# Intelligenza artificiale

sottotitolo o riferimento normativo

***Tecniche e sistemi di cybersecurity***

*Ing. Paola Rocco*

*L.A. ISO/IEC 27001,L.A. ISO 22301, DPO UNI 11697:2017,*
*Valutatore Privacy UNI 11697:2017, PRINCE2®*
*CTU presso il Tribunale di Roma*

*9 Aprile 2025*

## Indice degli argomenti:

- **Conoscere le normative europee e le best practice per la sicurezza.**
- Comprendere i rischi e le minacce legate all'IA e alla sicurezza dei dati
- Valutare le misure di sicurezza da implementare

# La pubblicazione dell’AI Act dell’UE

- L’intelligenza artificiale (AI) deve essere affidabile, accessibile a tutti e sicura, soprattutto, più che l’uomo al centro, bisogna sempre ricordarci che questa tecnologia è uno strumento (dall’uomo sviluppato), è fondamentale essere formati per il suo corretto utilizzo, arrivando a sfruttarne le potenzialità e a riconoscerne i limiti, dettati dall’etica e dalla legge.

European Commission

European Parliament

Council of the European Union

# I driver

**Proporzionalità:**

approccio proporzionato basato sul rischio per non creare restrizioni inutili al commercio

**Trasparenza:**

i sistemi di IA devono essere progettati e sviluppati in modo da garantire la trasparenza delle loro decisioni e dei loro processi

**Sicurezza:**

i sistemi di IA devono essere progettati e sviluppati in modo da ridurre i rischi per la sicurezza e la salute delle persone, per i diritti fondamentali e per gli interessi pubblici

**Sussidiarietà**

solo per la competenza non esclusiva: un solido quadro normativo europeo assicura parità di condizioni e maggior tutela, rafforzando allo stesso tempo la competitività e la base industriale dell'Europa nel settore dell'IA

**Semplificazione**

normativa

# Definizione di Intelligenza Artificiale

- «Il sistema di intelligenza artificiale è un sistema basato su macchine progettato per funzionare con diversi livelli di autonomia e che può mostrare adattività dopo l'implementazione e che, per obiettivi espliciti o impliciti, deduce, dall'input che riceve, come generare output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali»

# Alcune date

| | | |
|---|---|---|
| 1° agosto 2024 | 20 giorni dopo la pubblicazione nella Gazzetta ufficiale | Entrata in vigore della legge |
| 2 febbraio 2025 | 6 mesi | Divieto sui sistemi di IA con rischio inaccettabile |
| 2 maggio 2025 | 9 mesi | Applicazione dei codici di condotta |
| 2 agosto 2025 | 12 mesi | Applicazione delle regole di governance e degli obblighi per l'AI di scopo generale |
| 2 agosto 2026 | 24 mesi | Inizio dell'applicazione dell'AI Act per i sistemi di IA (incluso l'Allegato III) |
| 2 agosto 2027 | 36 mesi | Applicazione dell'intero Regolamento per tutte le categorie di rischio (incluso l'Allegato II) |

# A chi si rivolge il nuovo AI act?

- Ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA nell'Unione, indipendentemente dal fatto che siano stabiliti nell'Unione o in un paese terzo
- Agli operatori di sistemi di IA che hanno il loro luogo di stabilimento o che sono ubicati nell'Unione
- Ai fornitori e agli operatori di sistemi di IA situati in un paese terzo laddove l'output prodotto dal sistema sia utilizzato nell'Unione.

A chi **non si applica:**

- Ai sistemi di IA sviluppati o usati per scopi esclusivamente militari o di difesa
- Ai sistemi di IA utilizzati solo a scopo di ricerca e innovazione
- Nel caso in cui l'intelligenza artificiale venga utilizzata nel corso di un'attività personale non professionale

# AI vietata se……

- (a) utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole al fine di distorcerne materialmente il comportamento in un modo che provochi o possa provocare un danno fisico o psicologico
- (b) sfrutta le vulnerabilità di uno specifico gruppo di persone, dovute all'età o alla disabilità fisica o mentale, al fine di distorcere materialmente il comportamento di una persona che appartiene a tale gruppo in un modo che provochi o possa provocare un danno fisico o psicologico
- (c) è usata da autorità pubbliche per la valutazione o classificazione dell'affidabilità delle persone fisiche per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali note o previste
- (d) fa uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, salvo che tale uso serva per la ricerca di vittime di reato o per la prevenzione di una minaccia specifica

# Art. 22 GDPR: divieto di processi automatizzati

*Articolo 22*

**Processo decisionale automatizzato relativo alle persone fisiche, compresa la profilazione**

1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona.

