

# Artificial Intelligence Regulatory Landscape

*Regulation around the Globe*



The background of the slide is a dark blue, futuristic digital landscape. It features a large, semi-transparent 'AI' logo in the upper left. The scene is filled with various glowing digital elements: a brain-like neural network structure, a 'NEWS' sign with an arrow, a user profile card with a lock icon, and several abstract data visualizations and network graphs. The overall aesthetic is high-tech and data-driven.

Regulatory landscape

---

Reactions from regulators about AI

---

Annex

*EU AI Act*

*US AI Bill or Rights*

*Supervision and AI initiatives*

*Abbreviations*

---

# 1 | Regulatory landscape

## General overview

Recent advances in Artificial Intelligence (AI) capabilities are pushing regulators<sup>1</sup> worldwide to establish regulations and different types of guidelines for the appropriate use of AI



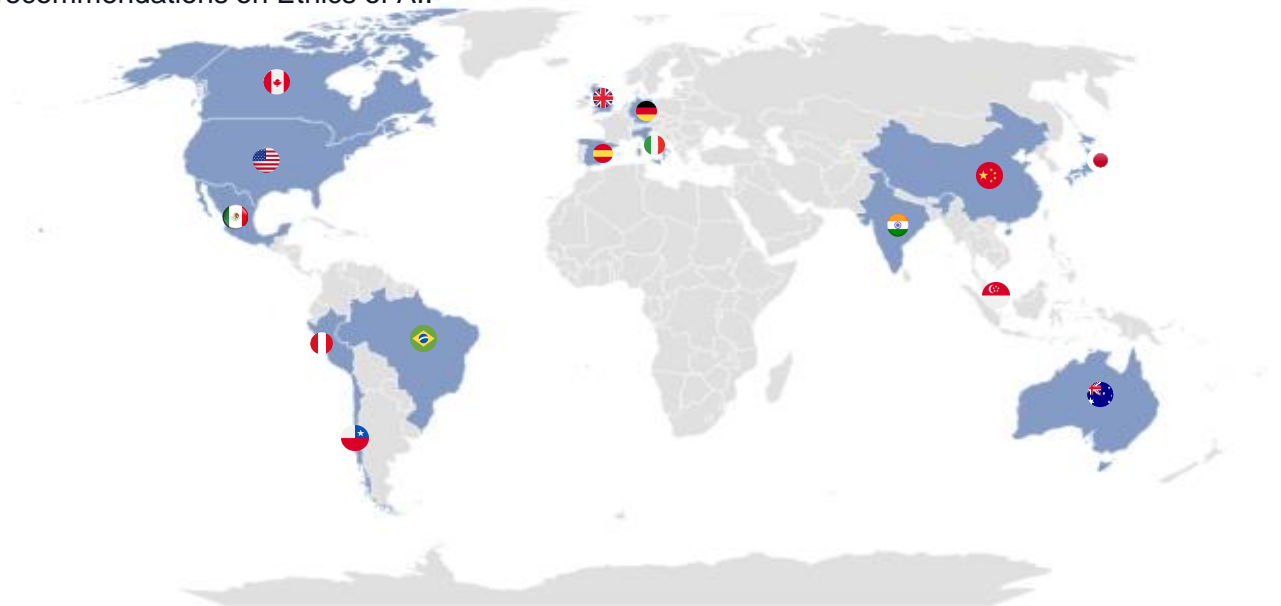
### General overview

- 1 Artificial Intelligence (AI) is the ability of a machine to display **human-like capabilities** such as reasoning, learning, planning and creativity<sup>2</sup>. An AI system means, thus, software that is developed with AI techniques and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with<sup>3</sup>.
- 2 Some AI technologies have been around for more than 50 years, but advances in computing power, the availability and storage capacity of enormous quantities of data and the development of new algorithms have led to **major AI breakthroughs in recent years**.
- 3 Most AI systems pose limited to no risk and can contribute to solving many societal challenges, certain AI systems create risks that **must be addressed** to avoid **undesirable outcomes**.



### Worldwide regulation

- The **European AI Act** is the first ever legal framework on AI. The US has also taken an approach towards AI through the issuance of the **AI Bill of Rights**. These two are considered the most relevant reference standards.
- Some **other countries** are also taking steps for regulating the AI (see map below).
- At **international level**, there are also some initiatives. For example, the OECD adopted some recommendations on AI, IOSCO adopted guidance on the use of AI, and UNESCO adopted recommendations on Ethics of AI.



(1) In the context of this technical note, the term "regulator" includes also supervisors and recognised standard setters.









(2) European Commission.

(3) EU IA Act. The set of what can be considered AI techniques is also described in this proposal for legislation, and includes machine learning approaches, logic and knowledge-based approaches, statistical approaches, Bayesian estimation, and search and optimization methods.

# 1 | Regulatory landscape

## EU and USA: Main characteristics of reference standards

**While the AI Act will set the legal framework in Europe, the AI Bill of Right in the US is a set of principles to help guide the design, use, and deployment of automated systems**

	 <b>AI Act (Europe)<sup>1</sup></b>	 <b>AI Bill of Rights (US)<sup>6</sup></b>
 <b>Objective</b>	<ul style="list-style-type: none"> <li>Improve the functioning of the internal market by laying down a <b>uniform legal framework</b> in particular for the development, marketing and use of AI in conformity with EU values</li> </ul>	<ul style="list-style-type: none"> <li>Help <b>guide</b> the <b>design, use, and deployment</b> of <b>automated systems</b> to protect the rights of the US public in the age of AI</li> </ul>
 <b>Publication Date</b>	<ul style="list-style-type: none"> <li><b>April 2021 (draft)<sup>2</sup></b></li> </ul>	<ul style="list-style-type: none"> <li><b>October 2022 (draft)</b></li> </ul>
 <b>Scope</b>	<ul style="list-style-type: none"> <li>AI system providers</li> <li>AI system users</li> <li>Deployers, importers and distributors of AI systems and affected persons located in the EU whose health, safety or fundamental rights were adversely impacted by the use of an AI system <sup>3</sup></li> </ul>	<ul style="list-style-type: none"> <li>Designers</li> <li>Developers</li> <li>Deployers of automated systems</li> </ul>
 <b>Enforceability</b>	<ul style="list-style-type: none"> <li><b>Mandatory</b> Act. when approved, it will be directly applicable in the Union's 27 countries</li> </ul>	<ul style="list-style-type: none"> <li><b>Voluntary</b> white paper</li> </ul>
 <b>Main content</b>	<p>It sorts AI applications into risk levels:</p> <ul style="list-style-type: none"> <li><b>Unacceptable risk</b> (prohibited practices)</li> <li><b>High-risk</b> (subject to a set of requirements and obligations to gain access to the EU market)</li> <li><b>Low or minimal risk</b></li> </ul> <p>Certain AI systems are subject to transparency obligations</p>	<p><b>5 principles:</b></p> <ul style="list-style-type: none"> <li>Safe and effective systems</li> <li>Algorithmic discrimination protection</li> <li>Data privacy</li> <li>Notice and explanation</li> <li>Human alternatives, consideration and fallback</li> </ul>
 <b>Next steps</b>	<ul style="list-style-type: none"> <li>Final version expected by the <b>end of 2023<sup>4</sup></b>, entering into force, in general terms, 2 years later (by the <b>end of 2025<sup>5</sup></b>)</li> </ul>	<ul style="list-style-type: none"> <li>There is no official calendar for next publications</li> </ul>

(1) For further information of the AI Act see [Annex 1](#)

(2) The first version of the Draft was published by the EC in April 2021. The proposal is now being discussed by the co-legislators (EP and the Council), so there are 3 different versions that have to be consolidated with the final version. Link to the [official comparison of versions](#). For more information go to [Annex 1](#).

