

Artificial Intelligence Risk Level Guidance

How to use this document.

This guidance describes when an AI system is “high-risk,” and includes considerations to help make that determination. The information in this document specifically addresses the minimum requirements to identify AI systems that pose a high risk to health, safety or fundamental rights. Agencies may prioritize other types of risks posed by AI systems. For example, an AI system could pose a high reputational or operational risk even if it does not impact health, safety or fundamental rights. Similarly, agencies may have different risk tolerances based on their operating context. Agencies are encouraged to consider additional types of risk.

What is a high-risk AI system?

A high-risk AI system is a system using AI technology that creates a high risk to natural persons' health, safety or fundamental rights. These risks include:

- Direct impacts, such as when an AI system is used to determine eligibility for benefits, and
- Indirect impacts, such as when an AI system is used to allocate resources in a way that impacts the people in a particular community.

Whether a system creates a high risk is dependent on (1) the magnitude of an impact to natural persons' health and safety or fundamental rights and (2) the likelihood of that impact occurring.

		Likelihood				
		1	2	3	4	5
Magnitude	5					
	4					
	3					
	2					
	1					

Magnitude	Likelihood
1 - Negligible. No foreseeable direct or indirect impact to natural persons.	1 - Remote or improbable. Very low chance of occurring.
2 - Low. Any impact to natural persons is very unlikely to impact health, safety or fundamental rights.	2 - Unlikely. Low chance of occurring.
3 - Moderate. Some impact to natural persons that may include indirect impact to health, safety or fundamental rights.	3 - Possible. Moderate chance of occurring.
4 - Significant. Major effect causing substantial harm or disruption to health, safety or fundamental rights. May include direct impact in individual circumstances, or indirect, systemic impacts.	4 - Likely. High chance of occurring.
5 - Severe or catastrophic. Extreme impact resulting in serious harm, injury or violation of fundamental rights.	5 - Probable. Very high chance of occurring.

Potential risk factors.

When determining whether an AI system is high-risk, consider at least:

- The intended use and operating context
- Data characteristics
- Systems characteristics and safeguards

The intended use and operating context

- What is the intended use of system outputs and who will benefit?
- Who might be adversely impacted and how? Consider at least the workforce, customers/residents, and vulnerable or underserved populations.
- What is the role of humans in reviewing system outputs? For example, do they have little to no review, do humans review in defined circumstances, or do humans verify, modify, or review all outputs?
- Who are the intended users? Is it intended for internal or external use? Is it intended for broad use or a small group of people?
- What laws apply to agency use of the system?
- What alternatives were considered and why is this the preferred option?

Data characteristics

- What is the highest data classification level for included information? Consider both information used for training and information processed during operation, including intended prompts, inputs, fine-tuning, augmentation or any other system customization.
- What type of information is involved? Will the system process personal information?

System characteristics and safeguards

- Is there evidence of biased, discriminatory, inaccurate or otherwise unreliable outputs? What measures are in place to address unreliable outputs?
- Are there built-in filters or monitoring to detect harmful outputs, manipulation, or misuse?
- How explainable is the system? Do users know the system uses AI? Does the system provide sources or otherwise help identify to users why it is producing a particular output?
- How transparent is system implementation? Are users trained or notified of appropriate use and system limitations?