Technical note

Draft guidelines on the classification of high-risk AI systems

Issued

19 May 2026

Issuing agencies

European Commission

Primary applicability

Providers and deployers of AI systems in the EU; competent market surveillance authorities

The draft Guidelines clarify the two classification routes for high-risk AI systems under the AI Act and provide practical criteria for providers, deployers and authorities

Context

- The **AI Act entered into force in August 2024**, establishing a risk-based framework for AI systems in the EU.
- Under **Article 6**, AI systems may be classified as high-risk through two routes: **Annex I** products or safety components, and **Annex III** use cases.
- The EC has published **three draft Guidelines for stakeholder consultation**, covering general principles, the Annex I route and the Annex III route.

Scope

- The draft Guidelines are limited to the **classification of high-risk AI systems under Article 6 of the AI Act**.
- They cover both classification routes: **Article 6(1) / Annex I** and **Article 6(2) / Annex III**.
- They are intended for **providers, deployers and market surveillance authorities** involved in the classification assessment.
- They do **not** address the full set of requirements and obligations applicable to high-risk AI systems, which will be covered by future Commission guidance.

Next steps

- **23 June 2026.** Deadline for stakeholders to submit feedback on the draft Guidelines.
- **2 December 2027.** Expected application date of the high-risk rules for systems classified under **Article 6(2)**, subject to the final AI Omnibus outcome.
- **2 August 2028.** Expected application date of the high-risk rules for systems classified under **Article 6(1)**, subject to the final AI Omnibus outcome.

Draft Guidelines on the classification of high-risk AI

General principles

- Clarifies the cross-cutting concepts that apply across both classification routes, including the definition of an **AI system**, the relevance of the system's **intended purpose**, the provider's role in the **self-assessment**, and the need for consistency across technical, legal and commercial documentation.

Annex I route

- Covers the **Article 6(1)** classification route, under which an AI system may be high-risk where it is itself a **product**, or a **safety component of a product**, covered by EU harmonization legislation and subject to third-party conformity assessment. It also clarifies that not every AI system embedded in a regulated product is automatically high-risk.

Annex III route

- Sets out the **Article 6(2)** classification route, under which an AI system may be high-risk where its intended purpose falls within one of the **Annex III use cases**.
- Addresses horizontal interpretation issues, including **human involvement**, **complex or agentic systems**, systems acting **on behalf of public authorities**, and the **Article 6(3) filter mechanism**.

The draft Guidelines clarify how to apply the high-risk classification rules in practice, especially for borderline cases, Annex III exclusions and complex decision-making systems

Expansion of high-risk perimeter through examples

- **Practical examples** bring additional borderline use cases within the scope of high-risk classification.
- Reduces ambiguity, especially in areas such as **biometrics**, **employment** and **access to essential services**.

Substance over form in intended purpose

- **Classification** depends on how the **system** is actually **positioned** and **used**, not only on formal disclaimers.
- Documentation, instructions and marketing materials become **key** to assess risk.

Human oversight does not mitigate classification

- The presence of human review **does not change** the classification if the AI system influences decisions.
- Human-in-the-loop is framed as a **control**, not as a way to avoid high-risk obligations.

Narrow interpretation of the Article 6(3) filter

- The exclusion filter is **restricted** to marginal or non-influential functionalities.
- Systems that support or shape decisions, such as **scoring**, **prioritisation** or **alerts**, **remain in scope**.

Profiling and system complexity trigger high-risk classification

- Profiling of natural persons and complex or agentic architectures are assessed **holistically**.
- **Splitting functionalities** or **modularising systems** does **not prevent** classification if outputs are combined.

Stronger accountability and ex-ante documentation

- Classification must be **justified** and **documented** before market deployment.
- Supervisory authorities may **challenge** assessments, **increasing** the need for robust evidence.

Key takeaways

- The main impact of the draft Guidelines is to make classification more evidence-based: organisations will need to assess AI systems based on their actual role, decision impact and supporting documentation, not only on formal descriptions.