(3) According to the last Draft of AI Act (May 2023)

(4) According to the [European Parliament](#)

(5) Provisions regarding high-risk systems and governance will apply 3 months following the entry into force

(6) For further information of the Blueprint for an AI Bill of Rights see [Annex 2](#)

# 1 | Regulatory landscape

## Main principles in AI Regulation

Proposals for AI regulation aim to address the potential risks, limitations, and ethical concerns associated with AI models while promoting their responsible development, implementation and use of these models

### Transparency and explainability

Explainability of AI systems and whether the AI model's decision-making process can be explained or whether it operates as a "black box." Draft regulation **requires ensuring the AI system outputs can be understood and evaluated by users and other stakeholders.**

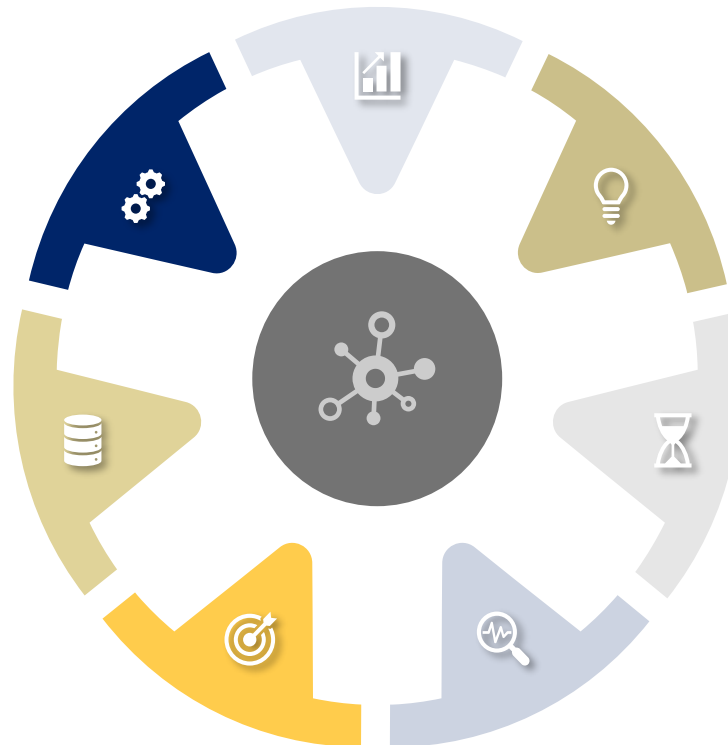
### Fairness and bias

Identifying and mitigating biases in AI models, and assessing whether the model's training data, algorithms, or decision-making processes introduce unfair advantages or disadvantages for specific groups or individuals. Draft regulation **may require measures to address bias, promote fairness, and ensure non-discriminatory outcomes.**

### Robustness and reliability

Assessing AI model performance under various conditions, including adversarial attacks, input variations, and edge cases. AI regulation may request **rigorous testing and validation, as well as human oversight and monitoring** to ensure that AI models behave reliably and consistently.

### Overview of AI regulatory requirements



### Accountability

This aspect involves **determining who is responsible for the actions and outcomes of AI systems**, including legal liability, roles of developers and operators, and **mechanisms for addressing harm or unintended consequences** caused by AI.

### Privacy and security

Evaluating how AI systems handle and protect user data during collection, storage, access, and the potential risks of unauthorized disclosure or misuse. Draft regulation may incorporate **requirements on data protection, informed consent, anonymization, and cybersecurity.**

### Ethics

Evaluating the broader ethical implications of AI models, such as the **impact on human rights, social values, and potential harm to individuals or society.**

### Risk assessment and governance

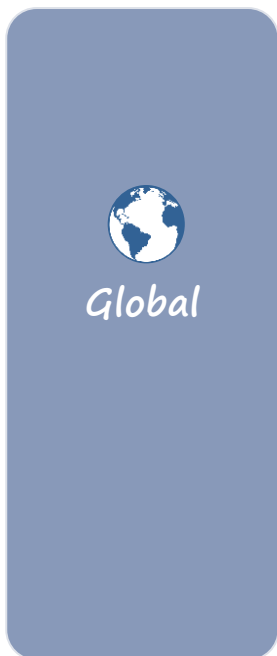
Assessing the potential risks associated with AI models, AI Regulation may **require mechanisms for risk assessment, certification, auditing, and regulatory oversight to ensure adequate governance and accountability** of AI systems.

# 2

## Reactions from regulators about AI

### Global

At global level, the OECD recommendations stand out as the first principles subscribed to by governments, and other organisms have also issued guidance and recommendations



#### Recommendation of the Council on Artificial Intelligence | OECD | May. 2019

G FV NB

This recommendation focuses on two building blocks. On the one hand, it sets out **Principles for responsible stewardship of trustworthy AI**: i) inclusive growth, sustainable development and well-being; ii) human-centred values and fairness; iii) transparency and explainability; iv) robustness, security and safety; and v) accountability. On the other hand, it sets out **recommendations for the integration of AI into national policies** and encourages international cooperation of governments for safe AI.

#### Use of AI and ML by market intermediaries and asset managers | IOSCO | Sep. 2021

I FV NB

The proposed guidance aims to help its members **regulate and supervise the use of AI and ML** by market intermediaries and asset managers. It also describe how regulators are addressing the challenges created by AI and ML and the guidance issued by supranational bodies in this area.

#### Ethics of Artificial Intelligence | UNESCO | Nov. 2021

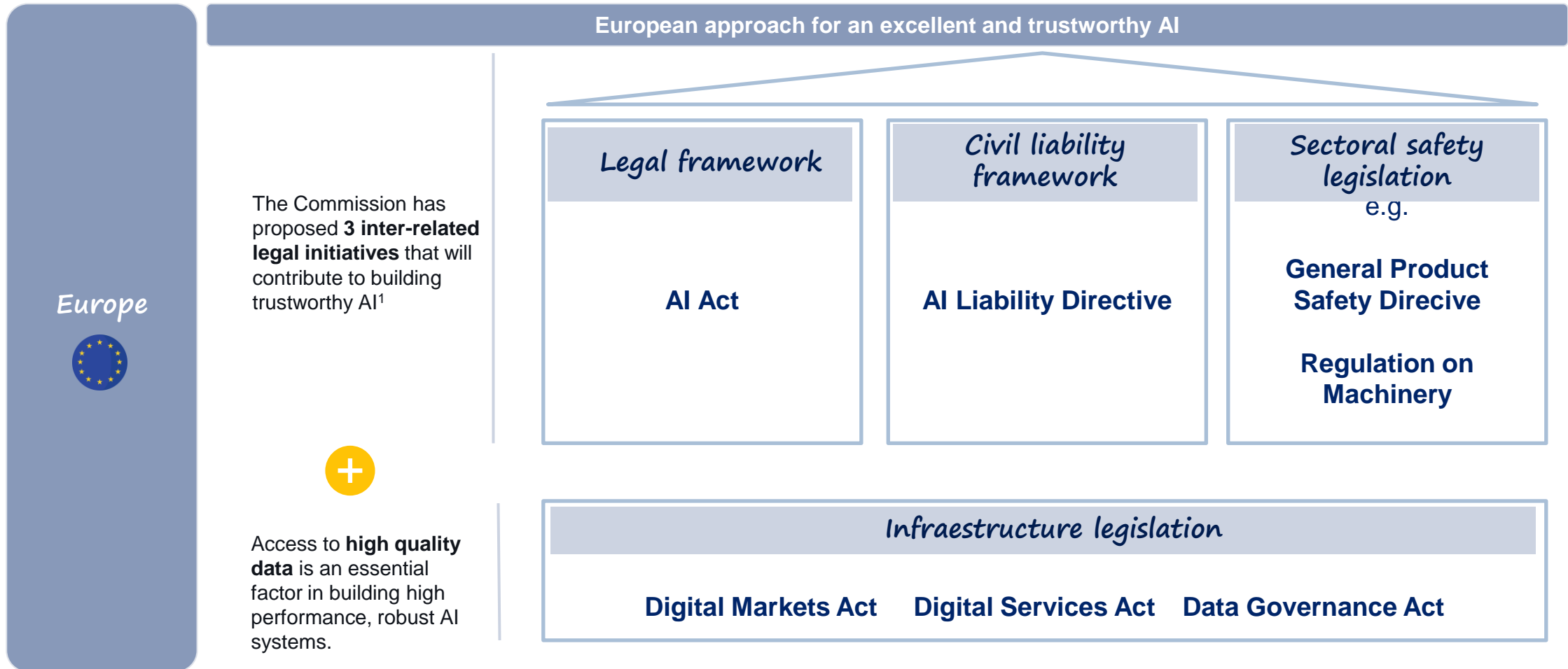
G FV NB

A set of values, **principles and recommendations on areas of policy action**, with the aim to provide a basis to make AI systems work for the good of humanity, individuals, societies and the environment and ecosystems, and to prevent harm, ensuring they contribute to a more inclusive, sustainable, and peaceful world.

# 2

## Reactions from regulators about AI Europe (incl. UK)

The European AI Strategy aims at making the EU a world-class hub for AI and ensuring that AI is human-centric and trustworthy



(1) In addition, in January 2023 the Committee on Artificial Intelligence of the Council of Europe has issued a draft Convention on Artificial Intelligence, human rights, democracy and the rule of law to establish principles, rules and rights aimed at ensuring that design, development and application of artificial intelligence systems is fully consistent with respect for human rights, the functioning of democracy and the observance of rule of law. This document will have to be negotiated in the Committee.

## 2 | Reactions from regulators about AI Europe (incl. UK)

In Europe, the Commission has proposed some inter-related legal initiatives that will contribute to building trustworthy AI and to address fundamental rights

Europe



### AI Act<sup>1</sup> | EC | Apr. 2021

G D B

The draft of Artificial Intelligence Regulation aims to **ensure a high level of trust in AI and its applications**, while laying the groundwork for innovation. It proposes a classification of AI practices into the following levels: i) prohibited practices; ii) high-risk AI systems; iii) low or minimal risk AI systems. Furthermore, it includes transparency obligations for certain AI systems that i) interact with humans; ii) are used to detect emotions; or iii) generate or manipulate content.