2. Il paragrafo 1 non si applica nel caso in cui la decisione:

a) sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento;

b) sia autorizzata dal diritto dell'Unione o dello Stato membro cui è soggetto il titolare del trattamento, che precisa altresì misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato;

c) si basi sul consenso esplicito dell'interessato.

3. Nei casi di cui al paragrafo 2, lettere a) e c), il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione.

4. Le decisioni di cui al paragrafo 2 non si basano sulle categorie particolari di dati personali di cui all'articolo 9, paragrafo 1, a meno che non sia d'applicazione l'articolo 9, paragrafo 2, lettere a) o g), e non siano in vigore misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato.

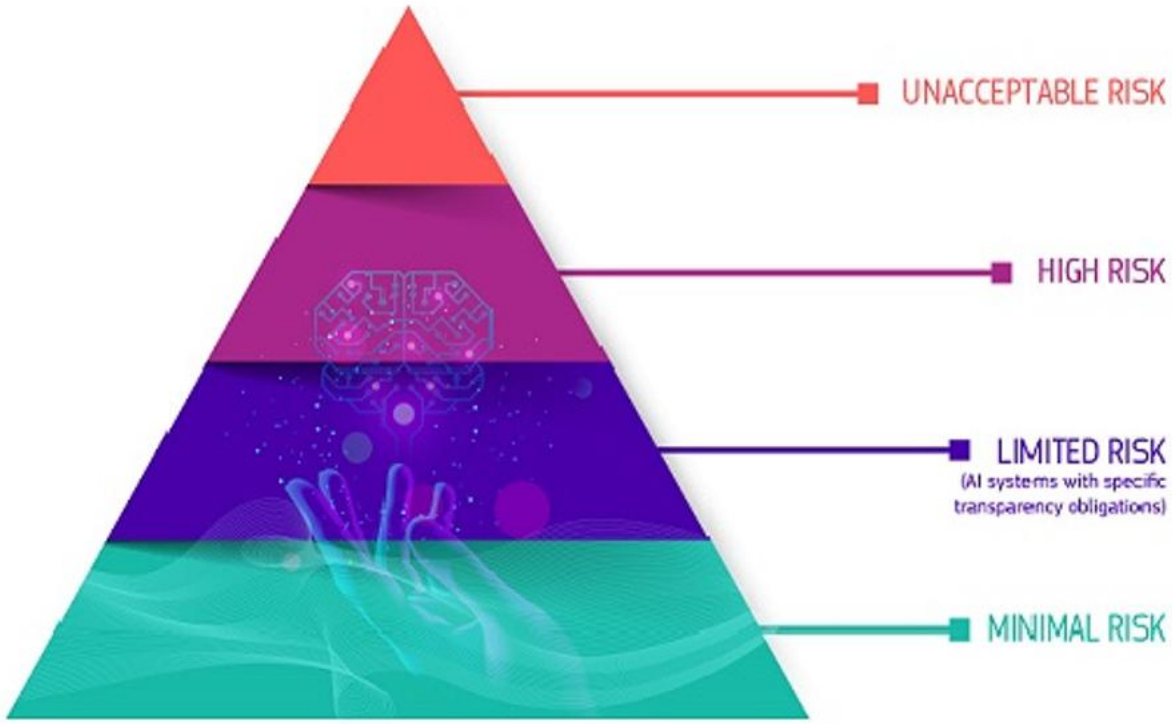
# Art. 22 GDPR: divieto di processi automatizzati

- diritto di non essere sottoposto a una (i) decisione (ii) basata unicamente sul trattamento automatizzato, compresa la profilazione, (iii) che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona
- il titolare del trattamento attua misure appropriate per tutelare i diritti, le libertà e i legittimi interessi dell'interessato, almeno il diritto:

  **di ottenere l'intervento umano da parte del titolare del trattamento**
  **di esprimere la propria opinione e di contestare la decisione**

- il titolare del trattamento deve informare l'interessato circa la logica della decisione

