Balancing Innovation and Regulation: A Detailed Analysis of the AI Act

The EU's new regulatory framework on AI is doubtlessly a pioneering legislative framework that could set a global benchmark, but does it strike the right balance between protection and innovation? Addressing the risks that AI technologies present to people's safety and fundamental rights is of major importance, but so is the need to foster AI innovation in the EU and the uptake of these transformative technologies that could help reinvigorate the EU economy. To be sure, most AI systems are expected to qualify as low/minimal risk and should thus not be (materially) affected by the new rules. On the other end, some AI applications that present unacceptable risks will now be banned while those classified as high-risk will be subject to a particularly heavy regulatory burden. In other words, the administrative burden of this Regulation predominantly falls on entities classified as providers of systems that qualify as high-risk AI systems.

A key concern is that the cost and complexity of compliance could stifle innovation, especially for SMEs and startups with limited means. Another concern is the legal uncertainty stemming from the broad definition of "AI system" that could catch systems and software that would typically not be thought of as artificial intelligence, such as certain manually constructed expert-rules-based systems. Furthermore, early hints from the new US administration suggest that the US may adopt a light-touch regulatory approach to AI, which could impact Europe's AI innovation and competitiveness.

As the February 2025 deadline for the application of Chapter II of the Act ("Prohibited AI Practices") moves closer, entities especially in industries more likely to be using AI, such as Financial Services, Marketing, E-commerce, Gambling, Gaming, Cybersecurity, and Healthcare should, as a matter of priority carry out a company-wide systems screening to identify and timely remove AI applications that could fall foul of the Chapter II ban. The Regulation is particularly tough on violations of the ban, providing for fines of up to €35 million or 7% of global turnover.

To better understand the extent of the administrative burden and the Act's impact on AI development and deployment in the EU, a closer look at the new rules and their practical implications is warranted.

1. Nature, Scope, and Key Definitions

The AI Act lays down rules with respect to AI systems and general-purpose AI models, employing mainly result-oriented provisions to be fleshed out at a later stage by technical specifications (harmonised industry technical standards, codes of practice, and possibly also Commission common specifications). The Act is moreover, technology-neutral; the goal is not to regulate technology (which is neither inherently beneficial nor harmful) but to regulate particular uses of technology that pose material risks to people's safety and fundamental rights. This approach makes the Act considerably future-proof and provides flexibility for adapting the framework to changes in uses or in the state of the art. The new framework applies across sectors and is without prejudice to existing EU laws such as GDPR, consumer protection, employment, and product safety.

Scope: The Regulation has a broad territorial scope, catching even 3rd country AI systems that affect the EU. In particular, it applies to:

SCOPE	**The Provider** (person or entity that develops an AI system or a general-purpose AI model) who - places on the market or puts into service an AI system, or - places on the market a general-purpose AI model under its own name or trademark in the EU, irrespective of being established in the EU or in a 3rd country;
	The Deployer, defined as the person or entity using an AI system under its authority, having its place of establishment in the EU;
	Providers and Deployers of AI systems established in a 3rd country, where the output produced by the AI system is used in the Union;
	Other Operators, where applicable: Legal representatives of 3rd country providers, importers and distributors of AI systems in the EU, and certain product manufacturers.

⇨ A US developer of an AI credit scoring system that places that system on the EU market (i.e, downloadable from its EU website for a price), is considered a Provider.

An EU bank that buys this system and uses it to screen loan applicants is considered a Deployer

⇨ Under certain conditions a deployer (or importer or distributor) may be considered to be a provider, such as when he affixes his name or trademark on a high-risk AI system already placed on the market or put into service or substantially modifies such a system

It is important to bear in mind that the AI Act applies only if an AI system or model has been placed on the market or put into service or used in the EU; the stages relating to its development and testing do not fall within the scope of the Act. Other out-of-scope cases include use of AI systems for military, defence or national security purposes, AI systems and models for scientific research purposes, certain AI systems released under a free and open-source license, and cases where the deployer is a natural person using an AI system for a personal non-professional activity.

➢ **What is an AI System** (Article 3(1) & recital 12)**:**

A machine-based system → With varying levels of **autonomy** **&** That may exhibit **adaptiveness** after deployment **&** That **infers** from the input it receives, how to generate outputs (i.e, content, decisions)

- **Autonomy:** the system must have some level of autonomy, hence probably even a low level of autonomy (in the sense of some independence from human involvement) would do.
- **Adaptiveness:** The word "may" can be understood to mean that this is not an essential feature (note that the great majority of machine learning systems learn by analysing large amounts of data during a training phase, which precedes deployment. Hence, in a strict sense, only those very few systems that continue to learn post deployment would qualify as truly adaptive).

- **Capability to infer:** the core of the definition is the capability to infer outputs from received inputs. An AI system is typically considered to be a system trained to detect patterns on a large data set (i.e, x-rays) during its training stage, which can then -at deployment stage -recognise these patterns in new data that it has not seen before (new x-rays) and draw accurate conclusions or predictions in a way that resembles human logic. The Act's definition however, read in conjunction with recital 12, is quite broad and can catch traditional automated systems or software that are not sophisticated or "intelligent" in the sense commonly associated with AI, yet they may be said to "infer" outputs from inputs in a way that constitutes basic or rudimentary "modelling" or "reasoning".

This broad definition, while helpful in keeping it future-proof and less prone to circumvention, casts the regulatory net beyond what would typically be considered AI and likely catches less sophisticated systems. In the case of some systems, operators may struggle to determine whether to classify them as AI or not. More clarity on this should result from the guidelines that the Commission is expected to issue on the application of the definition.

AI System

Vs

General Purpose AI Model

It is important to be mindful of the crucial distinction between an AI system and a general-purpose AI model (GPAI model), and their different treatment under the Regulation. According to recital 97 "*Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components such as for example a user interface, to become AI systems*". For instance, in the case of Open AI, the GPAI model is GPT 4.0, while the AI system is ChatGPT 4.0.

➢ A Risk-Based Regulatory Approach

Systems that qualify as AI systems are subject to a risk-based regulatory treatment consisting of 4 risk levels. Those falling in the top category are considered of "unacceptable risk" (i.e, social scoring) and are banned; those in the high-risk category (i.e, credit scoring) are subject to stringent regulation; AI systems classified as limited risk (i.e, AI chatbots) are only subject to certain disclosures. All other AI systems are considered minimal risk and are not subject to obligations (save for the AI literacy provision of Art.4). According to the European Commission, the vast majority of AI systems currently used in the EU fall in this latter category.

Image: European Commission

Most of the obligations and requirements of the Act fall on providers of AI systems that qualify as high-risk. Deployers of such systems are also affected, albeit to a much lesser extent. Providers of GPAI models, must also comply with several requirements (see GPAI section below), since these models are particularly powerful,

versatile, and can be integrated and form the basis of a large number of AI systems. Models that qualify as systemic risk GPAI models (most likely GPT-4.0 or Gemini core AI models), are subject to additional more stringent rules.

2. Prohibited AI Practices

Chapter II of the Regulation sets out 8 specific AI use cases that are considered particularly harmful and are banned in the EU.

- **Manipulation & Deception**: AI systems that deploy subliminal **or** purposefully manipulative or deceptive techniques with the objective **or effect** of materially distorting a person's behaviour thereby causing it to take a decision he would not otherwise have taken, causing or likely to cause significant harm

  ⇨ A manipulative or deceptive technique could involve a betting or ecommerce site that aggressively uses personalized adaptive time-limited discounts to push the customer to buy. Note that intention is not necessary as it's enough for the technique to have the effect of distorting a person's behaviour in the manner described. On the meaning of terms such as significant harm and manipulative practices, GDPR and the Unfair Commercial Practices Directive can be a useful guide.

- **Exploiting the Vulnerable**: AI systems that exploit the vulnerabilities of people due to their age, disability, **or** specific social or economic situation, with the object or effect of materially distorting their behaviour and causing or likely to cause significant harm

  ⇨ Note that the vulnerability can be permanent or temporary. Persons can be deemed vulnerable due to factors other than age and disability, such as low income, financial distress, or health issues (think of an AI system that preys on low education and low-income households).

- **Detrimental Social Scoring**: AI systems that evaluate or classify people based on their social behaviour or other personal/personality characteristics, resulting in detrimental treatment in unrelated contexts or treatment that is disproportional to the social behaviour

  ⇨ For instance, an AI system that assigns a score to people based on their social media data, which is then used to determine suitability for a loan or a job.

- **Predictive Policing**: AI system that assesses the risk of an individual committing a crime, based solely on the profiling of that individual (the movie "Minority Report")

- **Untargeted Scraping of Facial Images** from the internet or CCTV footage for creating or expanding facial recognition databases

- **Emotions Recognition**: AI systems that infer individuals' emotions at the workplace or at educational institutions, on the basis of their biometric data.

  ⇨ The prohibition only applies to emotions; it does not apply to physical states such as pain or fatigue (i.e, a system monitoring a driver's fatigue) nor to expressions (such as frowns) or gestures (thumbs up) or movements (hands up). For example, a bank's AI system that recognises if its employees are happy or sad or angry would be prohibited. A system that recognises the same for viewers of an ad at a mall would not be prohibited (but would classify as high risk)

- **Biometric Categorisation**: AI systems that, on the basis of individuals' biometric data (i.e, face, voice, fingerprint), categorise them according to sensitive attributes such as race, religion and political opinions (exceptions apply)

- **Real-Time Remote Biometric Identification** of persons in publicly accessible spaces by law enforcement bodies

  ⇨ For instance, an AI system deployed by the police at an airport or a mall to remotely face-scan people in real-time or near real-time, in search of a particular person. Narrow exceptions apply for specific law enforcement purposes, such as when searching for abduction victims, potential terrorists, and persons who are suspected of 16 specified very serious crimes. Member States that wish to make use of such AI systems are required to introduce specific national rules.

The rules concerning prohibited practices become applicable on 2 February 2025. The Commission is expected to issue guidance on the application of the rules prior to that date.

3. High-Risk AI Systems

Systems qualifying as AI systems that meet the conditions of Chapter III, are classified as high-risk AI systems and are subject to extensive requirements. There are two cases in which an AI system can qualify as high-risk:

I. The AI system is used as a safety component of a product covered by the Union Harmonisation legislation listed in Annex I or the AI system is itself a product that is covered by that harmonising legislation, and the product is required to undergo a 3rd party conformity assessment

Or

II. It is one of the high-risk AI systems listed in Annex III (see below), such as specific uses of AI in education, employment, biometric identification, banking and insurance.

Products covered by the Union Harmonisation legislation include medical devices, toys, lifts, civil aviation, and other. For example, in the case of an AI system that constitutes a safety component of a medical device subject to 3rd party conformity assessment, the system will be classified as high-risk.

Exceptions: It is important to bear in mind that according to Article 6(3), an AI system falling within Annex III can escape the high-risk classification if it does not pose a significant risk, and particularly where it does not materially influence decision-making, such as when it performs a narrow procedural or preparatory task (i.e classifying and preparing essays for grading but not doing the grading). The exception cannot apply in the case of systems performing profiling of natural persons. If a provider considers that he can make use of the exception, she must document her assessment and provide it to the competent authorities upon request, and also register the AI system in the EU Database for High-Risk AI Systems.

➢ High-Risk AI Systems Listed in Annex III

Annex III lists 8 high-risk AI cases. Practical guidance, including examples of use cases will be issued by the Commission by February 2026.

1. **Biometrics**: AI systems, insofar as they do not constitute prohibited practices, that are used for:

   (a) **remote biometric identification**, such as a real-time remote biometric identification system deployed by the police at an airport on the basis of the law enforcement exception. Note that systems used solely for confirming a person's ID, such as for unlocking a smartphone, are not considered high-risk.

   (b) **biometric categorisation** of persons according to sensitive attributes or characteristics such as gender, age, language, religion. For example, an advertising AI system that analyses biometric data to categorise people according to gender and age and promote products accordingly. Note that biometric categorisation systems that are a purely ancillary feature intrinsically linked to another service are not considered high-risk (i.e, an app that shows how you look with different make-up styles).

   (c) **emotion recognition**, such as an AI system that performs facial recognition to infer whether or not spectators of content on a mall screen are happy with what they are being shown.

2. **Critical Infrastructure**: AI systems intended to be used as safety components of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity. Think of an AI system embedded in a system that regulates traffic lights with the aim of optimising the flow of traffic.

3. **Education and Vocational Training**: AI systems intended to be used to determine a person's access or admission to educational and vocational training institutions (i.e, assessing entry applications), to assess learning outcomes (i.e, grading tests or essays), or to monitor and detect prohibited behaviour by students during tests.

4. **Employment**: AI systems intended to be used for the placing of targeted job advertisements and for analysing job applications and evaluating candidates (i.e, scoring, ranking CVs). Also, systems used to make work-related decisions, such as promoting, demoting or terminating an employee, allocating tasks (i.e, the tasks arriving at an IT help desk), and assessing worker behaviour and performance.

5. **Access to Essential Public and Private Services and Benefits**: AI systems intended to be used for –

   - Assessing eligibility of persons for public benefits such as unemployment benefit or housing
   - Assessing the creditworthiness of persons or establishing a credit score, excluding AI systems intended for the detection of financial fraud (such as fraudulent transactions)
   - Risk assessment and price-setting in relation to persons in the case of life and health insurance
   - Assessing and classifying calls to emergency services such as police, firefighters and medical aid

6. **Law Enforcement**: AI systems intended to be used for –

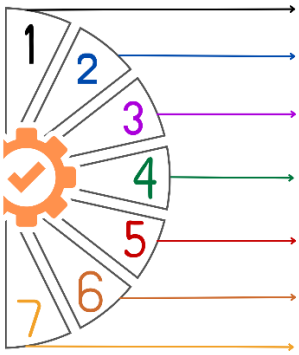
   - Assessing the risk of a person becoming the victim of a criminal offence
   - The operation of polygraphs or similar tools
   - Evaluating the reliability of evidence, such as DNA, fingerprints or digital evidence
   - Evaluating the risk of a person offending or re-offending, not solely based on profiling
   - AI systems used for the profiling of persons in the course of investigating criminal offences

7. **Migration, Asylum, Border Management**: AI systems intended to be used as polygraphs or similar tools, for assessing security risks such as the risk of irregular migration into a Member State, for evaluating asylum or visa applications, and for detecting or identifying natural persons.

8. **Administration of Justice**: AI systems intended to be used –
   - By or on behalf of a judicial authority to assist in researching and interpreting facts and the law and in applying the law to a concrete set of facts (i.e, a system that searches databases of statutes and case law and delivers recommendations to judges)
   - For influencing the outcome of an election or referendum or the voting behaviour of persons

4. *Requirements for High-Risk AI Systems*

Providers of high-risk AI systems have the greatest compliance burden, as these systems are subject to extensive **mandatory requirements** under section 2 of Chapter III, organised under 7 areas. The Commission is expected to issue guidelines on the application of these requirements.



Risk Management System

Data Governance

Technical Documentation

Record-Keeping (Logging)

Transparency to Deployers

Human Oversight

Accuracy, Robustness & Cybersecurity

High-risk AI Systems that are in conformity with harmonised technical standards will be presumed to be in conformity with these seven mandatory requirements.

The Commission has mandated the two European Standardisation Organisations CEN and CENELEC to develop appropriate standards.

1) **Risk Management System** (Article 9)

A risk-management system must be established, documented, and regularly reviewed and updated. Known and reasonably foreseeable risks must be identified, assessed in terms of likelihood and impact, and managed through targeted measures, particularly during the design stage. Where the provider is obliged to have risk management procedures under other EU laws (i.e financial institutions), these requirements may be integrated into those procedures.

2) **Data Governance** (Article 10)

Data sets used for training, validation and testing of the AI models that underpin high-risk AI systems must meet quality, relevance and completeness criteria. Data governance arrangements must be in place, governing issues such as data sources, collection, annotation, cleaning and enriching, and identifying and removing data gaps and bias. Moreover, the data sets must also be sufficiently representative, taking into account the persons in relation to whom the system will be used.

3) **Technical Documentation** (Article 11)

It must be drawn up before placing the system on the market or putting it into service, kept updated, and drawn in such a way as to demonstrate compliance of the system with the mandatory requirements of section 2. It should follow the structure and include the elements shown in Annex IV of the Regulation.

Among others, it should include information on the system's general characteristics, capabilities and limitations, purpose, forms of distribution, instructions for the deployer, logic and architecture, data sets used, accuracy levels, human oversight tools, description of the risk management system, and a copy of the EU declaration of conformity. The Commission shall establish a simplified technical documentation form for SMEs and startups.

4) **Record-Keeping (Logging)** (Article 12)

The system must automatically record events by means of log files to enable traceability of its functioning and post market monitoring. With respect to high-risk AI systems for remote biometric identification, minimum logging requirements include (i) recording the start and end of each use/session, (ii) the input data that triggered a decision, (iii) the reference database against which input data was checked, and (iv) the identification of persons accessing the system for the verification of results. An example would be an AI system used by the police that deploys CCTV at a train station to remotely identify a suspect of a serious crime.

5) **Transparency to Deployers - Instructions Manual** (Article 13)

A high-risk AI system must be designed in a way that deployers can understand how it works, its capabilities, limitations and risks, the way it should or shouldn't be used, how outputs should be interpreted, and what human oversight measures are available. To that effect, providers must accompany the system with instructions of use consisting of the elements set out in Article 13. Note that most of these elements are also found in the technical documentation mentioned under point (3) above.

6) **Human Oversight** (Article 14)

High-risk AI systems must be designed is such a way that they can be effectively overseen by natural persons. Oversight measures must be risk-based and should include measures built into the AI system by the provider, and/or measures that are implemented by the deployer.

In essence, the provider should provide to the deployer with an AI system that enables the deployer's staff that is tasked with overseeing the system (i) to understand and monitor the system and detect and address anomalies, (ii) to avoid over-relying on the system's output (automation bias), (iii) to correctly interpret the system's output, (iv) to decide to disregard, override or reverse the output, and (v) to interrupt the system through a "stop button".

Given the very high risks involved in the use of remote biometric identification systems, Article 14 provides that no action should be taken by the deployer on the basis of the identification resulting from the system unless separately verified by at least two persons with appropriate training and authority.

7) **Accuracy, Robustness & Cybersecurity** (Article 15)

High-risk AI systems must achieve appropriate levels of accuracy, robustness and cybersecurity and perform consistently in those respects throughout their lifecycle. The level of accuracy should be declared in the system's instructions manual. Concerning robustness, systems must be as resilient as possible regarding errors or faults; this should be ensured through technical and organisational measures such as backup or fail-safe plans. On cybersecurity, measures must be in place to prevent, detect and respond to attacks aiming to

manipulate the training data set (data poisoning) or the pre-trained components used (i.e, poisoning a pre-trained AI model) or feeding misleading input to cause mistakes (adversarial attacks).

➢ Additional obligations for providers of high-risk AI systems

Providers, in addition to ensuring compliance of their systems with the above mandatory requirements, must also comply with numerous other obligations. In particular, they are required to –

(i) indicate on the AI system their name, trade mark and address

(ii) have a Quality Management System in place with written policies and procedures

(iii) keep relevant documentation (i.e, the technical documentation, the documented quality management system, the EU declaration of conformity) for 10 years

(iv) keep the system's logs for at least 6 months

(v) subject the system to a conformity assessment (see below)

(vi) draw up a signed EU declaration of conformity according to Annex V

(vii) affix a CE marking, visibly, legibly and indelibly

(viii) register themselves and the system in the EU Database

(ix) demonstrate, upon a reasoned request of a competent authority, the conformity of the system with the mandatory requirements

Where a high-risk AI system is embedded as a safety component of a product that is covered by the Union harmonisation legislation listed in Section A of Annex I, and is not placed on the market or put into service independently of the product, the product manufacturer is considered to be the provider and bears responsibility for compliance with the Act's requirements.

On the requirement for a Conformity Assessment of Article 43, providers of high-risk AI systems listed in Annex III (excluding remote biometric identification systems) must follow the conformity assessment procedure based on internal control. This procedure, described in Annex VI, provides for self-verification by the provider and does not require the involvement of a notified body (a 3rd party conformity assessment body). As for remote biometric identification systems, providers have to follow the procedure of Annex VII (involving an assessment by a notified body), unless they have applied harmonised standards for demonstrating compliance with the mandatory requirements, in which case they can opt to apply Annex VI. Concerning high-risk AI systems that are products or safety components of products covered by the Union harmonisation legislation listed in Section A of Annex I, the provider must follow the conformity assessment procedure provided under that legislation. Note that according to Article 43, a high-risk system may be deployed for exceptional security or safety reasons without a conformity assessment.

Providers of high-risk AI systems are also expected to establish a post-market monitoring system to collect and analyse data on the performance of their systems in order to evaluate the continuous compliance of the

systems with the mandatory requirements. They are moreover required to report any serious incidents to the market surveillance authority of the Member State where the incident occurred.

➢ Obligations for deployers of high-risk AI systems

Deployers are required to take technical and organisational measures to ensure that they use high-risk AI systems according to their instructions of use. They must also ensure that the system is overseen and monitored by competent, trained staff, that input data fed to the system is relevant and sufficiently representative, and that the system's logs are kept for at least 6 months.

Where a high-risk AI system listed in Annex III makes decisions or assists in making decisions related to natural persons (say, a system used by a bank to assess applicants' eligibility for a loan), the deployer must inform these persons accordingly. If the effect on the person is significant, he has the right to obtain clear explanations from the deployer on the role of the AI system in the decision-making process. This right does not apply for high-risk systems relating to critical infrastructure.

A fundamental rights impact assessment must be carried out by deployers of high-risk AI systems listed in Annex III that perform creditworthiness assessments or price and risk assessments in life and health insurance. Deployers that provide public services must always perform such an assessment when using a high-risk AI system listed in Annex III, unless relating to critical infrastructure. The results of these assessments must be notified to the competent authority. To avoid overlaps, this assessment may be carried out in conjunction with a GDPR privacy impact assessment, where the latter is also required.

5. Transparency Obligations for Certain AI Systems

Certain AI systems, irrespective of being high-risk or not, can pose specific risks of deception and manipulation and are therefore subject to the transparency requirements set out below. It is noted that the Commission will issue guidelines on the practical implementation of these requirements.

Direct Interaction with Natural Persons - Disclosure
Providers of AI systems intended to interact directly with natural persons (such as chatbots), must ensure that the persons concerned are informed that they are interacting with an AI system, unless this is obvious.

Synthetic Content - Watermarking
Providers of AI systems, including GPAI systems, generating synthetic audio, image, video, or text content must ensure that the outputs are marked as artificially generated, in a machine-readable format. This could be a digital watermark using invisible pixels, thereby enabling other systems (i.e, social networking apps) to detect them. Note that the term "synthetic" is not defined, creating ambiguity on what should or should not be caught. Some clarity stems from the fact that the provision excludes assistive functions such as standard editing that does not significantly change the input data (i.e, submitting text to ChatGPT for proofreading).

Deep Fakes - Disclosure
Deployers of an AI system that generates or manipulates image, audio or video content that constitutes deep fakes must disclose that the content is artificially generated or manipulated. "Deep fake" is defined as an AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events **and** that would falsely appear to a person to be authentic or truthful. Hence, the mere fact that an image of an actor is AI-generated is not enough; it must also appear to be an authentic image of an existing actor. Milder rules apply with respect to creative, artistic, satirical expression. With respect to AI-generated or manipulated text, the deployer is required to disclose only in the case of "*text which is published with the purpose of informing the public on matters of public interest*".

Emotion Recognition - Disclosure
Deployers of an emotion recognition system or a biometric categorisation system are required to inform affected persons about the operation of the system.

6. General-Purpose AI Models

The Regulation clearly distinguishes between the notion of an AI system and that of a general-purpose AI model (GPAI model), setting specific rules for GPAI models in view of their extensive capabilities, versatility, and integration into a large number of downstream AI systems. The Act defines a GPAI model as follows:

> "an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that **displays significant generality** and is capable of **competently performing a wide range of distinct tasks** regardless of the way the model is placed on the market and that can be **integrated into a variety of downstream systems** or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market;

Recital 98 indicates that GPAI models (also known as foundation models) with at least a billion of parameters and trained with a large number of data using self-supervision at scale would satisfy the generality and competence requirements. It also mentions large generative AI models as typical examples of GPAI models (i.e, Open AI's GPT 4.0 and Meta's Llama). It is noted that the rules on GPAI models apply also when these models are integrated and form part of AI systems, resulting in general-purpose AI systems. The requirements apply once the model is placed on the market, and the responsibility for compliance rests solely with the model provider.

➢ Requirements for Providers of General-Purpose AI Models

Technical Documentation	Providers must draw up and keep updated the technical documentation of the model, containing at a minimum the information in Annex XI
Information to AI System Providers	Providers must draw up, keep updated and make available to downstream providers of AI systems that intend to integrate the GPAI model into their systems, information and instructions on the model's purpose, architecture, capabilities and other characteristics, as per Annex XII.
Compliance with Copyright Law	Providers must put in place a policy to comply with EU copyright law to ensure that data mining for the models' development and training does not infringe the rights of copyright holders.
Training Data	Providers must draw up and make publicly available a detailed summary of the content used for training the model (i.e, listing the main data sets used)

The first and second of the above requirements do not apply to providers of GPAI models released under a free and open-source license, provided that these models do not qualify as GPAI models with systemic risks.

➢ Requirements for Providers of General-Purpose AI Models with Systemic Risk

The Act provides that a GPAI model with systemic risk is one that (i) has high impact capabilities, or (ii) has been classified as such by a Commission decision on the basis of the criteria set out in Annex XIII. Art.51 provides in particular that a GPAI model will be presumed to pose systemic risks when trained using a total computing power of more than 10^25 flops (floating-point operations). It is understood that very few models currently meet that threshold (perhaps GPT4.0 and Gemini).

Providers of these models are subject to additional requirements, such as conducting state of the art model evaluations including adversarial testing, assessing and mitigating risks, documenting and reporting serious incidents, and ensuring cybersecurity of their models and related infrastructure.

Compliance of providers with the requirements for GPAI models, including those with systemic risks, will be facilitated by the creation of a Code of Practice, under the steering of the Commission.

7. Innovation, Supervision, Penalties & Timeline

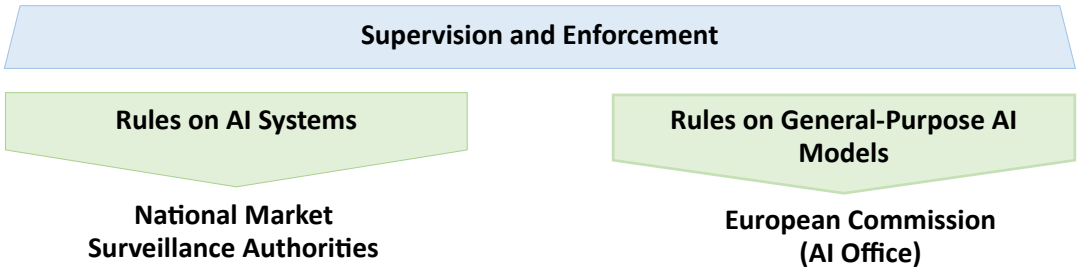
➢ Measures to Support Innovation

The Act seeks to foster innovation in AI by requiring national authorities to establish at least one AI regulatory sandbox at national level. These are arrangements that enable providers to develop and test AI systems in a controlled experimentation environment under regulatory oversight, before market release. SMEs, including startups, provided they meet the entry criteria, should enjoy priority access to the sandboxes free of charge.

In addition to sandboxes, the Act also provides for the possibility of pre-market real world testing of high-risk AI systems for a limited period, subject to authorisation by the competent market surveillance authority and with specific safeguards.

➢ Supervision and Enforcement

The AI Act creates a two-level governance framework: national authorities are responsible for the supervision and enforcement of the rules relating to AI systems, while at EU level the Commission will supervise and enforce the rules relating to general-purpose AI models. Member States are required to establish or designate a market surveillance authority for this purpose. Supervision and enforcement at EU level will be conducted via the Commission's recently established AI Office.



It is important to point out that, in respect of high-risk AI systems placed on the market, put into service, or used by EU **financial institutions** (i.e, banks, insurers, investment firms), their supervisory authority (i.e, the Central Bank) shall have the role of market surveillance authority for these institutions as regards the requirements of this Regulation (unless a Member State designates another authority). A similar approach applies concerning high-risk AI systems related to products covered by the Union harmonisation legislation listed in section A of Annex I; the market surveillance authority will be the one designated under that product legislation.

➢ Penalties

Penalties for infringement of the rules for AI systems:

- Non-compliance with the prohibited AI practices rules of Article 5: Fine of up to €35 million or up to 7% of total global annual turnover, whichever is higher (for SMEs/startups: the lowest).
- Non-compliance with other requirements of the Regulation: Fine of up to €15 million or up to 3% of total global annual turnover, whichever is higher (for SMEs/startups: the lowest).
- Providing incorrect, incomplete, or misleading information to authorities: Fine of up to €7,5 million or up to 1% of total global annual turnover, whichever is higher (for SMEs/startups: the lowest).

Penalties for Infringement of the rules for general-purpose AI models:

- The Commission can impose fines for violations of the rules for GPAI models of up to €15 million or up to 3% of total global annual turnover, whichever is higher.

➢ Liability Rules

Although the AI Regulation does not set out AI liability rules, recent developments concerning the new Product Liability Directive ("PLD") are particularly relevant and important to highlight in this context. One of the key updates in the PLD is that the term "product" is expanded to include standalone software (including AI software). Moreover, it introduces a right of disclosure that enables a potential claimant who can demonstrate the plausibility of his claim, to oblige product manufacturers to disclose information and evidence that he can use to support his case for compensation.

The most important update however relates to the possibility for reversal of the burden of proof from the claimant to the product manufacturer. Under certain conditions, the product defect and/or the associated damage to the claimant may be presumed, in which case the burden shifts from the claimant to the manufacturer to rebut the presumption. For instance, this may occur where the manufacturer fails to comply with a disclosure of information order. It could also occur where a claimant can demonstrate that the product's defect or the causal link between defect and damage is likely, but faces great difficulty in proving either of these due to the product's technical or scientific complexity.

Given these developments, providers of high-risk AI systems must pay special attention to the robustness of their system's technical documentation and record-keeping so that, if they are not at fault, they will be able to demonstrate this and effectively defend themselves.

➢ Application Deadlines

The general date of application of the AI Regulation is **2 August 2026**. Some of its provisions however, will apply earlier or later than that date, as shown below.

2 August 2024	⇨ Entry into Force
2 February 2025	⇨ General Provisions, AI literacy, Chapter II (Prohibited Practices)
2 August 2025	⇨ Obligations for General-Purpose AI Models, provisions on National Authorities, Governance, and Penalties
2 August 2026	⇨ General Date of Application
2 August 2027	⇨ Obligations for high-risk AI systems related to products listed in Annex I

Note that as regards operators (i.e, providers, deployers) of high-risk AI systems that have been placed on the market or put into service before 2 August 2026 (the general date of application of the Regulation), the **Regulation applies only if those high-risk systems undergo a significant modification after that date** (see definition at Article 3(23)). With respect to general-purpose AI models that have been placed on the market before 2 August 2025 (the date of application of the obligations for GPAI models), their providers must comply with the relevant rules by 2 August 2027.

8. Balancing between Protection and Innovation

The AI Act has been both commended and criticised for its effort to strike a balance between protection and innovation. Its approach of tailoring the regulatory requirements according to the risks that AI systems present seeks to ensure an appropriate level of protection for individuals while leaving ample space for AI innovation in the Union. Looking closely at the various requirements, it is clear that the obligations for providers of high-risk AI systems are quite strict and extensive, creating a heavy compliance burden, especially for SMEs and startups. Moreover, the definition of AI systems is quite broad, likely catching systems that are not sophisticated or "intelligent" in the sense commonly associated with AI. Also noteworthy is the view by some commentators that the framework may not be flexible or adaptable to the extent necessary given the rapid pace of technological progress. If the Act in fact leans significantly towards regulation, this can stifle AI innovation in the EU, and result in innovative startups, specialists and investments opting for less-regulated markets such as the US.

At the same time, it should be borne in mind that the large majority of AI systems will not qualify as high-risk and will thus not have to shoulder a heavy compliance burden. SMEs will, in a few cases, benefit from lighter procedures, such as simplified technical documentation forms. As for the broad definition of an AI system, the Commission guidelines are likely to bring some more certainty. On flexibility and adaptability, it is reminded that the Regulation consists of technology-neutral, results-oriented provisions, to be complemented by more easily adjustable industry-driven technical standards; moreover, many elements of the Act (such as the Annex III list of high-risk systems) can be modified by the Commission via delegated acts. Then there is harmonisation; the Regulation establishes a common EU-wide AI regulatory framework that grants legal certainty to operators and the ability to operate cross-border and scale-up. It also provides for the creation of regulatory sandboxes to encourage AI innovation, especially by SMEs and startups, in controlled environments. It follows that the Act sets a balance that, while not perfect, is overall satisfactory, with scope for adaptations as the use of technology and the state of the art evolve.

9. Next Steps

As a first step, companies should create an inventory of their IT software and applications, in order to screen and identify any systems that qualify as AI systems. This could be tasked to a cross-functional team consisting of IT, legal, compliance, and risk staff (and data scientists, if available), or outsourced to experts. For any identified AI system, the company's role with respect to that AI system must be determined, such as provider, deployer, importer, distributor. The company also needs to determine in which of the four risk categories the system falls: no/minimal risk, low risk, high risk, or unacceptable risk (prohibited practices).

The priority at this stage should be the identification of prohibited AI practices, in view of the application of the ban from February 2025; cancelling a prohibited AI practice, or changing it so as not to qualify as such, can take considerable time. With respect to AI systems that qualify as high risk or low risk, the company needs to carefully examine −depending on whether it is a provider or deployer −the relevant requirements and obligations under the Act, such as risk management, documentation, disclosures. It is noted that even where systems qualify as no or minimal risk, providers and deployers need to pay attention to the AI literacy requirements of Article 4 (as these apply from February 2025 to all AI systems), and take appropriate measures to raise AI awareness among staff.