### Follow-up report on Machine Learning for IRB models | EBA | Aug. 2023

I FV NB

The EBA Follow-up report provides an overview of the current use cases of machine learning techniques for internal ratings-based models. Furthermore, it analyses the interaction of the use of these techniques in credit risk models with two other legal frameworks: the General Data Protection Regulation and the Artificial Intelligence Act.

### AI Liability Directive | EC | Sep. 2022

G D NB

Proposal for a Directive on adapting non contractual civil liability rules to AI has the following objectives: i) **adapt non-contractual civil liability rules** to AI; ii) **promote the uptake of AI** and address the risks associated with its use; iii) **identify and address the specific challenges posed by AI** to existing civil liability rules; iv) establish a **civil liability regime** for AI that is suitable and effective; and v) ensure that victims of damage caused by AI-enabled products and services have access to **fair and efficient compensation**.

### General Product Safety Regulation (GPSR) | EC | May 2023

I FV B

It seeks to address the **product safety challenges of emerging technologies**, including use of AI and connected devices, and to establish clear obligations for online marketplaces, which consumers increasingly use for their online purchases.

(1) ECB has published its position regarding the AI Act. See [Annex 4](#).



# 2 | Reactions from regulators about AI Europe (incl. UK)

## This regulations are complemented with additional Acts to address the infrastructure and data access and systems



### Regulation on Machinery | EP and Council | Jun. 2023

I FV B

The regulation introduces a new **legal framework to the European machinery and equipment industry**. Manufacturers, importers, and distributors of all types of machinery will have to comply with extensive new obligations. The rules introduce new safety requirements for autonomous machines, human-machine collaboration and, for the first time, the safe use of AI systems in machinery.

### The Digital Markets Act (DMA) | EC | Oct. 2022

G FV B

The DMA establishes a set of **clearly defined objective criteria** to identify “gatekeepers”. Gatekeepers are large digital platforms providing so called core platform services, such as for example online search engines, app stores, messenger services. Gatekeepers will have to comply with the do’s (i.e. obligations) and don’ts (i.e. prohibitions) listed in the DMA. Its impact in the digital sphere could have implications for how AI technologies are used and regulated in the context of online platforms and the digital economy in the EU.

### Digital Service Act (DSA) | EC | Oct. 2022

G FV B

The DSA establishes **legal rules for online platforms operating in the EU**, including social media platforms, online marketplaces, and search engines. It seeks to make online platforms more accountable for the content they host and to strengthen user rights and protections which could be threatened by AI.

### Data Governance Act | EP and Council | May 2022

G FV B

It supports the set-up and development of **common European data spaces**, meaning an internal market for data in which data could be used irrespective of its physical storage location in the EU in compliance with applicable law, which, inter alia, could be pivotal for the rapid **development of AI technologies**.

## 2 | Reactions from regulators about AI Europe (incl. UK)

In Spain, the Royal Decree establishing a Sandbox for compliance with the AI Act stands out. Furthermore, white papers were published in Germany and the UK. Italy has adapted measures to regulate the use of ChatGPT



### Royal Decree establishing a Sandbox for compliance with the AI Act | MINECO | May 2023

G D B

The purpose is to **create a controlled testing environment** to test measures foreseen in the proposed EU Regulation on AI (AI Act). Artificial intelligence systems that imply risks that can affect health, safety and fundamental rights of persons will be screened out in order to design the principles that rule their design, validation and monitorization to mitigate those risks. See [Annex 3](#).

### Royal Decree approving the constitution of the Spanish AI Supervisory Agency | Spanish Gov. | Aug. 2023

I FV B

The purpose of this Agency is to **supervise the use of AI systems** to protect fundamental rights and minimise risks and it will collaborate with national and European authorities. See [Annex 3](#).



### AI white paper to turbocharge growth | UK Government | Mar. 2023

G D NB

The white paper outlines **5 clear principles** that regulators such as the Equality and Human Rights Commission and Competition and Markets Authority, should consider to best facilitate the safe and innovative use of AI in the industries they monitor: i) safety; ii) security and robustness; iii) transparency and explainability; iv) fairness; v) accountability and governance; and vi) contestability and redress.



### FS2/23 AI and Machine Learning | Bank of England | Nov. 2023

G D NB

The Discussion Paper (DP) 5/22 on Artificial Intelligence (AI) and Machine Learning provides further insight and deepens the dialogue on how AI may affect their respective objectives for prudential and conduct of business supervision of financial firms. The Feedback Statement (FS) provides a summary of the responses to DP5/22 with the objective of acknowledging the responses, identifying issues and providing an overall summary of the response.

## 2 | Reactions from regulators about AI Europe (incl. UK)

In Spain, the Royal Decree establishing a Sandbox for compliance with the AI Act stands out. Furthermore, white papers were published in Germany and the UK. Italy has adapted measures to regulate the use of ChatGPT



### Germany's Ethics and AI White Paper | Ministry for Economic Affairs and Energy | Sep. 2020

G FV NB

The aim of the project was to **present the current status of standardization** in the interdisciplinary field of AI and ethics and to identify possible future fields of action for standardization. The project looks at the interrelationships between ethics and AI and what role technical standards and norms can play in this context. In doing so, the project focuses on the areas of autonomous machines and vehicles.



### Measures for the Management of Generative AI Services | GPDP | Mar. 2023

G D B

The Guarantor for the Protection of Personal Data (GPDP) banned the use of Chat GPT the 30<sup>th</sup> of March 2023. It was issued to **guarantee the protection of personal data** and requested a number of concrete measures from OpenAI with a deadline of 30 April for the implementation of most of them by OpenAI. Just over a month later, the company had implemented the measures and Chat GPT is again available in Italy.

# 2 | Reactions from regulators about AI America

In the US, two non-binding initiatives stand out: the AI Bill of Rights and the AI Risk Management Framework. Canada and Mexico have also binding Acts to regulate AI systems



### AI Bill of Rights | WH | Oct. 2022

G FV NB

It sets out **five principles or citizen rights regarding AI**, including safe and effective systems, protection against discrimination by algorithms, data privacy, notification and explanation, and evaluation and correction by a human in the event of AI failure (fallback). These principles include the explainability of AI models, which requires plain language documentation in addition to technically valid, meaningful and useful explanations, and demonstrably clear, timely, understandable and accessible notices of use.

### AI Risk Management Framework | NIST | Jan. 2023

G FV NB

This framework aims to **offer a resource** to the organizations designing, developing, deploying, or using AI systems to **help manage the many risks of AI** and promote trustworthy and responsible development and use these systems. This framework is risk-specific, and use-case agnostic, providing flexibility to organizations of all sizes and in all sectors. This NIST risk management framework must be complemented by the AI Bill of Rights Blueprint to effectively protect citizens, according to experts.



### Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI | WH | Oct. 2023

G FV NB

This publication explores the impact of AI across sectors and helps agencies and consumers to reap the benefits of AI while mitigating the risks. Executive departments and agencies should, as appropriate and in accordance with applicable law, adhere to principles, included in the executive order AI should be safe, should promote responsible innovation, competition and collaboration. In addition, the responsible development and use of AI requires a commitment to supporting American workers.



### The AI and Data Act (AIDA) | Government of Canada | Nov. 2022

G D B

The proposed AIDA aims to **regulate and standardise international and interprovincial trade in AI systems** by requiring certain persons to take measures to reduce the risk of harm and biased outcomes associated with high performance AI systems. The AIDA proposes: i) to ensure that high-impact AI systems meet the same expectations with respect to safety and human rights to which Canadians are accustomed; and ii) to prohibit reckless and malicious uses of AI that cause serious harm to Canadians and their interests through the creation of new criminal law provisions.

## 2 | Reactions from regulators about AI America

### Peru and Chile have developed non-binding standards, and Brazil has issued a Draft for the regulation of AI



#### Artificial Intelligence and Robotics Ethical Regulation Act | La Camara de Diputados | May 2023

G FV B

Its purpose is to regulate the use of AI and robotics for governmental, economic, commercial, administrative, communicational and financial purposes, so that their use is always based on ethical and legal ethics.



#### Law promoting the use of AI for the economic and social development of the country | Congress | Jul. 2023

G FV NB

It aims to promote the use of AI in the framework of the national process of digital transformation privileging the individual and respect for human rights in order to promote the economic and social development of the country, in a safe environment that guarantees its ethical, sustainable, transparent, replicable and responsible use.



#### National AI policy | MinCiencia | Oct. 2021

G FV NB

This policy contains the strategic guidelines that the country should follow in this area over the next 10 years with the aim of empowering people in the use and development of AI tools, and participating in the debate on their legal, ethical, social and economic consequences. This roadmap is built around **three axes: enabling factors, use, and development** of AI in Chile, and ethical and security aspects.



#### Draft for the regulation of AI | Brazilian Senate | May 2023

G D B

The Bill seeks to **mitigate the risks involved in the system** taking into account freedom, equality and free development of the personality. The regulation is based on **three central pillars**: i) guaranteeing the **rights of people** affected by the system; ii) classifying the **level of risk**; and iii) predicting **governance** measures for companies that provide or operate the AI system.

# 2

## Reactions from regulators about AI Asia and Oceania

China has recently published draft binding AI standards, India a Bill on data protection, and Japan has published non-binding guidelines with social principles for AI



### Generative AI Measures | CAC | Jul. 2023

G FV B

In order to promote the **healthy development and standardized application** of generative AI technology, the Cyberspace Administration of China (CAC), along with six other agencies, collaborated to issue the official Interim Administrative Measures for Generative Artificial Intelligence Services. As the first comprehensive AI regulation in China, the official Interim Administrative Measures for Generative Artificial Intelligence Services (known as the Generative AI Measures) encompass a wide array of subjects pertaining to the development and provision of generative AI services. These regulations are set to impact Chinese technological exports and global AI research networks.



### Draft Bill on Digital Personal Data Protection | India Government | Jul. 2023

G D B

The Draft Bill on Digital Personal Data Protection has been approved by Cabinet. Its provisions are relevant to AI and directly challenge processing personal data that is enabled by it.



### Governance guidelines for implementing the AI principles | METI | Jan. 2022

G FV NB

The document sets **seven social principles** for AI that are to be **implemented in the society** as a whole: i) human-centric; ii) education/literacy; iii) privacy protection; iv) ensuring security; v) fair competition; vi) fairness, accountability and transparency; and vii) innovation.

## 2 | Reactions from regulators about AI Asia and Oceania

### Singapore and Australia have set up a voluntary framework for AI



#### AI model governance framework | PDPC | Jan. 2020

G FV NB

It focuses primarily on four broad areas: i) **internal governance structures and measures**; ii) **human involvement** in AI-augmented decision-making; iii) **operations management**; and iv) **stakeholder interaction & communication**.



#### Australia's AI Ethics Principles | Department of Industry, Science and Resources | Nov. 2019

G FV NB

They will **help achieve safer, more reliable and fairer outcomes** for all Australians. Principles will also help to reduce the risk of negative impact on those affected by AI applications; and businesses and governments to practice the highest ethical standards when designing, developing and implementing AI. The principles are **voluntary** and are intended to be **aspirational and complementary** to the existing AI regulations, they are: i) human, societal and environmental wellbeing; ii) human-centred values; iii) fairness; iv) privacy protection and security; v) reliability and safety; vi) transparency and explainability; vii) contestability; and viii) accountability.

# A | Annex 1

## EU AI Act

The Proposal for a Regulation on AI tabled by the EC on 21 April 2021 set harmonised rules for the development, placement on the market and use of AI systems in the EU following a proportionate risk-based approach. In May 2023, the EP adopted the draft with several amendments to the EC proposal



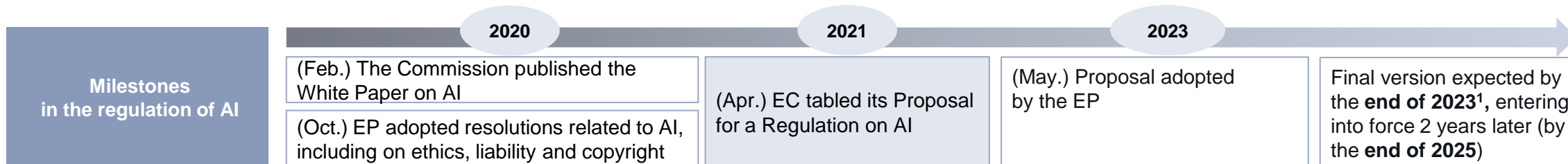
- On 19 February 2020 the EC published the White Paper on AI - **A European approach to excellence and trust**. The White Paper set out policy options on how to achieve the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of such technology.
- The EP and the EC have repeatedly expressed calls for legislative action to ensure a **well-functioning internal market** for AI systems where both **benefits and risks of AI are adequately addressed** at EU level.
- In October 2020, the EP **adopted a number of resolutions related to AI**, including on ethics, liability and copyright. In 2021, those were followed by resolutions on AI in criminal matters and in education, culture and added in the audio-visual sector.

In this context, in 2021, the EC put forward the **Proposal for a Regulation on AI**. The proposal has the following specific objectives:

- Ensure that AI systems placed and used on the EU market are safe and **respect existing law on fundamental rights and EU values**.
- Ensure **legal certainty** to facilitate investment and innovation in AI.
- Enhance **governance and effective enforcement of existing law on fundamental rights and safety requirements** applicable to AI systems.
- Facilitate the development of **a single market** for lawful, safe and trustworthy AI applications and **prevent market fragmentation**.

To achieve those objectives, this proposal presents a balanced and proportionate horizontal regulatory approach to AI.

In May 2023, the EP adopted this draft with several amendments to the EC proposal. The main amendments focus on: changes in the scope of application (expanded to any AI system), changes in the AI definition, and changes in the prohibited AI practices (amendments to the list of prohibited AI practices to include bans on intrusive and discriminatory uses of AI systems).





### Prohibited AI practices are mainly focused on subliminal techniques, exploiting vulnerabilities, misuse by public authorities and some biometric identification systems uses in public

#### Prohibited practices



#### The placing on the market, putting into service or use of AI ...

- ... that deploys **subliminal techniques** beyond a person's consciousness in order to materially distort a person's behaviour in a manner that causes or is likely to cause that person or another person significant harm.
- ... that exploits any of the **vulnerabilities of a specific group** of persons due to their age, physical or mental disability, in order to materially distort the behaviour of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm.
- ... that categorise natural persons according to sensitive or protected attributes or characteristics or based on the inference of those attributes or characteristics.
- ... by public authorities or on their behalf for the evaluation or **classification of the trustworthiness of natural persons** over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics, with the social score leading detrimental or unfavourable treatment of certain natural persons or whole groups thereof:
  - in social contexts which are unrelated to the contexts in which the data was originally generated or collected, or;
  - that is unjustified or disproportionate to their social behaviour or its gravity.
- ... for making risk assessments of natural persons or groups thereof in order to assess the risk of a natural person for offending or reoffending.
- ... that create or expand facial recognition databases through the untargeted scraping of facial images from the internet.
- ... to infer emotions of a natural person in the areas of law enforcement, border management, in workplace and education institutions.



#### Remote biometric identification systems

- ... The use of "real-time" **remote biometric identification systems in public spaces**.
- ... for the analysis of recorded footage of publicly accessible spaces through 'post' remote biometric identification systems, unless they are subject to a pre-judicial authorisation in accordance with Union law.

# A | Annex 1

## EU AI Act

The classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation. The classification as high-risk does not only depend on the function performed by the system, but also on the purpose and modalities for which that system is used

- There are specific rules for AI systems that create a high risk to the health and safety or fundamental rights of natural persons. In line with a risk-based approach, those **high-risk AI systems are permitted** on the European market subject to **compliance with certain mandatory requirements** and an ex-ante conformity assessment.

Classification of AI systems as high risk

- An AI system shall be considered **high-risk** where all the following conditions are fulfilled:
  - The AI system is intended to be used as a **safety component** of a product, or the AI system is itself a product, **covered by the EU harmonisation legislation**.
  - The **product whose safety component is the AI system**, or the AI system itself as a product, is required to **undergo a third-party conformity** assessment related to risks for health and safety, with a view to the placing on the market or putting into service of that product pursuant to the EU harmonisation legislation.
  - In addition to the previous high-risk AI systems, the regulation provides a list of AI systems, with mainly **fundamental rights implications**, that shall be considered high risk if they pose a significant risk of harm to the health, safety or fundamental rights of natural persons.
- Where providers consider that their AI system does not pose a significant risk, they shall submit a reasoned notification to the National Supervisory Authority. the National Supervisory Authority shall review and reply, directly or via the AI Office, within 3 months if they deem the AI system to be misclassified.
- Providers that misclassify their AI system and place it on the market before the deadline for objection by National Supervisory Authorities shall be responsible and be subject to fines.

# A Annex 1 EU AI Act

The intended purpose of the high-risk AI system and the risk management system shall be taken into account when ensuring compliance with those requirements. The providers of high-risk AI systems shall fulfill the obligations required

## Legal requirements for high-risk AI systems

- The proposed **minimum requirements are largely consistent with other international recommendations and principles**, which ensure that the proposed AI framework is compatible with those adopted by the EU's international trade partners.
- A risk management system **shall be established, implemented, documented and maintained in relation to high-risk AI systems**.
- The risk management system shall consist of a **continuous iterative process run throughout the entire lifecycle** of a high-risk AI system. It shall comprise the following steps:

Identification and analysis of the known and foreseeable risks associated with each high-risk AI system.

Estimation and evaluation of the risks that may emerge when the high-risk AI system is used in accordance with its intended purpose.

Evaluation of other possibly arising risks based on the analysis of data gathered from the post-market monitoring system.

Adoption of suitable risk management measures.

## Obligations of providers of high-risk AI systems

- Providers of high-risk AI systems shall:
  - **Ensure that their high-risk AI systems are compliant with the legal requirements.**
  - Indicate their name, registered trade name or registered trade mark, and their address and contact information on the high-risk AI system
  - Ensure that natural persons to whom human oversight of high-risk AI systems is assigned are specifically made aware of the risk of automation or confirmation bias;
  - Provide specifications for the input data, or any other relevant information
  - Have a **quality management** system in place.
  - Draw-up the **technical documentation** of the high-risk AI system.
  - When under their control, keep the logs **automatically generated** by their high-risk AI systems.
  - Ensure that the high-risk AI system undergoes the **relevant conformity assessment procedure** prior to its placing on the market or putting into service.
  - Comply with the **registration obligations**.
  - Take the **necessary corrective actions**, if the high-risk AI system is not in conformity with the legal requirements.
  - Inform the NCAs of the Member States in which they made the AI system available of the non-compliance and of any **corrective actions taken**.

## A framework is needed for the notified bodies to be involved as independent third parties in conformity assessment procedures

- The framework for notification authorities, procedures and bodies is divided in the following sections:

### Notifying authorities

- Each Member State shall designate or establish a **notifying authority responsible for setting up and carrying out the necessary procedures** for the assessment, designation and notification of conformity assessment bodies and for their monitoring. These procedures shall be developed in cooperation between the notifying authorities of all Member States.
- Notifying authorities **shall be established, organised and operated** in such a way that no conflict of interest arises with conformity assessment bodies and the objectivity and impartiality of their activities are safeguarded.
- Notifying authorities shall **not offer or provide any activities that conformity assessment bodies perform or any consultancy services** on a commercial or competitive basis.
- Notifying authorities shall have a sufficient number of competent personnel at their disposal for the proper performance of their tasks. Where applicable, competent personnel shall have the necessary expertise, such as a degree in an appropriate legal field, in supervision of fundamental rights.

### Notification procedure

- Notifying authorities shall notify the EC and the other Member States using the **electronic notification** tool developed and managed by the EC of each conformity assessment body.
- The notification shall include **full details of the conformity assessment activities**, the conformity assessment module or modules and the AI technologies concerned., as well as the relevant attestation of competence.

### Notified bodies

- Notified bodies shall perform the conformity assessment of the high-risk AI systems and **satisfy the organisational, quality management, resources and process requirements** that are necessary to fulfil their tasks as well as the minimum cybersecurity requirements set out for public administration entities identified as operators of essential services pursuant NIS2,
- Notified bodies shall be **independent of the provider** of a high-risk AI system in relation to which it performs conformity assessment activities This shall not preclude the use of assessed AI systems that are necessary for the operations of the conformity assessment body or the use of such systems for personal purposes.

**There is a conformity assessment procedure for each type of high-risk AI system. The procedure has the following key elements: harmonized standards, conformity assessments, certificates and registration**

## Key elements

- The conformity assessment approach aims to minimise the burden for economic operators as well as for notified bodies, whose capacity needs to be progressively ramped up over time

1

### Harmonised standards

- They aim to minimise the burden for economic operators as well as for notified bodies, whose capacity needs to be progressively ramped up over time.
- **High-risk AI systems** which are in **conformity with harmonised standards** or parts thereof shall be presumed to be **in conformity with the legal requirements** for high-risk AI systems.
- The EC shall issue standardisation requests covering all requirements of this Regulation

2

### Conformity assessment

- The provider shall follow the **conformity assessment procedure** based on **internal control** or follow the conformity assessment procedure based on **assessment of the quality management system** of the technical documentation, with the involvement of a notified body.

3

### Certificates

- Certificates issued by notified bodies shall be **drawn-up in an official Union language** determined by the Member State in which the notified body is established or in an official Union language otherwise acceptable to the notified body.
- Certificates shall be valid for the period they indicate, which shall not exceed **five years**.

4

### Registration

- Before placing on the market or putting into service a high-risk AI system referred the provider or, where applicable, the authorised representative shall register that system in the **EU database**.
- Before putting into service or using a high-risk AI the following categories of deployers shall register the use of that AI system in the EU database: deployers who are public authorities or Union institutions; deployers who are undertakings designated as a gatekeeper

### Certain AI systems require transparency obligations so that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use

Transparency obligations will apply for systems that:



- Interact with humans. Providers shall **ensure that AI systems are designed and developed in such a way that persons are informed that they are interacting with an AI system**.



- Are used to detect emotions or determine association with (social) categories based on biometric data. Users of an **emotion recognition system or a biometric categorisation** system shall inform of the operation of the system the natural persons exposed thereto.



- Generate or manipulate content ('deep fakes'). Users of an AI system that generates or **manipulates image, audio or video content** that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful, shall disclose that the content has been artificially generated or manipulated.



- However, the **transparency obligations** in relation to the systems that interact with **humans shall not apply** where the use is authorised by law to detect, prevent, investigate and prosecute **criminal offences**.

# A | Annex 1

## EU AI Act

**To keep a legal framework that is sustainable over time and is innovation-friendly, the EC encourages to set up regulatory sandboxes and sets a basic framework in terms of governance, supervision and liability**

---



Member States shall establish at least one AI regulatory sandbox at national level fosters innovation and facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan. Additional AI regulatory sandboxes at regional or local levels may also be established.

---



This is expected to take place under the direct **supervision and guidance** by the **CAs** with a view to ensuring **compliance with the requirements of this Regulation** and, where relevant, other Union and Member States legislation supervised within the sandbox.

---



All the **authorities competent in the protection of data** used in the innovative AI systems must be **included** in the operation of the **AI regulatory sandbox** of the same, which will be supervised by the member states.

---



**Any significant risks** to health and safety and fundamental rights, democracy and rule of law, health and safety or the environment **identified during the development and testing** of such systems shall result in immediate **mitigation**. CAs shall have the power to temporarily or permanently suspend the testing process, or participation in the sandbox if no effective mitigation is possible and inform the AI office of such decision.

---



Any **member state** establishing AI regulatory sandboxes is expected to **cooperate under the framework of the European Artificial Intelligence Board** through **annual reports**, starting one year after the establishment of the sandbox and then every year until its termination and a final report. Those reports shall provide information on the progress and results of the implementation of those sandboxes including experience obtained in all areas. Those annual reports or abstracts thereof shall be made available to the public, online.

---



Member States are expected to undertake measures to **reduce the regulatory burden on small and medium-sized enterprises SMEs and start-ups**.

### A governance system is established at both the Union and National level for the purpose of directing, controlling and executing this Regulation

#### Union Level

At Union level, the "**European Artificial Intelligence Board**" (the 'Board') is established for the purpose of **providing advice and assistance to the EC**. In order to coordinate, contribute and assist with matters covered by this Regulation.

#### Structure of the Board

The Board is expected to be composed of the **national supervisory authorities, and the European Data Protection Supervisor**. It should adopt rules of procedure by a **simple majority of its members**, following the consent of the EC. The rules of procedure shall also contain the **operational aspects related to the execution of the Board's tasks**.

The Board is expected to be chaired by the EC, which will provide administrative and analytical support for the Board's activities pursuant to this Regulation.

#### Tasks of the Board

- Monitor and ensure the effective and consistent application of this Regulation
- serve as a mediator in discussions about serious disagreements regarding the application of the Regulation
- contribute to the effective cooperation with the competent authorities of third countries and with international organisations.
- Collect and share expertise and best practices among Member States

The **European Data Protection Supervisor** will act as the competent authority for the **supervision** of the Union institutions, agencies and bodies when they fall **within the scope of this regulation**.

#### National Level

The **competent national authorities** are expected to be **designated** by each Member State for the purpose of **ensuring the implementation and enforcement** of this Regulation. Such authorities will be organized in such a way as to ensure the objectivity and impartiality of their activities and tasks.

Member States shall make publicly available and communicate to the AI Office and the Commission the national supervisory authority and information on how it can be contacted.



# A | Annex 1

## EU AI Act

The Regulation establishes the monitoring and reporting obligations for providers of AI systems with regard to post-market monitoring and reporting and investigating on AI-related incidents and malfunctioning controlled by Market surveillance authorities

### EU Database

To facilitate the monitoring work of the EC and national authorities, an EU-wide database is established **high-risk AI systems with mainly fundamental rights implications**. The database will be operated by the EC and **provided with data by the providers of the AI systems**, who will be required to register their systems before placing them on the market or otherwise putting them into service.

### Post-Marketing

#### Post-Market Monitoring

Providers are expected to **establish and document a post-market monitoring system** proportionate to the nature of the AI technologies and the risks of the high-risk AI system.

This system should actively and systematically **collect, document and analyze relevant data provided by users** on the **performance of high-risk AI systems** throughout their lifetime, and **allow the provider to evaluate the continuous compliance with the high-risk AI systems requirements**.

The EC is expected to **adopt an implementing act laying down detailed provisions** establishing a template for the post-market monitoring plan and the list of elements to be included in the plan.

#### Reporting incidents and malfunctions

**Providers and, where deployers have identified a serious incident, of high-risk AI systems** placed on the EU market should **report any serious incident** of those systems which constitutes a breach of obligations under Union law intended to protect fundamental rights to the national supervisory authority of the Member States where that incident or breach occurred.

#### Enforcement

**Market surveillance authorities would control the market and investigate compliance with the obligations and requirements for all high-risk AI systems already placed on the market.**

### Codes of conduct, which aim to encourage providers of non-high-risk AI systems to apply voluntarily the mandatory requirements for high-risk AI systems

#### Providers of non-high-risk AI systems

- Providers of non-high-risk AI systems may create and implement the codes of conduct themselves. Codes of conduct may include **voluntary commitments related to:**



- Environmental sustainability.



- Accessibility for persons with disability.






- Stakeholders' participation in the design and development of AI systems.






- Diversity of development teams.

- The **EC and the Member States** shall encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to AI systems other than high-risk AI systems.
- Codes of conduct may be drawn up by individual providers of AI systems or by organisations representing them or by both, including with the involvement of users and any interested stakeholders and their representative organisations. Codes of conduct may cover one or more AI systems taking into account the similarity of the intended purpose of the relevant systems.
- The **EC and the Board** shall take into account the specific interests and needs of the **small-scale providers and start-ups** when encouraging and facilitating the drawing up of codes of conduct.

There are three different AI Act proposals being discussed amongst the European Commission, European Parliament and the Council, which have to be consolidated into a final version

	AI Act (EC proposal) 	AI Act (EP mandate) 	AI Act (Council mandate) 
<b>Objective</b>	Original text  <ul style="list-style-type: none"> <li>Improve the functioning of the internal market by laying down a uniform legal framework in particular for the development, marketing and use of artificial intelligence in conformity with Union values.</li> </ul>	Modification of the objective of the regulation  <ul style="list-style-type: none"> <li>Promote the uptake of human centric and trustworthy AI to ensure a high level of protection of health, safety, fundamental rights, democracy and rule of law and the environment from harmful effects of AI systems in the EU while supporting innovation and improving the functioning of the internal market.</li> </ul>	Unchanged
<b>Scope of application</b>	Original text  <ul style="list-style-type: none"> <li>(1) AI system providers.</li> <li>(2) Users of AI systems.</li> <li>(3) Providers and users that are located in a third country</li> </ul>	Modification of the scope by adding the deployers of AI systems and removing users  <ul style="list-style-type: none"> <li>(1) [...]</li> <li><b>(NEW) AI systems deployers.</b></li> <li>(3) Providers and deployers of AI systems.</li> </ul>	Modification of the scope by adding importers and distributors  <ul style="list-style-type: none"> <li>(1) [...]</li> <li>(2) [...]</li> <li>(3) Providers and users of AI systems</li> <li><b>(NEW) Importers and distributors of AI systems</b></li> </ul>
<b>Definition of AI System</b>	Original text  <ul style="list-style-type: none"> <li>a) Machine learning approaches, including supervised, unsupervised and reinforcement learning [...]</li> <li>(b) Logic- and knowledge-based approaches [...]</li> <li>(c) Statistical approaches [...]</li> </ul>	Amendment of the definition of AI systems to align it with the definition agreed by the OECD (limited to ML approaches)  <ul style="list-style-type: none"> <li>AI system means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments.</li> </ul>	Amendment of the definition of AI narrowing it to systems developed through ML approaches and logic- and knowledge-based approaches  <ul style="list-style-type: none"> <li>AI system means a system that is designed to operate with elements of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of objectives using machine learning and/or logic- and knowledge based approaches.</li> </ul>

There are three different AI Act proposals being discussed amongst the European Commission, European Parliament and the Council, which have to be consolidated into a final version

	AI Act (EC proposal) 	AI Act (EP mandate) 	AI Act (Council mandate) 
	Original text	Amended the list of AI systems prohibited in the EU.	Minor changes
Prohibited systems	<p>The placing on the market putting into service or use of AI:</p> <ul style="list-style-type: none"> <li>• (1) That deploys subliminal techniques beyond a person's consciousness.</li> <li>• (2) That exploits any of the vulnerabilities of a specific group of persons due to their age, physical or mental disability.</li> <li>• (3) for the evaluation or classification of the trustworthiness of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics.</li> <li>• (4) The use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement, unless and in as far as such use is strictly necessary for one of the objectives defined:</li> </ul>	<p>The placing on the market putting into service or use of AI:</p> <ul style="list-style-type: none"> <li>• (1) That deploys subliminal techniques beyond a person's consciousness or purposefully manipulative with the objective of materially distorting a person's behaviour.</li> <li>• (2) That exploits any of the vulnerabilities of a person or a specific group of persons, including characteristics of such person's known or predicted personality traits or social or economic situation.</li> <li>• <b>(NEW) use of biometric categorisation systems that categorise natural persons according to sensitive or protected attributes or characteristics or based on the inference of those attributes or characteristics.</b></li> <li>• (3) for the social scoring evaluation or classification of natural persons or groups thereof over a certain period of time based on their social behaviour.</li> <li>• <b>(NEW) for making risk assessments of natural persons.</b></li> <li>• <b>(NEW) that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage.</b></li> <li>• (4) the use of 'real-time' remote biometric identification systems in publicly accessible spaces.</li> </ul>	<ul style="list-style-type: none"> <li>• (1) [...]</li> <li>• (2) that exploits any of the vulnerabilities of a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behaviour of a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person physical or psychological harm;</li> <li>• (3) for the evaluation or classification of natural persons over a certain period of time based on their social behaviour or known or predicted personal or personality characteristics.</li> <li>• (4) [...]</li> </ul>



# Annex 1

## EU AI Act: comparison of versions<sup>1</sup>



There are three different AI Act proposals being discussed amongst the European Commission, European Parliament and the Council, which have to be consolidated into a final version

	AI Act (EC proposal)	AI Act (EP mandate)	AI Act (Council mandate)
	Original text	Adds the additional requirement that the systems must pose a 'significant risk' to qualify as high-risk.	Focus on the third-party conformity assessment and clarification of the requirements of AI systems
High-risk AI systems	<ul style="list-style-type: none"> <li>(1) AI system shall be considered high-risk where both of the following conditions are fulfilled: i) the AI system is intended to be used as a safety component of a product, or is itself a product; ii) the product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment</li> <li>(2) High-risk systems are also the ones referred in Annex III of this Regulation</li> </ul>	<ul style="list-style-type: none"> <li>(1) [...]</li> <li>(2) High-risk systems are also the ones referred in Annex III of this Regulation if they pose a significant risk of harm to the health, safety or fundamental rights of natural persons.</li> <li><b>(NEW) Where providers falling use cases referred to in Annex III consider that their AI system does not pose a significant risk they shall submit a reasoned notification to the national supervisory authority</b></li> </ul>	<ul style="list-style-type: none"> <li>(1) If it is required to undergo a third-party conformity assessment</li> <li>(2) An AI system intended to be used as a safety component of a product covered by the legislation shall be considered as high risk if it is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product.</li> </ul>
High-risk systems Annex III	<ul style="list-style-type: none"> <li>(5) Access to and enjoyment of essential private services and public services: a[...] (b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;</li> </ul>	<ul style="list-style-type: none"> <li>(5) [...] (b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems used for the purpose of detecting financial fraud;</li> </ul>	<ul style="list-style-type: none"> <li>(5) [...] (b) AI systems intended to be used to evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by providers that are micro and small-sized Enterprises;</li> </ul>
Date of application	Original text	Unchanged	Delays the date of implementation
	<ul style="list-style-type: none"> <li>This Regulation shall apply from 24 months following the entering into force of the Regulation</li> </ul>	<ul style="list-style-type: none"> <li>[...]</li> </ul>	<ul style="list-style-type: none"> <li>This Regulation shall apply from 36 months following the entering into force of the Regulation</li> </ul>

# A | Annex 2

## US AI Bill or Rights

The Blueprint for an AI Bill of Rights is a set of five principles and associated practices to help guide the design, use, and deployment of automated systems to protect the rights of the US public in the age of AI

### Context

The White House Office of Science and Technology Policy published the Blueprint for an AI Bill of Rights in October 2022 which is an exercise in envisioning a future where the US public is protected from the potential harms, and can fully enjoy the benefits, of automated systems. It describes principles that can help ensure these protections. Some of these protections are already required by the US Constitution or implemented under existing US laws.

### Principles

1

Safe and effective systems



Automated systems should be developed with **consultation** from **diverse communities, stakeholders, and domain experts** to identify concerns, risks, and potential impacts of the system.

2

Algorithmic discrimination protections



Designers, developers, and deployers of automated systems should take **proactive and continuous measures** to **protect individuals and communities from algorithmic discrimination** and to use and design systems in an equitable way.

3

Data privacy



Designers, developers, and deployers of automated systems should seek for permission and **respect people's decisions regarding** collection, use, access, transfer, and deletion of **their data** in appropriate ways.

4

Notice and explanation



Designers, developers, and deployers should provide a clear description of: i) the overall system functioning and the role automation plays; ii) notice that such systems are in use; iii) the individual or organization responsible for the system.

5

Human alternatives, consideration  
and fallback



Opting for automated systems in favor of a human alternative, **where appropriate**. Appropriateness should be determined based on reasonable expectations in a **given context** and with a focus on ensuring **broad accessibility** and protecting the public from especially harmful impacts.