The Guidelines provide practical examples of systems that are high-risk, out of scope or potentially excluded through the Article 6(3) filter

	High-risk examples	Out of scope / potentially excluded examples
Biometrics	**Remote biometric identification** through CCTV or voiceprint matching against a reference database; **biometric categorization** based on sensitive attributes; **emotion recognition** based on biometric data.	**One-to-one biometric authentication**, such as smartphone unlocking or access control where the person actively participates; **presence detection** without identifying individuals.
Critical infrastructure	**AI safety components** used to monitor traffic and adjust traffic lights, detect risks in water or electricity systems, or protect the physical integrity of critical infrastructure.	**Network optimization**, **trouble-ticket management**, **demand forecasting**, cybersecurity monitoring or quality checks with no direct safety function.
Education	**Automated admissions systems**, **AI-enabled grading** contributing to final evaluation, adaptive placement tools, and **proctoring systems** detecting prohibited behaviour during tests.	**Admission chatbots** providing information, language-learning feedback, plagiarism checks outside live tests, **exam quality checkers** or application-processing support.
Employment	**CV ranking**, **candidate scoring**, targeted job advertisements, background checks producing risk scores, task allocation based on performance or behavioural signals, and **performance monitoring**.	**Scheduling interviews**, recognising and organising CV information into a database, checking discriminatory wording in job ads, employer reputation monitoring or candidate-managed CV tailoring.
Essential services	**Credit scoring** for consumer lending or mortgages, eligibility assessment for public benefits, **life/health insurance risk assessment or pricing**, and emergency-call triage.	**Customer segmentation or marketing** not used for creditworthiness, claims management **separate** from insurance risk/pricing, emergency-call transcription **without** classification, or scheduling medical appointments.

Key takeaways

- The examples help operationalise the classification criteria by showing when similar AI systems fall inside or outside the high-risk perimeter.

The Guidelines clarify that high-risk classification depends on the system's intended purpose and on whether it falls within one of the two Article 6 routes

Two pathways into the high-risk regime	Under Article 6 of the AI Act, an AI system is classified as high-risk in two scenarios: • **Article 6(1) and Annex I:** The AI system is intended to be used as a safety component of a product, or is itself a product, covered by Union harmonization legislation listed in Annex I, and the product is required to undergo a third-party conformity assessment. • **Article 6(2) and Annex III:** The AI system falls within one of the use cases listed in Annex III of the AI Act.
AI system threshold	Before a system can be classified as **high-risk**, it must **qualify as an "AI system"** within the meaning of **Article 3(1) of the AI Act:** • An AI system is a machine-based system that operates with a degree of autonomy and infers outputs, such as predictions, recommendations, decisions or content, that may influence physical or virtual environments.
Intended purpose as a core criterion	• The **intended purpose** of the AI system is central to classification. Providers must describe the envisaged use clearly and consistently across instructions for use, promotional or sales materials, statements and technical documentation, including the specific context and conditions of use. • For multi-purpose and general-purpose AI systems, high-risk uses may be deemed part of the intended purpose where they are **feasible and reasonably foreseeable**, unless the provider clearly, concretely and coherently limits or excludes such uses across all materials.
Provider self-assessment and value chain	• The classification assessment is the responsibility of the **provider**, subject to supervision by competent market surveillance authorities. Distributors, importers, deployers or other third parties may also become subject to provider obligations where they rebrand a high-risk system, substantially modify it, or change its intended purpose so that it becomes high-risk.

Key takeaways

- Providers should align legal, product, technical and marketing materials, as classification depends on the system's intended purpose, objective functionalities and reasonably foreseeable use, not only on formal disclaimers.

Under Article 6(1), an AI system is high-risk only where it is linked to a regulated product and the relevant product is subject to third-party conformity assessment

Classification Rationale	• Article 6(1) establishes a proportionate, safety-based classification mechanism for AI systems integrated into products already regulated under EU harmonization legislation. **Only AI systems that present significant risks to health, safety, or fundamental rights are classified as high-risk**; not all AI-enabled products are automatically high-risk.
The two cumulative conditions	1. **The AI system must be a product, or a safety component of a product, covered by Annex I legislation.** Annex I lists Union harmonization legislation covering products such as machinery, toys, lifts, pressure equipment, radio equipment, medical devices, in vitro diagnostic devices, and products in the automotive and aviation sectors. The AI system may be embedded in the product or placed on the market independently (e.g., as a software update or remote service). 2. **The product must be required to undergo a third-party conformity assessment.** The product must be deemed by Union harmonization legislation to present higher risks requiring independent third-party verification before being placed on the market.
Safety Component Definition	• Article 3(14) of the AI Act defines a "safety component" as a component of a product or of an AI system which fulfils a safety function for that product or AI system, or the failure or malfunctioning of which endangers the health and safety of persons or property. This is an autonomous concept, independent of how sectoral legislation defines safety components. Based on the above, **two alternative scenarios** trigger classification as a safety component: • The AI system is intended to fulfil a safety function (***intent-based***, determined by the provider). • The AI system's failure or malfunctioning would endanger health, safety, or property (**consequence-based**). Examples of safety components include AI-based collision detection in machinery, AI monitoring gas concentrations in explosive environments, and AI systems in vehicles providing lane assistance. By contrast, an AI system recommending music in a connected toy or performing yield forecasting in agriculture would not be a safety component.
Compliance Burden Minimization	• The AI Act allows economic operators to integrate AI-specific risk and quality management assessments into existing sectoral compliance frameworks, avoiding duplication where appropriate.

Key takeaways

- The commercial label of the system is not decisive: what matters is whether it performs a safety role or whether its failure could create a safety hazard.

Under Article 6(2), standalone AI systems are high-risk where their intended purpose falls within one of the Annex III use cases

Rationale	• **Article 6(2) classifies certain standalone AI systems as high-risk where, based on their intended purpose, they fall within one of the use cases listed in Annex III.** Classification as high-risk does not prohibit the use of such systems; rather, it subjects them to appropriate requirements ensuring accuracy, proper risk mitigation, and fundamental rights protection.
Eight Annex III areas	Annex III identifies eight broad areas susceptible to AI-related risks: • **Biometrics:** remote biometric identification, biometric categorization, and emotion recognition. • **Critical infrastructure**: AI systems as safety components in digital infrastructure, road traffic, and supply of water, gas, heating, or electricity. • **Education and vocational training:** systems determining access to education, evaluating learning outcomes, and monitoring student behavior during assessments. • **Employment, workers' management, and access to self-employment:** AI systems for recruitment, selection, task allocation, performance monitoring, and decisions on terms of employment or termination. • **Access to and enjoyment of essential private and public services and benefits:** systems evaluating eligibility for public assistance, creditworthiness, insurance risk/pricing, and emergency call triage. • **Law enforcement:** victim risk assessment, polygraphs, evidence evaluation, offending risk assessment, and profiling of persons. • **Migration, asylum, and border control management:** risk assessment of persons, examination of applications, polygraphs, and detection/identification of persons. • **Administration of justice and democratic processes:** AI assisting judicial authorities or in alternative dispute resolution, and systems influencing election outcomes or voting behavior.
Horizontal Principles for Annex III Classification	• **Human involvement does not alter classification.** Since human involvement cannot change the intended purpose of a system, it has no effect on whether the system is classified as high-risk. The provider cannot classify an AI system as 'low risk' simply by adding a requirement for human involvement. • **Natural persons.** Several use cases require that the AI system directly or indirectly evaluate "natural persons," which includes sole traders, self-employed persons, and individuals acting in a professional capacity. • **Complex and agentic AI systems.** AI systems operating as part of complex, interconnected configurations (including agentic AI) are assessed as a whole; exemptions do not apply to individual modules if the overall system materially influences key decisions. • **Intended to be used.** Classification depends on the system's intended purpose as defined by the provider; actual use is not required - the assessment occurs before placing the system on the market.

Key takeaways

- Annex III classification follows a functional approach: human involvement does not prevent high-risk classification where the AI system's intended purpose falls within a listed use case and materially influences individual decisions.

The Article 6(3) filter may exclude certain Annex III systems from high-risk classification, but only under narrow conditions

The Article 6(3) "Filter" Mechanism	Article 6(3) allows providers to exempt AI systems falling within Annex III use cases from high-risk classification where they do not materially influence decision-making and meet at least one of four conditions: • **Narrow procedural task**: the system transforms, categorizes, or structures data without performing value judgments (e.g., sorting applications by grade, detecting duplicates). • **Improves a previously completed human activity:** the system only refines or verifies an already-completed human result without reverting or replacing it (e.g., correcting grammar in a finalized decision). • **Detects decision-making patterns or deviations**: the system is used for quality assurance without replacing or influencing the previously completed human assessment (e.g., detecting grading inconsistencies). • **Preparatory task:** the system performs indexing, searching, or processing tasks prior to the assessment, with very low impact on the outcome (e.g., referencing relevant legal provisions).
Critical Limitation: Profiling	• The filter mechanism cannot be applied where the AI system performs profiling of natural persons within the meaning of Article 4(4) of the GDPR. Additionally, systems forming part of complex configurations whose combined outputs materially influence decisions in a high-risk use case cannot benefit from the filter.
Documentation Obligations	• Providers invoking the filter must document their assessment before placing the system on the market and must register the system in the EU database under Article 71 of the AI Act.
Anti-Circumvention	• Market surveillance authorities may evaluate whether a system has been misclassified as non-high-risk and impose penalties under Article 99 of the AI Act.

Key takeaways

- The Article 6(3) filter is a narrow exception: it only applies to procedural, preparatory or non-material tasks, and cannot be relied on where profiling or complex configurations materially influence high-risk decisions.

Management Solutions is experienced in reviewing and developing AI systems across all industries, while ensuring regulatory compliance and meeting supervisors' expectations.

1. **Specialized team.** MS has a team of **1,000+ Data Scientists** who combine **technical and quantitative skills with strong regulatory knowledge and certifications** in leading cloud providers (AWS, Azure and Google).

2. **AI models and regulatory practice.** MS has led the **development of numerous AI models** (supervised learning, unsupervised learning, NLP techniques, deep learning models...) with application in **multiple use cases**: fraud detection, risk classification, energy prediction, AML, XAI, and reputational risk or model risk measurement, among others. At the same time, MS has been involved in the implementation of various regulatory requirements across different industries (financial, telco, insurance...).

3. **Experience with regulators and supervisors.** MS is a "**highly qualified external service provider" to the European Central Bank**, with which it has signed 7 framework agreements in the last 6 years, and to national authorities. For the interpretability of advanced models, **MS works under the recommendations of the EBA in its "Report on Big Data and Advanced Analytics"**, according to its 7 elements of confidence for model development and interpretability.

4. **Interpretable models.** MS has **extensive experience in the development of interpretable models** and the application of interpretability techniques in the industries in which it operates: banking, insurance, energy, telecommunications and other industries.

5. **R&D area.** MS allocates **10% of its capacity to R&D,** allowing it to be at the forefront of Artificial Intelligence. **Co-founding of the iDANAE chair** (intelligence, data, analysis and strategy) **with the UPM** (Universidad Politécnica de Madrid), focused on the development of components that form part of the value cycle of the most important assets of today's society, such as information and knowledge.

6. **Close relationship with the RAC** (Royal Academy of Sciences) and active participation in several **research projects with AI applications in areas such as sustainability** (quantification of climate risk) **and efficient training of neural networks** (training optimization and interpretability in transfer learning architectures).

7. **In-house development of proprietary tools ModelCraft™,** with advanced AI/XAI techniques covering **multiple areas of advanced modeling, including dashboards and proprietary interpretability modules**, as well as management and **definition of architectures and cloud services**; **Gamma™,** a **model governance and MRM** tool, incorporating inventory, workflow management, document repository and MRM reporting; **and Hatari™,** a **reputational risk quantification** tool based on information from media and social networks, using **innovative artificial intelligence and NLP techniques.**

B. Abbreviations

Abbreviation	Meaning
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act
AML	Anti-Money Laundering
CER Directive	Critical Entities Resilience Directive
CJEU	Court of Justice of the European Union
EC	European Commission
EU	European Union
EUDPR	European Union Data Protection Regulation
FRIA	Fundamental Rights Impact Assessment
GDPR	General Data Protection Regulation
GPAI	General-Purpose AI
LED	Law Enforcement Directive
MRM	Model Risk Management

Abbreviation	Meaning
MSA	Market Surveillance Authority
NLF	New Legislative Framework
NLP	Natural Language Processing
OJEU	Official Journal of the European Union
RBI	Remote Biometric Identification
TFEU	Treaty on the Functioning of the European Union
UHL	Union Harmonisation Legislation
XAI	Explainable Artificial Intelligence

Management Solutions
Making things happen

International *One Firm*
Multiscope Team
Best practice *know-how*
Proven Experience
Maximum Commitment

Javier Calvo
Partner at Management Solutions
javier.calvo.martin@managementsolutions.com

Marta Hierro
Partner at Management Solutions
marta.hierro@msspain.com

For more information please visit

www.managementsolutions.com

Or follow us at: