EU AI Act

Risk-Based Approach (RBA) to AI Systems

Guidelines on the characteristics of a RBA to AI systems oversight, and the steps to be taken when conducting oversight on a risk-sensitive basis

Version 1.0

Last Updated: 22 November 2023

Contents

Preface

In the European Union ("EU"), there is now an obligation to have effective procedures in place to identify, manage and mitigation risks arising from AI systems. The EU Artificial Intelligence ("AI") Act (the "EU AI Act")[i], applying to deployers, developers and users of AI systems, originating in 2021, the Many of the procedures which will be appropriate to address these obligations are similar, and firms can often employ the same systems and controls to meet them.

Purpose of the guidance

The purpose of this guidance is to:

- outline the legal and regulatory framework for AI system requirements under the EU AI Act;
- interpret the requirements of the relevant law and regulations, and how they may be implemented in practice;
- indicate good industry practice in AI system procedures through a proportionate, risk-based approach; and
- assist firms to design and implement the systems and controls necessary to mitigate the risks arising from the development, deployment and use of AI systems.

Scope of the guidance

This guidance sets out what is expected of companies and their staff in relation to identifying, managing and mitigating the risks arising from AI systems under the EU AI Act, but allows them some discretion as to how they apply the requirements of the EU AI Act in the particular circumstances of the firm, and its AI systems.

This guidance relates solely to how firms should fulfil their obligations under the EU AI Act. It is important that firms understand that identification of the relevant AI system does not automatically qualify them for deployment of the AI system to the market; firms must also comply with other, commercial considerations in deciding whether particular AI systems should be deployed.

What are AI systems?

AI systems can be defined as, "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."

The risks posed by AI systems constantly evolve to match the level of innovation, and research and development, and the legislative/regulatory/law enforcement environment of the market in which the developer and deployer wishes to operate.

There are four main risk levels of AI systems

Term	Description
Minimal Risk	The proposal allows the free use of minimal-risk AI. This includes applications such as AI-enabled video games or spam filters. The vast majority of AI systems currently used in the EU fall into this category.
Limited Risk	Limited risk refers to AI systems with specific transparency obligations. When using AI systems such as chatbots, users should be aware that they are interacting with a machine so they can take an informed decision to continue or step back.

Term	Description
High-Risk	AI systems identified as high-risk include AI technology used in: • critical infrastructures (e.g. transport), that could put the life and health of citizens at risk; • educational or vocational training, that may determine the access to education and professional course of someone's life (e.g. scoring of exams); • safety components of products (e.g. AI application in robot-assisted surgery); • employment, management of workers and access to self-employment (e.g. CV-sorting software for recruitment procedures); • essential private and public services (e.g. credit scoring denying citizens opportunity to obtain a loan); • law enforcement that may interfere with people's fundamental rights (e.g. evaluation of the reliability of evidence); • migration, asylum and border control management (e.g. verification of authenticity of travel documents); • administration of justice and democratic processes (e.g. applying the law to a concrete set of facts).
Unacceptable Risk	All AI systems considered a clear threat to the safety, livelihoods and rights of people will be banned, from social scoring by governments to toys using voice assistance that encourages dangerous behaviour.

It is a offence under the EU AI Act to either deploy an unacceptable risk AI system on the market or, for high-risk AI systems, not comply with the pre- and post-market deployment requirements.

Firms increasingly look at AI system risks as part of an overall risk management strategy, and there are many similarities – as well as differences - between the two. When considering AI system risks, firms should consider their procedures against the EU AI Act and how these might reinforce each other. Where responsibilities are given to different departments, there will need to be strong links between those in the firm responsible for managing and reporting on these various areas of risk. When measures involving the public are taken specifically as a risk management measure, the distinction should be made clear.

Who is the guidance addressed to?

The guidance prepared by AI & Partners is addressed to firms classified as developers or deployers under the EU AI Act. All such firms – which, for the avoidance of doubt, may include regulated and unregulated firms- should have regard to the contents of the guidance.

The guidance will be of direct relevance to senior management, nominated officers and risk officers in any industry. The purpose is to give guidance to those who set the firm's risk management policies and its procedures for identifying, managing and mitigating AI system risks.

Although the guidance will be relevant to operational areas, it is expected that these areas will be guided by the firm's own, often more detailed and more specific, internal arrangements, tailored by senior management, nominated officers and risk officers to reflect the risk profile of the firm.

How should the guidance be used?

The guidance gives firms a degree of discretion in how they comply with the EU AI Act, and on the procedures that they put in place for this purpose.

It is not intended that the guidance be applied unthinkingly, as a checklist of steps to take. Firms should encourage their staff to 'think risk' as they carry out their duties within the legal and regulatory framework governing AI systems. Regulators have made clear its expectation that firms address their management of risk in a thoughtful and considered way, and establish and maintain systems and procedures that are appropriate, and proportionate to the risks identified. This guidance assists firms to do this.

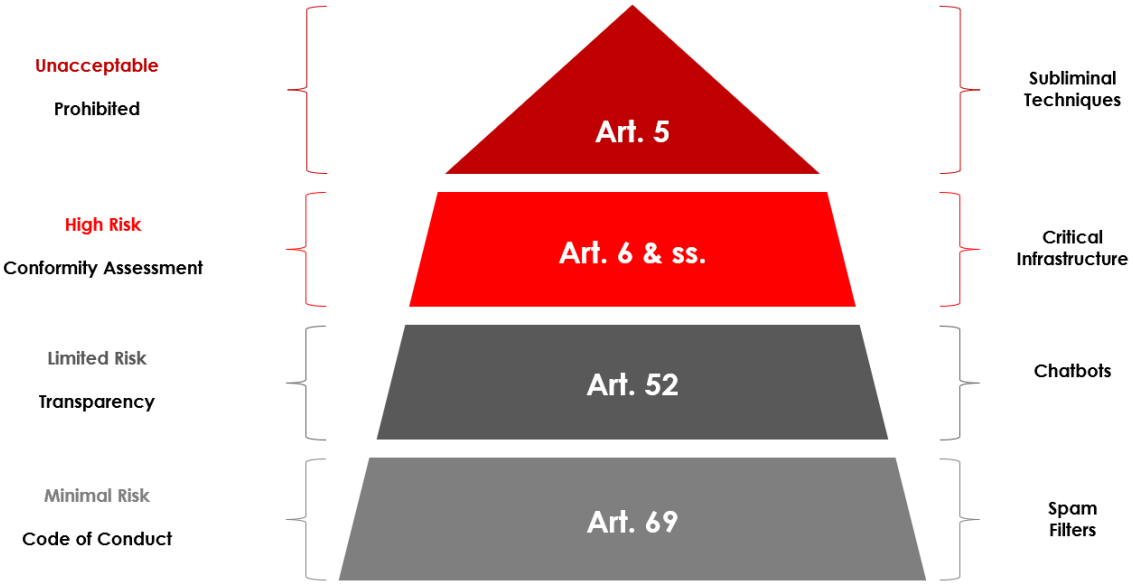
When provisions of the statutory requirements and of regulators' regulatory requirements are directly described in the text of the guidance, it uses the term 'must', indicating that these provisions are mandatory. In other cases, the guidance uses the term 'should' to indicate ways in which the statutory and regulatory requirements may be satisfied, but allowing for alternative means of meeting the requirements. References to 'must' and 'should' in the text should therefore be construed accordingly.

The content of the guidance

This guidance emphasises the responsibility of senior management to manage the firm's AI system risks, and how this should be carried out on a risk-based approach. It sets out a standard approach to the identification, management and mitigation of AI systems, separating out basic details from other aspects of AI system risk management measures, as well as giving guidance on the obligation to monitor AI systems.

The guidance incorporates a range of reference material which it is hoped that senior management, nominated officers and risk officers will find helpful in appreciating the overall context of, and obligations within, the EU AI Act.

Figure 1: Risk-Based Approach to AI Systems (Risk Pyramid)



What is a Risk-Based Approach?

In this document guidelines, the risk-based approach to AI systems is described as a cyclical process.

- Step 1 is the identification of an AI system's risk level based on the risks of harm posed to individuals health and safety, and fundamental rights, whereby firms obtain information on both domestic and foreign risks of harm affecting the relevant AI systems;
- Step 2 is the risk assessment, whereby firms obtain a holistic view of the risk level to which each subject of AI system assessment is exposed;
- Step 3 is the allocation of AI system supervisory resource in a way that is commensurate with the risk of harm identified. This includes decisions about the focus, depth, duration and frequency of supervision, and supervisory staffing needs, including technical expertise; and
- Step 4 is monitoring and review to ensure the risk assessment and associated allocation of supervisory resource remains up to date and relevant. This means that Step 4 can initiate again the identification of relevant information (Step 1), which may inform a new or updated risk assessment (Step 2), which in turn triggers new supervisory actions to mitigate those risks (Step 3).

Figure 2: Risk-Based Approach to AI Systems (Risk Process)

Definitions

For the purposes of this document, the following definitions apply.

Term	Description
Cluster	Means a group of AI systems of assessment having similar characteristics.
Firm	Means a developer or deployer of an **AI System**
Inherent Risk of Harm	Means the level of risk posed by an **AI System** to an individual's health and safety, or fundamental rights.
Risk-Based Approach (RBA)	Means an approach whereby **Firms** identify, assess and understand the risks of harm posed by **AI Systems** and take risk-sensitive measures that are proportionate to those risks
Risk of Harm	Means the likelihood and impact of risk of harm taking place. Risk refers to inherent risk.
Risk of Harm Factors	Means variables that, either on their own or in combination, may increase or decrease the risk of harm.
Risk Profile	Means the overall characteristics (including type and level) of risk that remains after mitigation.
Threat	Means the potential harm caused by a **AI System**. In the EU AI Act context, this includes the potential harm caused by **AI Systems**, as well as related past, present and future activities.

AI System	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.

Risk-Based Approach

Relevant Law/Regulation • EU AI Act **Other authoritative pronouncements which endorse a risk-based approach** • OECD Framework for the Classification of AI Systems[ii] • European Commission Policy and investment recommendations for trustworthy Artificial Intelligence[iii]
Core obligations • Identify and assess the risk of harm posed by AI systems to individuals, health/safety, security and fundamental rights • Appropriate systems and controls must reflect the degree of risk associated with the business and its customers • Determine appropriate systems and controls on a risk-sensitive basis, depending on the type of AI system • Take into account situations, factors and AI systems which by their nature can present a higher risk of harm
Actions required, to be kept under regular review • Carry out a formal, and regular, AI system risk assessment, including data modifications, and changes in AI systems, and the wider environment • Ensure internal policies, controls and procedures, including staff awareness, adequately reflect the risk assessment • Ensure AI system identification and adoption procedures reflect their risk characteristics • Ensure arrangements for monitoring systems and controls are robust, and reflect the risk characteristics of AI systems

Introduction and Legal Obligations

General

There are a number of discrete steps in assessing the most cost effective and proportionate way to manage and mitigate the AI system risks faced by the firm.

These steps are to:

- identify the money laundering, terrorist financing and proliferation financing risks that are relevant to the firm;
- assess the risks presented by the firm's particular AI systems;
- design and implement controls to manage and mitigate these assessed risks, in the context of the firm's risk appetite;
- monitor and improve the effective operation of these controls; and
- record appropriately what has been done, and why.

Whatever approach is considered most appropriate to the firm's AI system risk, the broad objective is that the firm should know at the outset of the relationship what its AI systems are, where they are located, what they do, who uses them, what risk level they are, and their expected level of activity with the firm. The firm then should consider how the profile of the AI system's capabilities builds up over time, thus allowing the firm to identify risks that that may pose a greater threat of harm.

Risk Assessment

(Art 3(1))

The EU AI Act requires firms to take appropriate steps to identify, assess, manage and mitigate the risks of harm posed by AI systems, taking into account:

- information on AI systems made available to them by regulators;
- risk factors, including factors relating to their AI systems.

In considering what steps are appropriate, firms must take into account the size and nature of its business. Firms that do not offer unacceptable or high-risk AI systems and that have limited or no EU exposure may not need an overly complex or sophisticated business risk assessment.

(Art 3(1))

The risk assessments carried out must be documented, kept up to date and made available to regulators on request. Regulators may decide that a documented risk assessment in the case of a particular firm is not required where the specific risks inherent in an AI system in which the firm deploys are clear and understood.

Obligation to adopt a risk-based approach

(Recital 14)

Senior management of most firms, whatever business they are in, manage the firm's affairs with regard to the risks inherent in the AI systems the firm deploys, those risks inherent in its AI systems and the effectiveness of the controls it has put in place to manage these risks.

To assist the overall objective to prevent the risk of harm posed by AI systems, a risk-based approach:

- recognises that these threats to individuals varies across AI systems;
- allows management to differentiate between their AI systems in a way that matches the risk in their particular business;
- allows senior management to apply its own approach to the firm's procedures, systems and controls, and arrangements in particular circumstances; and
- helps to produce a more cost-effective system.

(Recital 14)

A firm therefore uses its assessment of the risks inherent in its AI systems to inform its risk-based approach to the identification and verification of individual AI systems, which will in turn drive the level and extent of risk management measures appropriate to that AI system.

No system of checks will detect and prevent all risks associated with an AI system. A risk-based approach will, however, serve to balance the cost burden placed on individual firms and their AI systems with a realistic assessment of the threat of harm. It focuses the effort where it is needed and will have most impact.

The appropriate approach in any given case is ultimately a question of judgment by senior management, in the context of the risks they determine the AI systems poses.

Risk Assessment – Identification and Assessment of AI System Risks

(Art 3(1))

A firm is required to assess the risks inherent in its AI systems, taking into account risk factors including those relating to transparency and explainability (i.e. 'Know Your AI System' (KYAIS)).

(Art 3(1))

The risk environment posed by AI systems includes the wider context within which the firm operates – including the risks posed by the jurisdictions in which its AI systems are deployed. Risks are posed not only in relation to the extent to which the firm has, or has not, been able to carry out the appropriate level of due-diligence in relation to the AI system, nor by what the AI system is (are), but also in relation to the activities undertaken by the AI system – whether in the normal course of its business, or through its means of interaction.

(Art 3(1))

The business of many firms, and AI systems, can be relatively simple, involving few AI systems, with most AI systems falling into similar risk categories. In such circumstances, a simple approach, building on the risk the firm's AI systems are assessed to present, may be appropriate for most AI systems, with the focus being on those AI systems who fall outside the 'norm'. Other firms may have a greater level of business, but large numbers of their AI systems may be predominantly minimal or limited risk. Here, too, the approach for most AI systems may be relatively straightforward, building on the AI system risk.

(Art 3(1))

For firms which operate internationally, or which have AI systems deployed abroad, there are additional risk considerations relating to the position of the jurisdictions involved, and their reputation and standing as regards the inherent AI system risk, and the effectiveness of their EU AI Act enforcement regime.

(Art 3(1))

In identifying its AI system risk a firm should consider a range of factors, including the following AI system dimensions

- People & Planet;
- Economic Context;
- Data & Input;
- AI model; and
- Task & Output.

(Art 3(1))

The firm should therefore assess its AI system risks in the context of how they might most likely pose a threat of harm. In this respect, senior management should ask themselves a number of questions, for example:

- What risk is posed by the firm's AI systems?
- What risk is posed by a AI system's characteristics?
- How does the way the AI system is deployed by the firm affect the risk?
- What risk is posed by the AI system end-users interact with?

New technologies

(Recital 5)

In identifying and assessing the risk of harm, firms must take account of whether new AI systems and new business practices are involved, including new delivery mechanisms, and the use of new or developing technologies for both new and pre-existing new AI systems. As well as being specifically required in assessing whether there is a high risk of harm in a particular situation, such a risk assessment should take place prior to the launch of the new AI systems, or the use of new or developing technologies. Appropriate measures should be taken to manage and mitigate those risks, including where relevant in particular cases the application of appropriate risk management measures.

A Risk-Based Approach – Design & implement controls to manage and mitigate the risks

(Art [29(1a))

Once the firm has identified and assessed the risks posed by AI systems in respect of safety security, and fundamental rights, senior management must establish and maintain policies, controls and procedures to mitigate and manage effectively these risks identified in its risk assessment. These policies, controls and procedures must take account of the size and nature of the firm's business.

The policies, controls and procedures designed to mitigate assessed risk of harm, risks should be appropriate and proportionate to these risks, and should be designed to provide an effective level of mitigation.

(Art [29(1a))

Firms must obtain approval from their senior management for the policies, controls and procedures that they put in place and for monitoring and enhancing the measures taken.

A risk-based approach requires the full commitment and support of senior management, and the active co-operation of business units. The risk-based approach needs to be part of the firm's philosophy, and as such reflected in its procedures and controls. There needs to be a clear communication of policies, controls and procedures across the firm, along with robust mechanisms to ensure that they are carried out effectively, weaknesses are identified, and improvements are made wherever necessary.

(Art [29(1a))

The policies, controls and procedures must include, but are not limited to:

- risk management practices, reporting, record-keeping, internal controls, compliance management and AI system screening;
- where appropriate with regard to the size and nature of the business, an independent audit function to examine and evaluate the firm's policies, controls and procedures.
- for parent firms, policies on the sharing of information about AI systems.

(Art [29(1a))

The nature and extent of risk of harm, controls will depend on a number of factors, including:

- The nature, scale and complexity of the firm's business
- The diversity of the firm's operations, including geographical diversity
- The firm's AI system profile
- The distribution channels used
- The number and complexity of AI systems
- The extent to which the firm is dealing directly with the customer or is dealing through intermediaries, third parties, correspondents or non-face to face access.

(Art 9)

The application of risk management measures is intended to enable a firm to form a reasonable belief that it knows the true risk category of each AI system and, with an appropriate degree of confidence, knows the types of interactions the AI system is likely to conduct. The firm's procedures should include procedures to:

- Identify and verify the risk level of each AI system on a timely basis
- Identify and take reasonable measures to verify the risk level of any AI system
- Obtain appropriate additional information to understand the AI system's characteristics, including the expected complexity

(Art 9)

How a risk-based approach is implemented will depend on the firm's operational structure. For example, a firm that operates through multiple business units will need a different approach from one that operates as a single business. Equally, it will also be relevant whether the firm operates through branches or subsidiary undertakings; whether their business is principally face to face or online; whether the firm has a high staff/AI system ratio and/or a changing AI system base, or a small group of data scientists and a relatively stable AI system base; or whether their AI system is deployed internationally or largely domestic.

A Risk-Based Approach – AI System Risk Assessments

General

(Art 9)

Based on the risk assessment carried out, a firm will determine the level of risk management that should be applied in respect of each AI system. It is likely that there will be a standard level of risk management that will apply to the generality of AI system, based on the firm's risk appetite.

As regards risks of harm, managing and mitigating the risks will involve measures to verify the AI system's risk category; collecting additional information about the AI system; and monitoring their activity, to determine whether there are reasonable grounds for knowing or suspecting that the risk level may have changed. Firms must determine the extent of their risk management measures on a risk-sensitive basis depending on the type of AI system.

To decide on the most appropriate and relevant controls for the firm, senior management should ask themselves what measures the firm can adopt, and to what extent, to manage and mitigate these threats/risks most cost effectively, and in line with the firm's risk appetite. Examples of control procedures include:

- Introducing a AI system risk category identification programme that varies the procedures in respect of AI systems appropriate to their assessed risk of harm; and
- Monitoring AI systems. It is possible to try to assess the extent to which each AI system should be subject to each of these checks, but it is the balance of these procedures as appropriate to the risk assessed in the individual AI system to which they belong that is relevant.

A AI system risk identification programme that is graduated to reflect risk could involve:

- a standard information dataset to be held in respect of all AI systems;
- a uniform verification requirement for all AI systems; and
- an approach to monitoring AI systems that reflects the risk assessed posed by the AI systems.

AI System Risk Assessments

(Art 9)

Although the risks of harm fundamentally arise through its AI system, the nature of their businesses and their activities, a firm must consider its AI system risks in the context of the wider environment inherent in the business and jurisdictions in which the firm's AI systems are deployed operate.

The risk of harm posed by an individual AI system may be assessed differently depending on where it is deployed.

In reaching an appropriate level of satisfaction as to whether the risk of harm posed by the AI system is acceptable and able to be managed, requesting more and more risk identification is not always the right answer – it is sometimes better to reach a full and documented understanding of what the AI system does, and the activities it is likely to undertake. Some business lines' use of AI systems carry an inherently higher risk of harm than others.

(Art 9)

If a firm cannot satisfy itself as to the risk category of a AI system; verify that AI system; or obtain sufficient information on the nature and intended purpose of the AI system, it must not deploy it and must terminate an existing one.

While a risk assessment should always be performed at pre-deployment of the AI system, for some AI systems, a comprehensive risk profile may only become evident once the AI system has begun interacting with end-users, making the monitoring of transactions and on-going reviews a fundamental component of a reasonably designed risk-based approach ("RBA"). A firm may also have to adjust its risk assessment of a particular AI system based on information received from a competent authority.

Some other firms, however, often (but not exclusively) those dealing in wholesale markets, may offer a more 'bespoke' AI systems to end-users, many of whom are already subject to extensive risk management by AI service providers for reasons other than safety, safety and fundamental rights. In such cases, the business of risk classifying the AI system will be more complex, but will take account of the considerable additional information that already exists in relation to the prospective AI system.

General Principles – Use of Risk Categories and Factors

(Art 9)

In order to be able to implement a reasonable RBA, firms should identify criteria to assess potential risks posed by an AI system. Identification of the risks of harm, to the extent that such risk can be identified, of AI systems, will allow firms to design and implement proportionate measures and controls to mitigate these risks.

(Art 9)

Risks of harm may be measured using a number of factors. Application of risk categories to AI systems can then provide a strategy for managing potential risks by enabling firms to subject AI systems to proportionate controls and oversight. The key risk criteria are: health and safety; and fundamental rights. The weight given to these criteria (individually or in combination) in assessing the overall risk of harm may vary from one institution to another, depending on their respective circumstances. Consequently, firms have to make their own determination as to the risk weights. Parameters set by the EU AI Act may limit a firm's discretion.

(Art 9)

Annex contains a fuller list of illustrative risk factors a firm may address when considering the risk of harm posed by AI systems.

(Art 9)

When assessing the risks of harm relating to types of AI systems, a firm should take into account risk variables relating to those risk categories. These variables, either singly or in combination, may increase or decrease the potential risk posed, thus impacting the appropriate level of risk management measures. Examples of such variables include:

- The purpose of an AI system
- The number of end-users to be serviced by a AI system
- The regularity or duration of the AI system

(Art 9)

When assessing risk, firms should consider all relevant risk factors before determining what is the overall risk category and the appropriate level of mitigation to be applied.

(Art 9)

A risk assessment will often result in a stylised categorisation of risk: unacceptable, high, minimal and limited. Criteria will be attached to each category to assist in allocating AI systems to risk categories, in order to determine the different treatments of identification, verification, additional risk management requirements and monitoring for each category, in a way that minimises complexity.

Weighting of Risk Factors

(Art 9)

When weighting risk factors, firms should make an informed judgment about the relevance of different risk factors in the context of a particular AI system. This often results in firms allocating different 'scores' to different factors – for example, firms may decide that a AI system's links to a business function associated with higher risk of harm is less relevant in light of the features of the AI system they seek.

(Art 9)

Ultimately, the weight given to each of these factors is likely to vary from AI system to AI system, and from one firm to another. When weighting factors, firms should ensure that:

- Weighting is not unduly influenced by just one factor;
- Economic or profit considerations do not influence the risk rating;
- Weighting does not lead to a situation where it is impossible for any AI system to be classified as high risk;
- Situations identified by national legislation or risk assessments as always presenting a high risk of harm cannot be over-ruled by the firm's weighting; and
- Firms are able to override any automatically generated risk scores where necessary. The rationale for the decision to override such scores should be documented appropriately.

(Art 9)

Where a firm uses automated systems, purchased from an external provider, to allocate overall risk scores to categorise AI system, it should understand how such systems work and how it combines risk factors to achieve an overall risk score. A firm must always be able to satisfy itself that the scores allocated reflect the firm's understanding of risk of harm, and it should be able to demonstrate this to the regulator if necessary.

Minimal/Limited Risk

(Art 9)

Many AI systems, by their nature or through what is already known about them by the firm, carry a lower risk of harm. These might include:

- AI systems that do not employ certain profiling techniques; and
- AI systems with a long-term relationship with the firm.

(Art 9)
Annex contains a fuller list of illustrative risk factors a firm may address when considering the risk of harm posed by AI systems.

(Art 9)
Having a lower risk of harm for risk identification purposes does not automatically mean that the same AI system is lower risk for all types of risk management measures, in particular for ongoing monitoring of AI systems.

(Art 9)
Firms should not, however, judge the level of risk solely on the nature of the AI system. Where, in a particular situation, the AI system is considered to carry a higher risk of harm, the overall risk of the AI system should be considered carefully. Firms need to be aware that allowing a higher risk AI system to interact with end-users on the basis of a verification standard that is appropriate to a lower risk AI system, can lead to a requirement for further risk management requirements, particularly if the AI end-user wishes subsequently to interact with a higher risk AI system.

High-Risk

(Art [6)
Where higher risks of harm are identified, firms are required to take appropriate risk-management measures to manage and mitigate the risks. Profiling AI systems have been identified as high-risk.

(Art [6)
Where a AI system is assessed as carrying a higher risk, then depending on its purpose, it will be necessary to apply robust risk management measures in respect of the AI system.

(Art [6)
Where the risks of harm are higher, firms must conduct enhanced risk management measures consistent with the risks identified.

(Art [6)
Identifying a AI system as carrying a higher risk of harm does not automatically mean that it is harmful. Similarly, identifying a AI system as carrying a low risk of harm does not mean that the AI system is not. Staff therefore need to be vigilant in using their experience and common sense in applying the firm's risk-based criteria and rules.

A Risk-Based Approach – Monitor and Improve the Effective Operation of the Firm's Controls

(Art 9)
The policies, controls and procedures should be approved by senior management, and the measures taken to manage and mitigate the risks of harm (whether higher or lower) should be consistent with national requirements and with guidance from competent authorities.

Independent testing of, and reporting on, the development and effective operation of the firm's RBA should be conducted by, for example, an internal audit function (where one is established), external auditors, specialist consultants or other qualified parties who are not involved in the implementation or operation of the firm's EU AI Act compliance programme.

(Art 9)

The firm will need to have some means of assessing that its risk mitigation procedures and controls are working effectively, or, if they are not, where they need to be improved. Its policies, controls and procedures will need to be kept under regular review. Aspects the firm will need to consider include:

- appropriate procedures to identify changes in AI system characteristics, which come to light in the normal course of business;
- reviewing ways in which different AI system may be used for harmful purposes, and how these ways may change, supported by typologies/law enforcement feedback, etc;
- adequacy of staff training and awareness;
- monitoring compliance arrangements (such as internal audit/quality assurance processes or external review);
- where appropriate, the establishment of an internal audit function;
- the balance between technology-based and people-based systems;
- capturing appropriate management information;
- upward reporting and accountability;
- effectiveness of liaison with other parts of the firm; and
- effectiveness of the liaison with regulatory and law enforcement agencies.

A Risk-Based Approach – Monitor and Improve the Effective Operation of the Firm's Controls

(Art 29 (1a))

Firms must document their risk assessments in order to be able to demonstrate their basis, keep these assessments up to date, and have appropriate mechanisms to provide appropriate risk assessment information to competent authorities.

(Art 29 (1a))

Annex contains illustrative examples of systems and controls a firm might have in place in order to keep its risk assessments up to date.

(Art 29 (1a))

The responses to consideration of the issues set out above, or to similar issues, will enable the firm to tailor its policies and procedures on the prevention of risks of harm. Documentation of those responses should enable the firm to demonstrate to its regulator and/or to a court:

- how it assesses the threats/risks of being used in connection with risks of harm;
- how it agrees and implements the appropriate systems and procedures, including risk management requirements, in the light of its risk assessment;
- how it monitors and, as necessary, improves the effectiveness of its systems and procedures; and
- the arrangements for reporting to senior management on the operation of its control processes.

(Art 29 (1a))

In addition, on a case-by-case basis, firms should document the rationale for any additional risk management measures it has undertaken (or any it has waived) compared to its standard approach, in view of its risk assessment of a particular AI system.

Risk Management is Dynamic

(Art 9)

Risk management generally is a continuous process, carried out on a dynamic basis. A AI system risk assessment is not a one-time exercise. Firms must therefore ensure that their risk management processes for managing risks of harm are kept under regular review.

There is a need to monitor the environment within which the firm operates. Success in preventing risk of harm in one area of operation or business will tend to drive malicious actors to migrate to another area, business, or product stream. Periodic assessment should therefore be made of activity in the firm's market place. If evidence suggests that displacement is happening, or if AI system behaviour is changing, the firm should be considering what it should be doing differently to take account of these changes.

In a stable business change may occur slowly - most businesses are evolutionary. AI systems' activities change (without always notifying the firm) and the firm's AI systems – and the way these are deployed or marketed to end-users – change. The AI systems interacting with end-users will also vary as aspects of their relative vulnerability change.

There is, however, a balance to be achieved between responding promptly to environmental changes, and maintaining stable systems and procedures.

A firm should therefore keep its risk assessment(s) up to date. An annual, formal reassessment might be too often in most cases, but still appropriate for a dynamic, growing business. It is recommended that a firm revisit its assessment at least annually, even if it decides that there is no case for revision. Firms should include details of the assessment, and any resulting changes, in the risk officer's annual report.

Annex A

Considerations in Assessing the Level of Risk of Harm of Different AI Systems

This Annex is designed to assist firms by setting out how they might approach their assessment of AI systems, to determine their level of risk of harm. The Annex discusses AI systems where there may be a presumption of low risk, and those where such a presumption may not be appropriate without further investigation. It then discusses issues that a firm should consider in all cases when coming to a judgment on the level of risk of harm implicit in any particular AI system.

Implications of an Assessment as Minimal or Limited Risk

Assessment of a AI system as limited or minimal risk only allows for some easement of the level of risk management carried out – it is not a complete exemption from the application of risk management measures in respect of AI systems. It does not exempt the firm from carrying out ongoing monitoring of AI systems, nor from the need for such other procedures as may be necessary to enable a firm to fulfil its responsibilities under the EU AI Act.

Although the judgment on the risk level is one to be made by each firm in the light of the particular circumstances, senior management is accountable for this judgment – either to its regulator, or, if necessary, to a court. It is therefore important that the reasons for concluding that a particular

jurisdiction is limited or minimal risk (other than those in respect of which a presumption of limited or minimal risk may be made) are documented at the time the decision is made, and that it is made on relevant and up to date data or information.

Annex B

Illustrative Risk Factors Relating to AI Systems

Note: These are risk factors that may be relevant for consideration during the course of risk assessments but do not automatically indicate a higher risk.

System Purpose:

High-risk AI systems were defined based on their intended purpose, particularly those with potential significant societal or individual impact.

What specific societal or individual impacts are anticipated from the AI system's intended purpose, and how are these impacts evaluated for potential risks?

Have potential unintended consequences or negative impacts on society or individuals been thoroughly assessed in relation to the AI system's purpose?

Technical Robustness and Accuracy:

The proposal emphasized the need to assess the technical robustness and accuracy of AI systems to ensure their reliability and minimize errors or biases.

How is the technical robustness of the AI system measured, and what measures are in place to continually assess and enhance its reliability?

In what ways does the AI system address and minimize errors or biases during its operation, and how are these measures integrated into its design?

Data Quality and Governance:

The quality of training data and the governance processes related to data usage were considered to avoid biases and inaccuracies in AI system outcomes.

What processes are implemented to ensure the quality of the training data used by the AI system, and how is data governance structured to prevent biases and inaccuracies?

How does the AI system handle and mitigate potential biases in the data, and what steps are taken to ensure fair and accurate outcomes?

Transparency and Explainability:

The transparency and explainability of AI systems were highlighted to enable users to understand how decisions are made and to ensure accountability.

How does the AI system provide transparency in its decision-making process, and what mechanisms are in place to explain its outputs to users and stakeholders?

What efforts are made to enhance the explainability of the AI system's decisions, especially in complex or critical scenarios?

Human Oversight:

The degree of human oversight and control in AI systems was a crucial factor to prevent overreliance on automated decision-making and to allow for human intervention when necessary.

To what extent does the AI system incorporate human oversight, and how is this oversight designed to prevent overreliance on automated decision-making?

What mechanisms are in place to facilitate human intervention when the AI system encounters ambiguous situations or unforeseen circumstances?

Legal and Ethical Compliance:

Compliance with legal and ethical standards was a key consideration, ensuring that AI systems operate within the bounds of existing regulations and ethical guidelines.

How is the AI system designed to ensure compliance with existing legal regulations relevant to its operation?

In what ways does the AI system address ethical considerations, and how are ethical guidelines embedded in its design and deployment?

Annex C

Considerations in Keeping Risk Assessments Up To Date

Firms should keep their assessment of risk of harm associated with individual AI systems, as well as the underlying factors, under review to ensure their assessment of risk of harm remains up to date and relevant. Firms should assess information obtained as part of their ongoing monitoring of AI systems and consider whether this affects the risk assessment.

Firms should also ensure that they have systems and controls in place to identify emerging risks of harm and that they can assess and, where appropriate, incorporate these in their business-wide and individual AI system risk assessments in a timely manner.

Examples of systems and controls firms should put in place to identify emerging risks include:

- processes to ensure internal information is reviewed regularly to identify trends and emerging issues, both in relation to individual AI systems and the firm's AI systems;
- processes to ensure the firm regularly reviews relevant information sources. This should involve, in particular:
  - regularly reviewing media reports that are relevant to the AI systems the firm interacts with;
  - regularly reviewing law enforcement alerts and reports;
  - ensuring that the firm becomes aware of changes to relevant AI system incident alerts as soon as they occur, for example by regularly reviewing AI incident alerts; and

    - regularly reviewing thematic reviews and similar publications issued by competent authorities.
- processes to capture and reviewing information on risks relating to new AI systems;
- engagement with other industry representatives and competent authorities (such as round tables, conferences and training) and processes to feed back any findings to relevant staff; and
- establishing a culture of information sharing within the firm and strong company ethics.

Examples of systems and controls firms should put in place to ensure their individual and business-wide risk assessment remains up to date include:

- setting a date at which the next risk assessment update takes place, e.g. on the 1 March every year, to ensure new or emerging risks of harm are included in the risk assessment. Where the firm is aware that a new risk of harm has emerged, or an existing one has increased, this should be reflected in the risk assessment as soon as possible; and
- carefully recording issues throughout the year that could have a bearing on the risk assessment, such as internal AI incident reports, compliance failures and intelligence from staff.

Like the original risk assessments, any update of a risk assessment and adjustment of accompanying risk management measures should be proportionate and commensurate with the risk of harm.

---

[i] https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF

[ii] https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm

[iii] https://digital-strategy.ec.europa.eu/en/library/policy-and-investment-recommendations-trustworthy-artificial-intelligence