### Automated systems should be developed with consultation from diverse communities, stakeholders, and domain experts to identify concerns, risks, and potential impacts of the system

<b>1</b> <i>Protect the public from harm in a proactive and ongoing manner</i>	
Consultation	Public should be consulted in the <b>design, implementation, deployment</b> , acquisition, and maintenance phases of <b>automated system</b> development.
Testing	Undergo <b>extensive testing before deployment</b> . This testing should follow <b>domain-specific best practices</b> .
Risk identification and mitigation	Before deployment, and in a proactive and ongoing manner, <b>potential risks should be identified and mitigated</b> .
Ongoing monitoring	Ongoing monitoring procedures to <b>ensure that performance does not fall below an acceptable level over time</b> , based on changing real-world conditions or deployment contexts, post-deployment modification, or unexpected conditions.
Clear organizational oversight	Include <b>clearly-stated governance procedures</b> before deploying the system, as well as <b>responsibility</b> of specific individuals or entities to oversee ongoing assessment and mitigation.
<b>2</b> <i>Avoid inappropriate, low-quality, or irrelevant data use and the compound harm of its reuse</i>	
Relevant and high-quality data	<b>Data used as part of any automated system's</b> creation, evaluation, or deployment should be relevant, of high quality, and tailored to the task at hand.
Carefully track and review derived data sources	<b>Data that is derived from other data</b> through the use of algorithms, such as data derived or inferred from prior model outputs, should be identified and tracked.
Data reuse limits in sensitive domains	<b>Data reuse</b> , and especially data reuse in a new context, can result in the <b>spreading and scaling of harms</b> . Accordingly, such data should be subject to extra oversight to ensure safety and efficacy.
<b>3</b> <i>Demonstrate the safety and effectiveness of the system</i>	
Independent evaluation	Independent evaluators, should be given <b>access to the system and samples</b> of associated data, in a manner consistent with privacy, security, law, or regulation in order to perform such evaluations.
Reporting	Provide <b>regularly-updated reports</b> , including: i) an overview of the system; ii) system goals; iii) any human-run procedures.

## Algorithms should not be discriminatory, and systems should be used and designed in an equitable way

### 1 *Protect the public from algorithmic discrimination in a proactive and ongoing manner*

Proactive assessment of equity in design

**Review potential input data**, associated historical context, accessibility for people with disabilities, and societal goals to identify potential discrimination and effects on equity resulting from the introduction of the technology.

Representative and robust data

Any data used should be **representative of local communities**, reviewed for bias based on the historical and societal context of the data, and sufficiently robust to identify and help to mitigate biases and potential harms.

Guarding against proxies

**Identify proxies** by testing for correlation between demographic information and attributes in any data used.

Ensuring accessibility during design, development & deployment

Consideration of a **variety of disabilities**, adherence to relevant accessibility standards, and user experience research to identify and address any accessibility barriers to the use or effectiveness of the automated system.

Disparity assessment

Test systems by using **demographic performance measures**, overall and subgroup parity assessment, and calibration measures to assess whether the system components produce disparities.

Disparity mitigation

Evaluate multiple models and select the one that has the **least adverse impact**, modify data input choices, or identify a system with fewer disparities. If this is not possible, then the use of the automated system should be reconsidered.

Ongoing monitoring and mitigation

**Regularly monitor automated systems** to assess algorithmic discrimination that might arise from unforeseen interactions of the system with inequities not accounted.

### 2 *Demonstrate that the system protects against algorithmic discrimination*

Independent evaluation

**Allow independent evaluation** of potential algorithmic discrimination caused by automated systems they use or oversee.

Reporting

Provide reporting of an appropriately designed algorithmic impact assessment, with clear **specification** of who performs the assessment, who evaluates the system, and how **corrective actions** are taken in response to the assessment.



## Users should be protected from abusive data practices via built-in protections and have agency over how data about the user is used

1

*Protect the privacy by design and by default*

Privacy by design and by default

Automated systems should be **designed** and built with privacy protected by default.

Data collection and use-case scope limits

**Data collection** should be **limited in scope**, with specific, narrow identified goals.

Risk identification and mitigation.

**Attempt** to proactively **identify harms** and seek to manage them when collecting, using or storing sensitive data.

Privacy-preserving security

Entities creating, using, or governing automated systems should **follow privacy** and security best practices designed to ensure data and metadata do not leak beyond the specific consented use case.

2

*Protect the public from unchecked surveillance*

Heightened oversight of surveillance

Surveillance or monitoring systems should be subject to **heightened oversight** that includes at a minimum assessment of potential harms during design.

Limited and proportionate surveillance

**Surveillance should be avoided** unless it's necessary to achieve a legitimate purpose and it's proportionated to the need.

Scope limits on surveillance to protect rights and democratic values

**Civil liberties** and **civil rights** must **not** be **limited** by the **threat of surveillance** or harassment facilitated or aided by an automated system.

## Users should be protected from abusive data practices via built-in protections and have agency over how data about the user is used

### 3 Provide the public with mechanisms for appropriate and meaningful consent, access, and control over their data

Use-specific consent.

Consent practices should **not allow for abusive surveillance** practices.

Brief and direct consent requests.

Short, plain language consent requests should be used so that users understand for what **use contexts, time span, and entities** they are providing data and metadata consent.

Data access and correction.

People whose data is collected, used, shared, or stored by automated systems should be able to **access data and metadata** about themselves.

Consent withdrawal and data deletion.

Entities should **allow withdrawal** of data access consent.

Automated system support.

Entities designing, developing, and deploying automated systems should **establish and maintain the capabilities** that will allow individuals to use their own automated systems.

### 4 Demonstrate that data privacy and user control are protected

Independent evaluation.

Entities should allow **independent evaluation** of the claims made regarding data policies.

Reporting

When members of the public wish to know what data about them is being used in a system, the entity responsible for the development of the system should **respond quickly** with a report on the data it has collected or stored about them.

## Users should be notified of the use and understand how and why the automated system contributes to outcomes that impact them

### 1 Provide clear, timely, understandable, and accessible notice of use and explanations

Generally accessible plain language documentation

The entity responsible for using the automated system should ensure that **documentation** describes the overall system.

Accountable

Notices should clearly identify the **entity responsible** for designing each component of the system and the entity using it.

Timely and up-to-date

Users should **receive notice of the use of automated systems** in **advance** of using or while being impacted by the technology.

Brief and clear

**Notices** and **explanations** should **be assessed**, such as by research on users' experiences, to ensure that the people using or impacted are able to easily find notices and explanations, read them quickly, and understand and act on them.

### 2 Provide explanations as to how and why a decision was made or an action was taken by an automated system

Tailored to the purpose

Explanations should be **tailored to the specific purpose** for which the user is expected to use the explanation, and should clearly state that purpose.

Tailored to the target of the explanation

Explanations should be targeted to **specific audiences** and clearly state that audience. An explanation provided to the subject of a decision might differ from one provided to an advocate, or to a domain expert or decision maker.

Tailored to the level of risk

An **assessment** should be done to determine the level of risk of the automated system.

Valid

The explanation provided by a system should **accurately reflect the factors** and the influences that led to a **particular decision**, and should be meaningful for the particular customization based on purpose, target, and level of risk.

### 3 Demonstrate protections for notice and explanation

Reporting

**Document** the determinations made based on the above considerations.

### Users should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter

#### 1 Provide a mechanism to opt out from automated systems in favor of human alternative

Brief, clear, accessible notice and instructions.	Those impacted by an automated system <b>should be given a brief</b> , clear notice that they are entitled to opt-out, along with clear instructions for how to opt-out.
Human alternatives provided when appropriate	When automated systems make up part of the attainment process, <b>alternative timely human-driven</b> processes should <b>be provided</b> .
Timely and not burdensome human alternative	<b>Opting out</b> should be <b>timely</b> and not unreasonably burdensome.

#### 2 Provide timely human consideration and remedy by a fallback and escalation system if an automated system fails

Proportionate	The availability of <b>human consideration</b> and fallback should be <b>proportionate</b> to the potential of the automated system.
Accessible	Mechanisms for human consideration and fallback should be <b>easy to find</b> .
Convenient	Mechanisms for human consideration and fallback <b>should not be unreasonably burdensome</b> as compared to the automated system's equivalent.
Equitable	Consideration should be given to <b>ensure outcomes of the fallback</b> and escalation system are equitable.
Timely	Human consideration and fallback are only useful if they are conducted and concluded in a <b>timely manner</b> .
Effective	Organizational structure surrounding processes for consideration and fallback should be designed so that if the human <b>decision-maker</b> determines that it should be overruled, the new decision <b>will be effectively enacted</b> .
Maintained	Human consideration and fallback process and any associated automated processes should be <b>maintained and supported</b> as long as the relevant automated system continues to be in use.

# A | Annex 3

## Supervision and initiatives - ECB Position (1/2)



### In the banking industry, the ECB as prudential supervisor for credit institutions has published its position regarding the AI Act requirements

#### General observations

- The ECB welcomes the objective of the proposed regulation and **acknowledges the importance of setting armonised requirements for AI systems**, especially in the banking sector.
- The proposed regulation **integrates obligations and procedures established in the Directive 2013/36/EU (CRD)** with regard to risk management and governance, however, **further clarification is requested to clarify supervisory expectations on internal governance**.
- The ECB considers the obligation for provider to have a **quality management system and to monitor the AI systems is already fulfilled by complying with the CRD**.
- The proposed regulation should be without prejudice to the more specific or stringent prudential obligations of credit institutions set out in sectoral regulation and supplemented by supervisory guidance (e.g., effective control of outsourcing as specified in the CRD).
- **The ECB follows a technology-neutral approach.**
- **The ECB's role under the proposed regulation should be clarified:** (1) the ECB's prudential supervisory competences generally, and in relation to market surveillance and conformity assessment; and (2) the application of the proposed regulation to the performance of the ECB's tasks under the Treaty

#### Classification of AI systems

- Under the proposed regulation, credit scoring activities making use of AI systems would be subjected to the minimum requirements for high-risk AI systems. The **ECB suggests methods such as decision-trees a logistic regressions are not considered high-risk** provided that the impact of such approaches to the assessment of natural persons' creditworthiness or credit score is minor.
- For credit scoring, the **ECB suggests to delay the entry into force until there are specifications on the conditions to verify conformity with the applicable requirements**, and define when AI systems should be considered as 'put into service by small scale providers for their own use'.
- The ECB suggests to consider **updating the list of high-risk AI systems to consider other AI systems put into place by financial institutions** such as AI data modelling linking sales, transactions, and performance data to ensure a clear, overview of conduct risk in a certain area. Similarly, AI systems might be used in the real time monitoring of payments, or profiling of clients or transactions, for anti-money laundering and counter-terrorist financing purposes...

# A | Annex 3

## Supervision and initiatives - ECB Position (1/2)



### In the banking industry, the ECB as prudential supervisor for credit institutions has published its position regarding the AI Act requirements

Clarification on the ECB's role under the proposed regulation

- **Clarification of the ECB's prudential supervisory competences in relation to market surveillance**
  - The ECB understands that, under the proposed regulation, the ECB is not in any way a market surveillance authority.
  - The ECB considers that market surveillance does not aim to ensure the safety and soundness of credit institutions, but focuses instead on protecting the interests of individuals and proposes the text should be modified to clarify this.
  - Member States might consider the designation of national competent authorities involved in the supervision of credit institutions as responsible for market surveillance in the context of the proposed regulation, insofar as permitted by their mandate.
  - The ECB notes that the market surveillance provisions of the proposed regulation do not adequately address situations in which an AI system is put into service for own use.
- **Clarification of the ECB's prudential supervisory competences in the area of conformity assessment**
  - The Union legislator is invited to consider the extent to which several elements of the conformity assessment might not be prudential in nature insofar as they largely concern the technical assessment of AI systems to safeguard the health and safety of persons and ensure that fundamental rights.
  - The highlights the need to designate relevant competent authorities as responsible for the supervision of the conformity assessment for requirements on health, safety and fundamental rights.
  - Certain requirements for high-risk AI systems are not entirely clear or specific enough to provide a sufficient understanding to inform supervisory expectations (e.g., train, validation and testing data to be relevant and representative).
  - The ECB considers that the proposed regulation should be amended to reflect the ex-post nature of the specific assessment as part of the SREP.
- **Clarification of the ECB's prudential supervisory competences:** The ECB may be considered a competent authority only insofar as necessary for it to carry out the tasks conferred on it under the SSM Regulation.
- **Clarification of the ECB's independence in the performance of its tasks under the Treaty:** The ECB understands that when acting as a provider placing on the market or putting into service AI systems, or as a user, it (or the NCBs) may be subject to the proposed regulation, while maintaining their independence to carry out the tasks conferred on it by the Treaty.



# A | Annex 3

## Supervision and initiatives - AI Sandbox in Spain



The Spanish Government has launched an initiative, in collaboration with the European Commission, to implement an Artificial Intelligence regulatory sandbox in the EU

### Objective



- The aim of this collaboration is to **connect the competent authorities with Artificial Intelligence development companies to jointly define best practices when implementing the future European regulation** on Artificial Intelligence, promoted by the European Commission.
- The result of these tests will be **compiled in a best practices guide**, which will be presented during the Spanish Presidency of the Council of the EU in the second half of 2023.
- The guide will be **accessible to all Member States and the European Commission**.

### Scope



As the proposed **Artificial Intelligence Regulation focuses in particular on the obligations to be fulfilled by so-called high-risk artificial intelligence systems**, participation in the regulated controlled environment focuses on those:

- Artificial intelligence systems that are classified as high risk;
- General purpose artificial intelligence systems;
- Foundational models;
- Generative artificial intelligence systems.

### Benefits



Provide **clarity on the new requirements for AI systems** set out in the AI Regulation:  
Transfer **compliance expertise on the implementation of the forthcoming AI legislation to entities** developing AI solutions;  
**Encourage innovation** and enable the development of innovative and reliable AI systems;  
**Build capacity** and initiate consultations in Spain that will eventually lead to the creation of a National Supervisory Authority;  
**Test future obligations and requirements in a controlled environment** and provide practical learning experience to support the development of standards, guidance and tools at national and European level.





## Annex 3

# Supervision and initiatives - AI Supervisory Authority in Spain



**In Spain, the Royal Decree 729/2023 approving the constitution of the Spanish AI Supervisory Agency, as a state agency attached to the Ministry of Economic Affairs and Digital Transformation**

*Spanish Agency for the Supervision of Artificial Intelligence (AESIA)*

### AESIA's constitution

- The Agency will have **its own legal personality, management autonomy and administrative powers** to fulfil its purposes. Its headquarters will be in A Coruña.
- The **AESIA will supervise the use of artificial intelligence systems to protect fundamental rights and minimise risks**. It will collaborate with national and European authorities.
- Its competences include: promoting test environments, fostering ethical and sustainable use of AI, strengthening trust in the technology, coordinating with other actors, training and raising awareness of responsible use of AI.
- The **AESIA will have governing bodies** (Presidency and Governing Board), executive bodies (Directorate, General Secretariat, two Sub-Directorates) and control bodies (Control Commission).
- The AESIA's personnel, economic-financial, budgetary, patrimonial and contracting regime is regulated.
- Legal assistance will be provided by the State Attorney General's Office.

### Main Competences

**The main competences of the Spanish Agency for the Supervision of Artificial Intelligence (AESIA), as detailed in the Royal Decree, are:**

- Promote regulated test environments so that AI systems can be tested safely and in compliance with the law.
- Promote ethical, sustainable and environmentally friendly use of AI.
- Create a voluntary certification system to ensure technical standards and responsible design of AI solutions.
- Identify trends and assess the social impact of AI through studies and reports.
- Coordinate with other public and private initiatives related to AI.
- Generate knowledge, training and awareness of ethical and humanistic AI.
- Dynamise the market to foster innovative AI practices.
- Collaborate with the private sector to foster a humanistic development of AI.
- Monitor and, where appropriate, sanction the use of AI systems to ensure compliance with European and national regulations.
- Provide technical assistance to judges and courts in AI-related legal cases.
- Other functions related to the supervision of AI that may be attributed to it due to regulatory or technological changes.



## A

## Annex 4 Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
EBA	European Banking Authority
CRR	Capital Requirements Regulation
CAC	Cyberspace Administration of China
EC	European Commission
EU	European Union
GPDP	Guarantor for the Protection of Personal Data
Nat Cat	International Organization of Securities Commissions
IOSCO	National Competent Authority
ML	Machine Learning
MINECO	Ministry of Economic Affairs and Digital Transformation
METI	Ministry of Economy, Trade and Industry
OECD	Organisation for Economic Cooperation and Development
IRB	Internal ratings-based
PDPC	Personal Data Protection Commission

**MSO**  
**Management Solutions**  
*Making things happen*



**International  
One Firm**



**Multiscope  
Team**



**Best practice  
know-how**



**Proven  
Experience**



**Maximum  
Commitment**

**Javier Calvo Martín**

Partner at Management Solutions

Javier.calvo.martin@managementsolutions.com

**Manuel Ángel Guzmán**

Partner at Management Solutions

Manuel.guzman@managementsolutions.com

**Marta Hierro**

Partner at Management Solutions

Marta.Hierro@msspain.com

© Management Solutions, 2023

All rights reserved. Cannot be reproduced, distributed, publicly disclosed or transformed, whether totally or partially, free of charge or at no cost, in any way or by any means, without the express written authorization of Management Solutions.

The information contained in this publication is merely to be used as a guideline, is provided for general information purposes and is not intended to be used in lieu of consulting with our professionals. Management Solutions is not liable for any use that third parties may make of this information. The use of this material is not permitted without the express authorization of Management Solutions.

For more information please visit

[www.managementsolutions.com](http://www.managementsolutions.com)

Or follow us at: