumption) of compliance with the Artificial Intelligence Regulation in relation to the following ten areas:

1) Risk management system for AI systems
2) Governance and quality of the datasets used to build AI systems
3) Record keeping through the logging capability of AI systems
4) Transparency and information to users of AI systems
5) Human oversight of AI systems
6) Accuracy specifications for AI systems
7) Robustness specifications for AI systems
8) Cybersecurity specifications for AI systems
9) Quality management systems for providers of artificial intelligence systems, including post-marketing follow-up processes
10) Conformity assessment of AI systems

10 In previous drafts of the AI ACT, it was referred to as the "AI Committee".

Prohibited AI practices - Unacceptable risk

The AI ACT prohibits the following AI systems from being marketed, installed and used:

1) AI system that **uses subliminal, manipulative or deceptive techniques** to alter the behaviour of an individual or group.

2) AI system that **exploits vulnerabilities** (age, disability, social or economic status) to alter the behaviour of an individual or group.

3) AI systems for **social scoring**.

4) AI systems for **assessing the risk of a person committing crimes**, solely **based** on **their profile** or **personality**.

   Exception: AI systems to support human assessment already based on objective facts directly related to a criminal activity.

5) AI systems to create **facial recognition databases through the non-selective extraction of facial images** from the internet.

6) AI systems for **inferring a person's emotions in workplaces and educational institutions**.

   Exception: AI system intended to be installed or introduced in the market for medical or security purposes.

7) **Biometric categorisation** systems that individually classify individuals based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation (special categories of data).

   This prohibition does not include the tagging or filtering of biometric datasets acquired lawfully, such as images, based on biometric data or the categorisation of biometric data for law enforcement purposes.

The AI ACT prohibits the use of:

8) **"Real-time" remote biometric identification systems in publicly accessible spaces for law enforcement purposes**, subject to certain defined exceptions and to a number of safeguards.

These practices will be prohibited as of 02/02/2025.

As a reminder, they correspond to AI systems of high risk to health, safety, or fundamental rights (these AI systems are subject to regulatory requirements).

High-risk AI systems are classified into **two types**:

a) On the one hand, AI systems associated with regulated products (that is, products that are subject to product safety regulations), which may have a high impact on the health or safety of people.

b) And, on the other hand, AI systems that are included in a list of use cases that are considered high risk due to their high impact on people's fundamental rights.

a) AI systems associated with regulated products:

In turn, there are two types:

1. **AI system that is a regulated product** and this product must undergo a **third party conformity assessment**. For example, an autonomous robot or a diagnostic system.
2. **AI system that is a safety component of a regulated product** and this product is subject to a **third party conformity assessment**. For example, a lift brake system.

The regulated products affected are as follows:

- Machinery | Toys | Jet skis | Lifts and safety components for lifts | Protective apparatus and systems for use in potentially explosive atmospheres | Radio equipment | Pressure equipment | Cable transport installations | Personal protective equipment | Apparatus burning gaseous fuels | Medical devices | In vitro diagnostic medical devices.

- Civil aviation | Vehicles | Marine equipment | Railway system | Unmanned aircraft.

Therefore, it will be necessary to analyse whether or not, according to the regulations governing the product in question (of which the AI system is a safety component or is the product itself), the product must undergo a third party conformity assessment, so that, if the answer is yes, the AI system is considered to be high risk.

The provisions of the AI ACT in relation to this type of AI system will be applicable as of 02/08/2027.

b) AI systems included in the list of high-risk use cases

Eight areas are established, identifying a number of AI system use cases or applications, which are considered to be high risk:

1. Biometrics
2. Critical infrastructures
3. Education and vocational training
4. Employment, workforce management and access to self-employment

5. Access to essential services and benefits
6. Law enforcement
7. Migration, asylum and border control management
8. Administration of justice and democratic processes

The provisions of the AI ACT in relation to this type of AI system will be applicable as of 02/08/2026.

Detailed below are the use cases of AI systems that are considered high risk in each of the 8 areas:

(1) Biometrics

a) Remote biometric identification systems (excluding biometric verification).

b) AI systems for biometric categorisation according to sensitive characteristics based on the inference of these characteristics.

c) AI systems for the recognition of emotions (outside workplaces or educational institutions).

(2)Critical infrastructures

AI systems that are safety components in the management and operation of critical digital infrastructures, road traffic, or water, gas, heat or electricity supply.

(3)Education and vocational training

a) AI systems to determine access or admission of individuals to educational and vocational training institutions.

b) AI systems for assessing learning outcomes.

c) AI systems to assess the appropriate level of education a person will receive or be able to access.

d) AI systems for monitoring and detection of prohibited student behaviour during exams.

(4)Employment, workforce management and access to self-employment

a) AI systems for recruitment or selection of staff.

b) AI systems for decision-making on working conditions, promotion or termination of employment contracts, allocation of tasks based on individual behaviour or personal traits or characteristics, or for monitoring and evaluating the performance and behaviour of individuals in the framework of employment relationships.

(5)Access to and enjoyment of essential services and benefits

a) AI systems to assess the eligibility of individuals to benefit from essential public assistance services and benefits, including healthcare services, and to grant, reduce, withdraw or claim back such services and benefits.

b) AI systems to assess the creditworthiness of individuals or to establish their credit rating, with the exception of AI systems used to detect financial fraud.

c) AI systems for risk assessment and pricing of life and health insurance.

d) AI systems for the assessment and classification of emergency calls made by individuals or for dispatching or prioritisation of dispatching first responders in emergency situations, for example police, fire and medical assistance services, and in patient triage systems in the context of emergency healthcare.

(6) Law enforcement (police purposes)

a) AI systems to assess the risk of an individual being a victim of crime.

b) AI systems such as polygraphs or similar.

c) AI systems to assess the reliability of evidence during the investigation or prosecution of crimes.

d) AI systems to assess the risk of an individual committing a crime or reoffending or to assess personality traits and characteristics or past criminal behaviour of individuals or groups.

e) AI systems for profiling individuals, during the detection, investigation or prosecution of criminal offences.

(7) Migration, asylum and border control management

a) AI systems such as polygraphs or similar tools.

b) AI systems to assess the risk posed by an individual entering the territory of a Member State.

c) AI systems to assist in the examination of applications for an asylum, visa or residence permit.

d) AI systems to detect, recognise or identify individuals, with the exception of the verification of travel documents.

(8) Administration of justice and democratic processes

a) AI systems to assist a judicial authority in the investigation and interpretation of facts and law, as well as in the application of the law to a particular set of facts.

b) AI systems to influence the outcome of an election or referendum or the voting behaviour of individuals exercising their right to vote in elections or referenda.

However, an AI system included in the list of use cases **will not be considered high-risk if it does not pose a significant risk** of causing harm to the health, safety or fundamental rights of

individuals, for example, if it does not substantially influence **the outcome of decision-making**. Thus, **this exception applies** to the following AI systems:

a) To perform a **limited procedural task**.

b) To improve the **outcome of a previous human activity**.

c) To **detect patterns of decision-making or deviations** from patterns of decision-making and not replace or influence prior human assessment.

d) To perform a **preparatory task**.

However, there is an **exception to the exception**, as AI systems included in the list of use cases **will always be considered as high risk when the AI system profiles** individuals.

Which agents in the AI value chain are subject to the AI ACT?

The agents that are subject to the AI ACT are those that are part of the value chain of high-risk and limited-risk AI systems and general-purpose AI models:

- AI system providers
- MIAUG providers
- Authorised representatives of non-Union based providers
- Importers of AI systems
- Distributors of AI systems
- Deployers of AI systems
- Providers of tools and components that are integrated into AI systems
- Manufacturers of products comprising AI systems

What are the requirements for high-risk AI systems?

High-risk AI systems must meet a number of regulatory requirements (and it is up to the provider to ensure that they are met):

1. Risk management system

   A risk management system must be established, implemented, documented and maintained, and implemented throughout the life cycle of the system, with regular reviews.

   Among other aspects, the risks that the system may pose to health, safety or fundamental rights when used in accordance with its **intended purpose**, and also when a **foreseeable misuse** occurs, must be analysed.

The risks to be managed are those that can be mitigated or eliminated through the design or development of the system or the provision of appropriate technical information.

High-risk AI systems must be **tested** before being placed on the market or commissioned to verify that they operate in a manner consistent with their intended purpose and meet regulatory requirements.

2. Data and data governance

   High-risk AI systems that involve training AI models with data must be developed from datasets (training, validation and test) that meet certain quality criteria:

   1. **Data governance and management** suitable for the intended purpose of the high-risk AI system. For example: **assessing** the availability, quantity and adequacy of the required datasets, or **examining possible biases** that may affect the health and safety of individuals, adversely affect fundamental rights or lead to prohibited discrimination, or identifying gaps or deficiencies in the data.
   2. They must be **relevant, sufficiently representative and**, to the greatest extent possible, **free of errors and complete** in view of their intended purpose. They must also have the **appropriate statistical properties**.
   3. They must take into account the **particular characteristics of the specific environment** (geographical, contextual, behavioural or functional) in which the high-risk AI system is intended to be used.

3. Technical documentation

   This is information that is intended to be provided to national competent authorities and, where appropriate, to notified bodies[11].

   The **technical documentation of a high-risk AI system**:

   1. It must be **drawn up before** it is placed on the market or commissioned, and kept up to date.
   2. It must be **written** in a way that (i) **demonstrates that** the system **complies with the regulatory requirements** and (ii) **provides** the information necessary **for the national competent authorities and notified bodies** to assess the compliance of the AI system with those requirements.
   3. It must include the minimum content established by the AI ACT.

4. Record keeping

11 Notified bodies are the certification bodies that conduct the conformity assessment of the AI system when an independent third party is required to do so.

High-risk AI systems must technically allow the **automatic recording of events** (log files) throughout the life cycle of the system. For example, to **detect situations** that could lead to the system posing a **risk** to the health, safety or fundamental rights of individuals.

5. Transparency and communication of information to deployers

   High-risk AI systems must be designed and developed with sufficient **transparency** for their

   (i) **output results to be correctly interpreted and used by the deployers**, and

   (ii) for the **provider and the deployer to fulfil their obligations** established in the AI ACT.

   High-risk AI systems must be accompanied by instructions for use, with clear and understandable information for deployers; including, among other aspects:

   - the intended purpose;
   - the **characteristics, capabilities and limitations of the operation** of the high-risk AI system;
   - **information allowing deployers to interpret the output results** of the high-risk AI system and use them appropriately;
   - **human oversight measures** (including technical measures) to facilitate deployers' interpretation of the output results.

6. Human oversight

   In line with the "human in command" premise, high-risk AI systems must be designed and developed in such a way that they can **be monitored by human beings during their use**, which requires an appropriate **human-machine interface**.

   The oversight measures must be proportionate to the risks, the level of autonomy and the context of use of the high-risk AI system.

   Oversight is ensured through system-integrated technical measures and organisational measures implemented by the deployment manager.

   The high-risk AI system must be provided to the deployment manager in such a way that **the individuals tasked with human oversight** are able to:

   a) **understand the capabilities and limitations of the system and monitor its operation**;

   b) **be aware of "automation bias"** to avoid the tendency to automatically rely on the output results generated by the system;

   c) **correctly interpret the output results** of the system;

   d) **not use the system in a particular situation or discard the output results** it generates;

   e) **stop the operation** of the system.

7. Accuracy, robustness and cybersecurity

   High-risk AI systems must be designed and developed to achieve an **appropriate level of accuracy, robustness and cybersecurity.**

Before introducing the AI system to the market or commissioning it, the provider must demonstrate that the system meets the above requirements, as follows:

a) AI systems of biometrics use cases: by the application of standardisation standards together with the conformity assessment by internal control, or through the conformity assessment conducted by a notified body[12].

b) AI systems of the remaining use cases: through the conformity assessment by internal control.

c) AI systems associated with regulated products: through the conformity assessment by an independent third party, according to their applicable legislation.

What are the obligations in the development of high-risk AI systems?

Providers of high-risk AI systems have the following obligations:

1) To ensure that their systems comply with the regulatory requirements for high-risk AI systems (indicated in the previous section).

2) To indicate their name, registered trade name or registered brand and their contact address.

3) To have a quality management system that complies with the requirements established in the AI ACT.

4) To retain the documentation related to the system established in the AI ACT for a period of 10 years.

5) To retain the log files automatically generated by their systems for at least 6 months.

6) To submit their systems to the relevant conformity assessment procedure[13] before being placed on the market or commissioned.

7) To draw up the EU Declaration of Conformity[14].

8) To affix the CE marking on the system.

9) To register their system and themselves in the EU database.

12 Independent third party officially authorised to conduct this conformity assessment.

13 By internal control or by independent third party assessment, as the case may be

14 If the system involves the processing of personal data, it must contain the declaration that the AI system complies with the GDPR.

10) To take the necessary corrective measures when they consider that their system does not comply with the AI ACT or poses a risk to health, safety or fundamental rights and to inform operators and market surveillance authorities.

11) To demonstrate, at the request of the national competent authority, the compliance of the high-risk AI system with the regulatory requirements.

12) To ensure that the high-risk AI system meets accessibility requirements in accordance with the applicable directives.

13) To take measures to ensure that their staff and others involved in the operation and use of AI systems have a sufficient level of AI literacy so that they are aware of the opportunities and risks posed by AI and the harm it can cause.

What are the obligations in the use of high-risk AI systems?

Deployers of high-risk AI systems must comply with the following obligations:

1) To take appropriate technical and organisational measures to operate the system in accordance with the instructions for use.

2) To entrust human oversight to individuals with the necessary competence, training and authority.

3) To ensure that the input data are relevant and sufficiently representative taking into account the intended purpose of the system (to the extent that they have control over such data).

4) To monitor the operation of the system based on the instructions for use.

5) To notify the provider and the market surveillance authority if the system poses a risk to health, safety or fundamental rights, and to suspend the use of that system.

6) To inform the provider and the market surveillance authority if a serious incident is detected.

7) To retain the log files automatically generated by the system for at least 6 months (to the extent that these files are under their control).

8) In the case of employers, to inform workers' representatives and affected workers that they will be exposed to the use of the high-risk AI system.

9) To use the information provided by the provider in the instructions for use to conduct, where appropriate, the data protection impact assessment.

10) In the case of high-risk AI systems included in the list of use cases, which make decisions related to individuals, to inform these individuals that they are exposed to the use of high-risk AI systems.

11) To take measures to ensure that their staff and others involved in the operation and use of AI systems have a sufficient level of AI literacy so that they are aware of the opportunities and risks posed by AI and the harm it can cause.

12) To conduct, in some cases*, an assessment of the impact that the use of the system may have on fundamental rights, and to notify the market surveillance authority of the results.

*Deployers (i) that are bodies governed by public law, or private entities providing public services, or (ii) that use AI systems to establish credit ratings for individuals or to price life and health insurance.

What are the obligations when supplying components that are integrated into a high-risk AI system?

The provider of a high-risk AI system and the third-party provider of tools or components to be integrated into this system must enter into a written agreement that establishes:

- the information,
- the capabilities,
- technical access, and
- other assistance

necessary to allow the provider of the high-risk AI system to meet the obligations set forth in the AI ACT. This obligation does not apply to third parties who make tools or components other than general-purpose AI models available to the public under a free and open source licence.

Transparency obligations for some AI systems - Risk of manipulation, spoofing or deception

Because of the risks of manipulation, spoofing or deception they pose, certain AI systems are subject to a number of transparency obligations. If these systems are high risk, they must also comply with the obligations established for high-risk systems.

Some transparency obligations are placed on the providers of the AI system, others on those responsible for the deployment:

AI systems	Transparency obligations	Obligated agent
Intended to interact directly with individuals	Individuals must be informed that they are interacting with an AI system (unless it is obvious).	Provider

That generate synthetic audio, image, video, video or text content	The output results must be marked and must be detectable as synthetic.	Provider
Emotion recognition Biometric categorisation	To inform exposed persons of the operation of the system To process their personal data in accordance with data protection regulations.	Deployer
That generate deep fake image, audio or video content	To publicise that the content is synthetic.	Deployer
That generate text informing the public on matters of public interest	To disclose that the text has been artificially generated.	Deployer

Other AI systems - Minimal risk

Providers of artificial intelligence-based systems that fall outside the AI ACT (because they do not fit the legal definition of an AI system, because they are not high risk, or because they do not require transparency measures) may voluntarily apply Codes of Conduct.

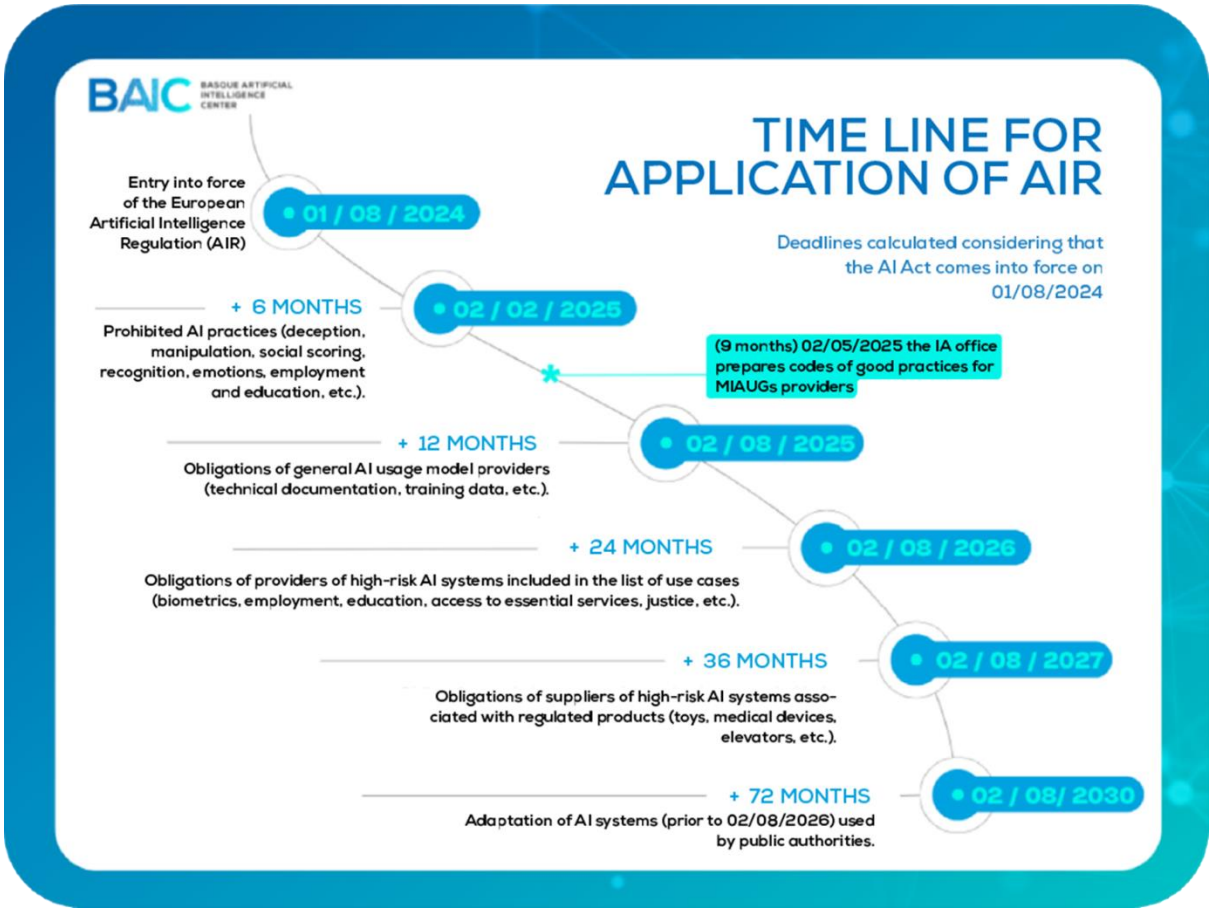
The European Union encourages the establishment of Codes of Conduct for the voluntary implementation of all or part of the requirements applicable to high-risk AI systems, adapted taking into account the intended purpose of the systems and the lower risk they pose and considering the available technical solutions and best practices in the industry[15].

When will the AI ACT apply?

The AI ACT became effective on 1 August 2024. However, the implementation periods will be staggered depending on the AI system and/or its purpose.

15 For example, the model card.

Figure 3: Timeline for implementing the AI ACT



What happens to high-risk AI systems and general-purpose AI models prior to the implementation of the AI ACT?

If the AI system already placed on the market or commissioned is prohibited, it must be withdrawn by 02/02/2025.

In all other cases, the AI ACT applies to operators of high-risk AI systems that have been placed on the market or commissioned before 02/08/2026, if after 02/08/2026, their design is significantly modified.

In any case, providers and deployers of high-risk AI systems intended for use by public authorities must take the necessary measures to comply with the requirements and obligations of the AI ACT by 02/08/2030 at the latest.

What about before the implementation of the AI ACT?: AI Pact

The Commission is proposing a so-called AI Pact, in order for organisations to voluntarily commit to start implementing the requirements of the AI ACT before its mandatory implementation. To do so, organisations participating in this initiative commit to the following key actions:

- Adopt an AI governance strategy to foster AI adoption in the organisation and work towards future compliance with the AI ACT;
- Conduct a mapping of AI systems developed or used in areas that would be considered high risk under the AI ACT;
- Promote awareness and AI literacy of staff and others involved in the deployment of AI systems on their behalf, taking into account their technical knowledge, experience, education and training and the context in which AI systems will be used, and considering the individuals or groups of individuals affected by the use of AI systems.

In addition to the above, the AI Pact contains additional commitments that the organisation may make, which vary depending on whether the organisation develops AI systems or uses AI systems.

Recommendations for addressing compliance with the AI ACT

BAIC

Recommendations for addressing compliance with the AI ACT

It is recommended that organisations start taking steps to comply with the European Artificial Intelligence Regulation.

What actions can organisations take to adapt to the AI ACT?

1) Identify AI within the organisation.

   It is very likely that there are already existing uses and/or developments of artificial intelligence within the organisation. In order to be able to manage and control these uses and/or developments, they must first be identified and inventoried, and for this purpose there must be a person in charge of this task.

2) Analyse whether or not the identified AI uses/developments fall within the scope of the AI ACT.

   Once identified, it is necessary to check whether the uses and/or developments of AI in the organisation, by their purpose, are excluded from the scope of the AI ACT (military, defence or national security purposes, sole and specific purpose of scientific research and development, research, testing or development activities concerning AI systems or AI models prior to their introduction to the market or commissioning).

3) Analyse whether the identified uses/developments of AI are considered AI systems for the purposes of the AI ACT.

   In this section we will analyse whether or not the uses of AI fall within the definition of AI systems, and, if so, whether they are AI models, or AI components or solutions that are integrated within AI systems.

4) Analyse whether the identified AI is a prohibited practice/high-risk AI system.

5) Analyse the role of the organisation with respect to the identified AI (AI provider/acquirer/user).

   Once the existing uses/developments of AI in the organisation have been identified, we must analyse what role the organisation adopts with respect to them (if it uses AI, if it develops it, if it trains AI models, if it acquires AI elements, if it provides AI solutions), because depending on the role assumed, the organisation will have to fulfil certain obligations.

   For example, before using or commissioning a high-risk AI system in the workplace, the organisation must inform workers' representatives and affected workers that they will be exposed to high-risk AI systems.

6) Establish an AI governance model in the organisation (roles, functions, etc.).

In order to comply with the regulatory framework applicable to AI, it is necessary for the organisation to have an AI governance model in place, setting out the necessary roles (for example, AI Ethics Committee, AI Manager, Data Controller, persons in charge of human oversight, involvement of the Data Protection Officer, Privacy Officer/Committee, etc.), and the functions of each of them.

7) Establish an AI management system (including AI risk management).

An AI management system needs to be established, implemented and maintained within the organisation, including an AI risk management system and the other requirements set out in the AI ACT. A high-level AI policy must be approved and deployed through procedures, policies and protocols applicable within the organisation. The AI governance model will be part of the AI management system, as well as everything related to the AI procurement process and AI value chain.

8) Develop protocols for the use of AI (including generative AI).

In the framework of the AI management system, one of the key elements is the development of protocols for the ethical, legal and responsible use of artificial intelligence, including the use of generative AI. The organisation will have to decide what uses are authorised and not, if any, under what conditions, etc.

9) AI training and awareness for staff (AI literacy).

AI awareness and literacy of staff and others involved in the deployment of AI systems on their behalf must be promoted, taking into account their technical knowledge, experience, education and training and the context in which AI systems will be used, and considering the individuals or groups of individuals affected by the use of AI systems.

10) Establish a compliance roadmap.

Based on each organisation's starting point, establish an action plan to achieve compliance with the AI ACT, with deadlines and responsible parties.