# Livelli di rischio



# Come definire il rischio

- Insieme delle possibilità di un evento e delle sue conseguenze sugli obiettivi (cfr. Norma UNI 11230)

| | |
|---|---|
| **Rischio** | Insieme delle possibilità di un evento e delle sue conseguenze sugli obiettivi |
| **Obiettivo** | Risultato da raggiungere |
| **Evento** | Accadimento di una serie di circostanze |
| **Conseguenza** | Esito di un evento |
| **Probabilità** | Misura o stima della possibilità che un evento ha di verificarsi |
| **Controllo** | Misura che mantiene e/o modifica il rischio |
| **Minaccia** | Causa o origine di un danno o perdita potenziali |
| **Vulnerabilità** | Debolezza di un asset o controllo che può essere sfruttata da una o più minacce |

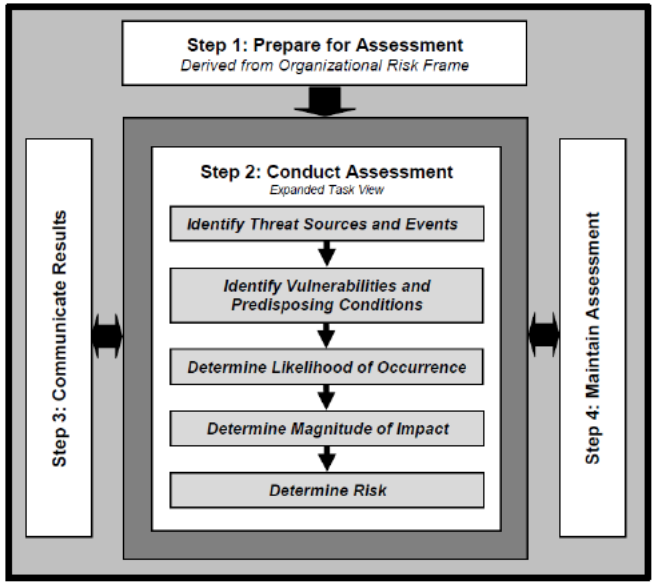
Definizioni simili si trovano nelle norme ISO 31000:2018 e ISO 27000:2018.

# ISO 31000:2018

- La norma ISO 31000:2018 Risk management —Guidelines definisce tutte le attività concrete necessarie per sviluppare una metodologia efficiente per la gestione del rischio

# Risk assessment

- Valutare il rischio significa analizzare le minacce e le vulnerabilità dell' infrastruttura in esame con lo scopo di determinare la probabilità che si verifichino eventi che potrebbero avere un impatto negativo sull'organizzazione




*NIST SP 800-30r1, Guide for Conducting Risk Assessments*

# Risk assessment

Il rischio è una combinazione di probabilità e di gravità:

$$R = P \times Vu \times Val$$

P = Probabilità dell'attacco

Vu = Vulnerabilità all'attacco

Val = Valore del danno provocato nel caso in cui l'attacco abbia successo

# Risk treatment

- Consiste nel definire un piano di trattamento tramite un elenco di controlli per affrontare i rischi
- Il piano di trattamento coinvolge misure per ridurre, conservare o evitare i rischi, oltre a misure per la valutazione dell'effettiva efficacia delle misure di trattamento messe in atto

# Risk Acceptance

- Consiste nella decisione di accettare i rischi e nella definizione delle responsabilità correlate
- L'organizzazione stabilisce un elenco di rischi consapevolmente accettati, con un'eventuale giustificazione per i rischi che non soddisfano i criteri di accettazione del 1°step

# Livelli di rischio

| LEGENDA DI VALUTAZIONE DEL RISCHIO | BASSA | MEDIO | ALTO | ESTREMO |
|---|---|---|---|---|
| | 0<br>ACCETTABILE | 1<br>ALARP<br>(As Low As Reasonably Practicable) | 2<br>GENERALMENTE INACCETTABILE | 3<br>INTOLLERABILE |
| | OK A PROCEDERE | INTRAPRENDERE SFORZI DI MITIGAZIONE | INTRAPRENDERE SFORZI DI MITIGAZIONE | METTERE L'EVENTO IN SOSPESO |

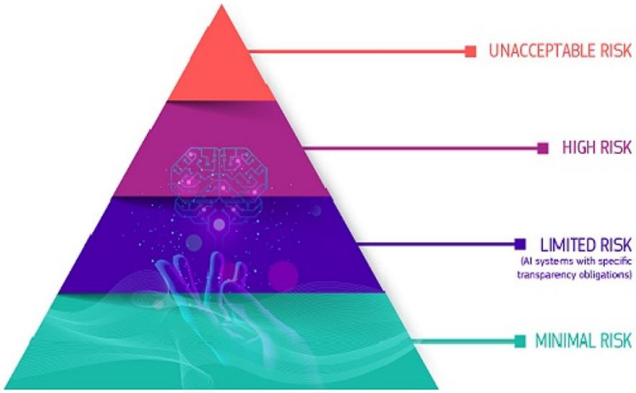
# Matrici di rischio

| | GRAVITÀ | | | |
|---|---|---|---|---|
| | ACCETTABILE<br>POCO O NESSUN EFFETTO SULL'EVENTO | TOLLERABILE<br>GLI EFFETTI SONO PERCEPITI, MA NON SONO CRITICI PER IL RISULTATO | NON DESIDERABILE<br>GRAVE IMPATTO SULLA LINEA D'AZIONE E RISULTATO | INTOLLERABILE<br>POTREBBE COMPORTARE UN DISASTRO |
| PROBABILITÀ | | | | |
| IMPROBABILE<br>IL RISCHIO È IMPROBABILE CHE SI VERIFICHI | BASSO<br>– 1 – | MEDIO<br>– 4 – | MEDIO<br>– 6 – | ALTO<br>– 10 – |
| POSSIBILE<br>IL RISCHIO PROBABILMENTE SI VERIFICHERÀ | BASSO<br>– 2 – | MEDIO<br>– 5 – | ALTO<br>– 8 – | ESTREMO<br>– 11 – |
| PROBABILE<br>IL RISCHIO SI VERIFICHERÀ | MEDIO<br>– 3 – | ALTO<br>– 7 – | ALTO<br>– 9 – | ESTREMO<br>– 12 – |

# Rischio (1/4)

## Rischio Inaccettabile:

vi rientra tutto ciò che rappresenta una minaccia per la sicurezza, i mezzi di sussistenza e i diritti delle persone. In questi casi, l'utilizzo dell'IA è vietata, in quanto in contrasto con i principi dell'Unione europea e i diritti fondamentali dell'uomo.



# Rischio Inaccettabile

## Esempi

Esempi di questa categoria includono un sistema di **social score**. In un sistema di social scoring immaginario, i cittadini guadagnano o perdono punti basati su comportamenti specifici. Punteggi alti possono portare a privilegi come prestiti agevolati e accesso prioritario ai servizi pubblici. Comportamenti negativi, come multe non pagate o comportamenti scorretti online, possono abbassare il punteggio, risultando in restrizioni come l'accesso limitato ai trasporti pubblici o a certi lavori. Questo punteggio influisce quindi sulla vita quotidiana, modellando l'accesso a risorse e opportunità basandosi sulla condotta individuale potenzialmente ledendo diritti che consideriamo fondamentali. Per questo i sistemi di IA che generano questo tipo di punteggio sono vietati.

# Rischio Inaccettabile

Tutti i sistemi di IA considerati una chiara minaccia alla sicurezza, ai mezzi di sussistenza e ai diritti delle persone sono vietati. La legge sull'IA vieta otto pratiche, vale a dire:

1.manipolazione e inganno dannosi basati sull’IA

2.sfruttamento dannoso delle vulnerabilità basato sull’IA

3.punteggio sociale

4.Valutazione o previsione del rischio di reato individuale

5.raschiatura non mirata di materiale di Internet o CCTV per creare o espandere database di riconoscimento facciale

6.riconoscimento delle emozioni nei luoghi di lavoro e negli istituti di istruzione

7.categorizzazione biometrica per dedurre determinate caratteristiche protette

8.identificazione biometrica remota in tempo reale a fini di contrasto in spazi accessibili al pubblico

# Rischio inaccettabile

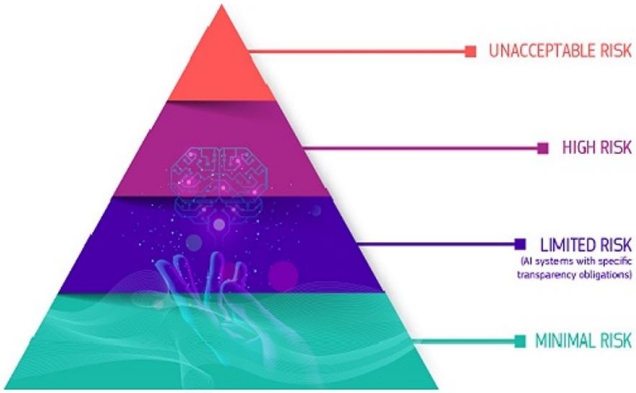
**Rischio inaccettabile: le pratiche vietate**

- la manipolazione comportamentale cognitiva
- lo ***scraping*** non mirato delle immagini facciali da Internet o da filmati di telecamere a circuito chiuso
- il **riconoscimento delle emozioni** sul luogo di lavoro e negli istituti di istruzione
- il punteggio sociale
- la **categorizzazione biometrica** per dedurre dati sensibili
- **polizia predittiva**

# Rischio (2/4)

## rischio elevato:

corrisponde ad un rischio significativo di danno alla salute, alla sicurezza o ai diritti fondamentali. Questi sistemi, tuttavia, sono consentiti ma devono rispettare un insieme di requisiti tecnici specifici prima di poter essere immessi nel mercato



# Rischio elevato

## Esempi

a seconda categoria considerata è quella di rischio elevato: rientra in questa categoria un'ampia gamma di sistemi che potrebbero causare danni significativi in caso di malfunzionamento: sistemi di identificazione biometrica e riconoscimento facciale, sistemi di valutazione del credito e scoring, chatbot medici, sistemi di guida autonoma.

Un esempio specifico in questa categoria è un sistema di riconoscimento facciale utilizzato per la sorveglianza in tempo reale in luoghi pubblici, In una grande città metropolitana, le autorità hanno installato telecamere di sorveglianza dotate di tecnologia di riconoscimento facciale in stazioni di metropolitana, aeroporti e piazze principali, potenzialmente prevenendo crimini o facilitando l'arresto di sospetti, ma l'uso di questi sistemi può essere legale solo nei termini strettamente previsti dalle norme.

# Rischio elevato - esempi

- Componenti di sicurezza dell'IA nelle infrastrutture critiche (ad esempio i trasporti), il cui guasto potrebbe mettere a rischio la vita e la salute dei cittadini

- Soluzioni di IA utilizzate negli istituti di istruzione, che possono determinare l'accesso all'istruzione e il corso della vita professionale di una persona (ad esempio il punteggio degli esami)

- Componenti di sicurezza dei prodotti basati sull'IA (ad es. applicazione dell'IA nella chirurgia assistita da robot)

- Strumenti di IA per l'occupazione, la gestione dei lavoratori e l'accesso al lavoro autonomo (ad esempio software di selezione dei CV per l'assunzione)

- Alcuni casi d'uso dell'IA utilizzati per dare accesso a servizi pubblici e privati essenziali (ad esempio il credit scoring che nega ai cittadini l'opportunità di ottenere un prestito) le sentenze dei tribunali)

# Rischio elevato - esempi

•Sistemi di IA utilizzati per l'identificazione biometrica remota, il riconoscimento delle emozioni e la categorizzazione biometrica (ad esempio un sistema di IA per identificare retroattivamente un taccheggiatore)

- casi d'uso dell'IA nelle attività di contrasto che possono interferire con i diritti fondamentali delle persone (ad esempio valutazione dell'affidabilità delle prove)

- Casi d'uso dell'IA nella gestione della migrazione, dell'asilo e del controllo delle frontiere (ad esempio esame automatizzato delle domande di visto)

- Soluzioni di IA utilizzate nell'amministrazione della giustizia e dei processi democratici (ad esempio soluzioni di IA per preparare le sentenze dei tribunali)

# Rischio elevato - esempi

•Alcuni esempi sono:

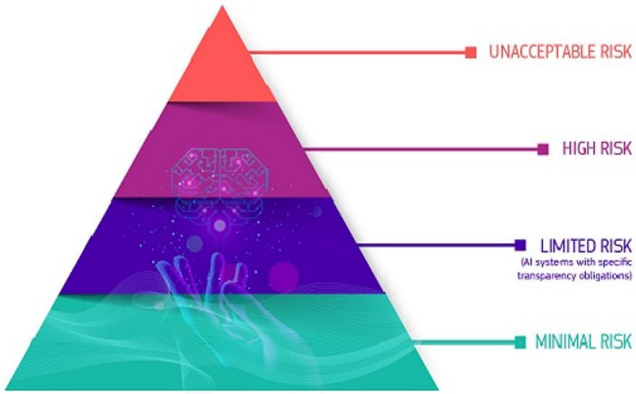
•sistemi di IA utilizzati come componenti di sicurezza in determinate infrastrutture critiche, ad esempio nei settori del traffico stradale e della fornitura di acqua, gas, riscaldamento ed elettricità;
•sistemi di IA utilizzati nel settore dell'istruzione e formazione professionale, ad esempio per valutare i risultati dell'apprendimento, orientare il processo di apprendimento e monitorare i comportamenti disonesti;
•sistemi di IA utilizzati nei settori dell'occupazione, della gestione dei lavoratori e dell'accesso al lavoro autonomo, ad esempio per pubblicare annunci di lavoro mirati, analizzare e filtrare le candidature e valutare i candidati;

•sistemi di IA utilizzati per l'accesso a servizi e a prestazioni pubblici e privati essenziali (ad esempio l'assistenza sanitaria), la valutazione dell'affidabilità creditizia delle persone fisiche e la valutazione dei rischi e la determinazione dei prezzi in relazione ad assicurazioni sulla vita e assicurazioni sanitarie;
•sistemi di IA utilizzati nei settori delle attività di contrasto, della migrazione e del controllo delle frontiere, nella misura in cui sono consentiti, nonché nell'amministrazione della giustizia e nei processi democratici;
•sistemi di IA utilizzati per l'identificazione biometrica, la categorizzazione biometrica e il riconoscimento delle emozioni, nella misura in cui sono consentiti.

# Rischio (3/4)

## rischio limitato:

per questi sistemi sono previsti soltanto requisiti minimi di trasparenza.



# Rischio limitato

## Esempi

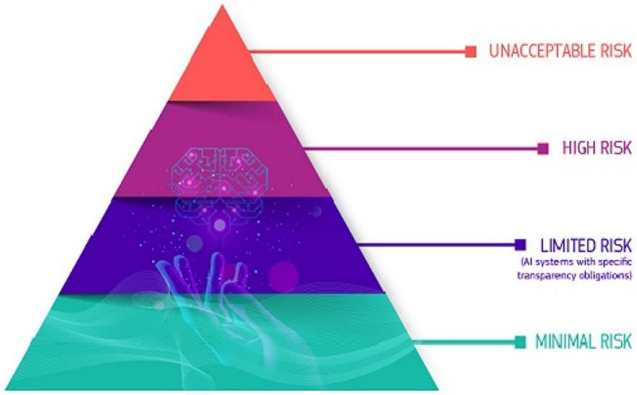
Ad esempio, un'azienda sviluppa un chatbot generico per fornire assistenza clienti online. Il chatbot è programmato per rispondere a domande frequenti riguardanti orari di apertura, disponibilità di prodotti e politiche di reso. Non ha accesso a dati sensibili o personali degli utenti, limitando così le implicazioni etiche e i rischi di privacy. Un altro esempio potrebbe essere un software di elaborazione delle immagini utilizzato per migliorare la qualità visiva delle foto nei cataloghi online.

Questo software ottimizza automaticamente luminosità e contrasto, ma non identifica né memorizza informazioni identificative sulle persone nelle immagini. Entrambi questi sistemi presentano un rischio minore e per questo rientrano nella categoria di sistemi a rischio limitato

# Rischio (4/4)

## rischio minimo o nullo:

è consentito il libero utilizzo (es. applicazioni come videogiochi abilitati all'intelligenza artificiale o filtri antispam).

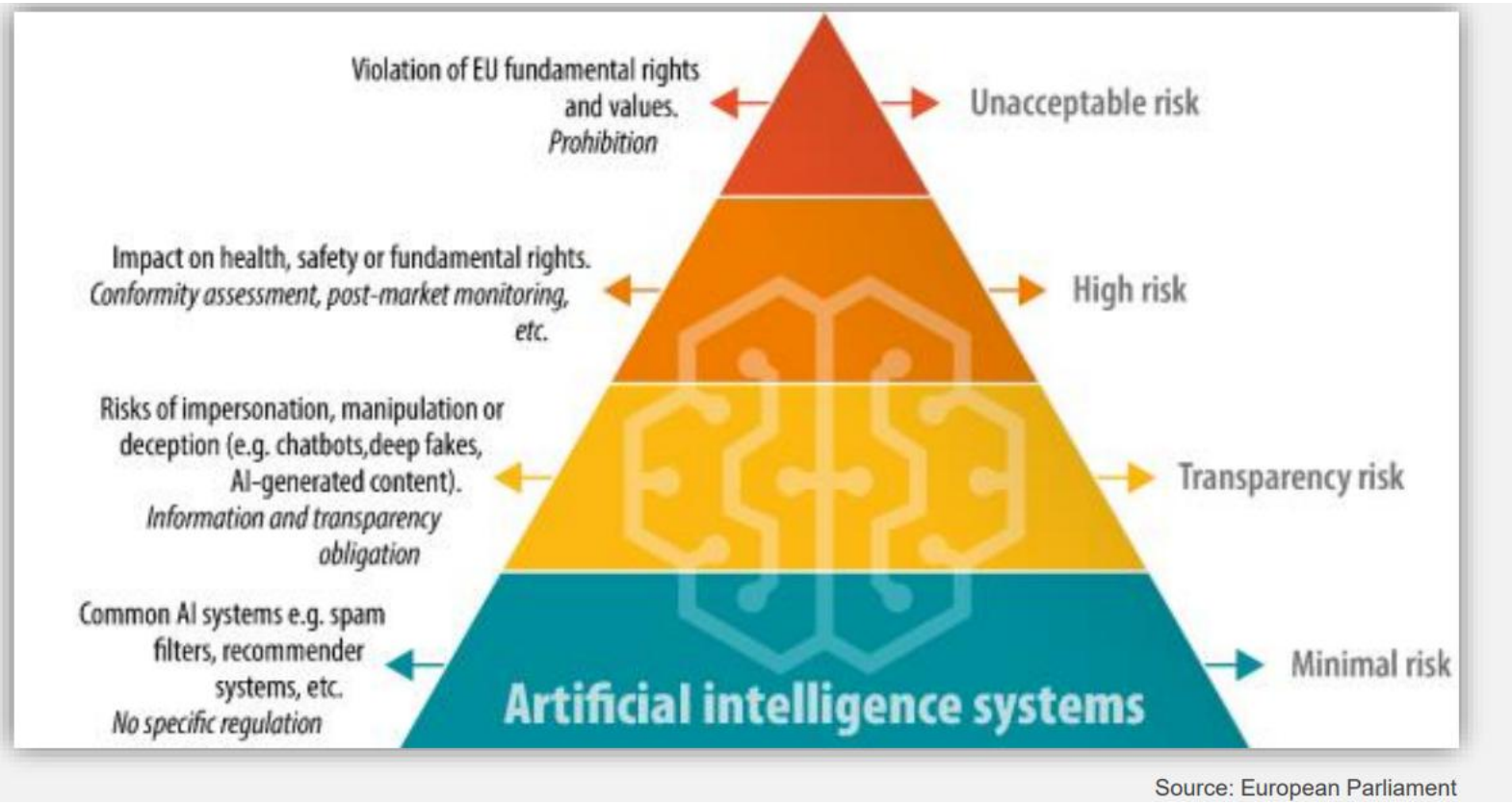


# Rischio minimo o nullo

## Esempi

Questa categoria include sistemi con un impatto minimo o nullo sui diritti o la sicurezza delle persone, come calcolatrici o videogiochi semplici. Un esempio comune è una calcolatrice basata sull'intelligenza artificiale per smartphone. Un altro esempio potrebbe essere un'applicazione per la gestione di liste della spesa. Questa app permette agli utenti di creare e organizzare liste della spesa, impostare promemoria per gli acquisti e condividere le liste con altri utenti. Non richiede né tratta dati personali sensibili, limitandosi a informazioni generiche sui prodotti. Inoltre, l'app non interagisce con altri sistemi critici o infrastrutture, minimizzando qualsiasi tipo di rischio legato alla privacy o alla sicurezza degli utenti.

# Livelli di rischio



## Sistemi ad alto rischio – misure di sicurezza

- Sistema di gestione dei rischi, processo iterativo continuo eseguito nel corso dell'intero ciclo di vita di un sistema di IA ad alto rischio, che richiede un aggiornamento costante e sistematico

- Governance dei dati, ad es. i set di dati di addestramento, convalida e prova sono soggetti ad adeguate pratiche di governance e gestione dei dati

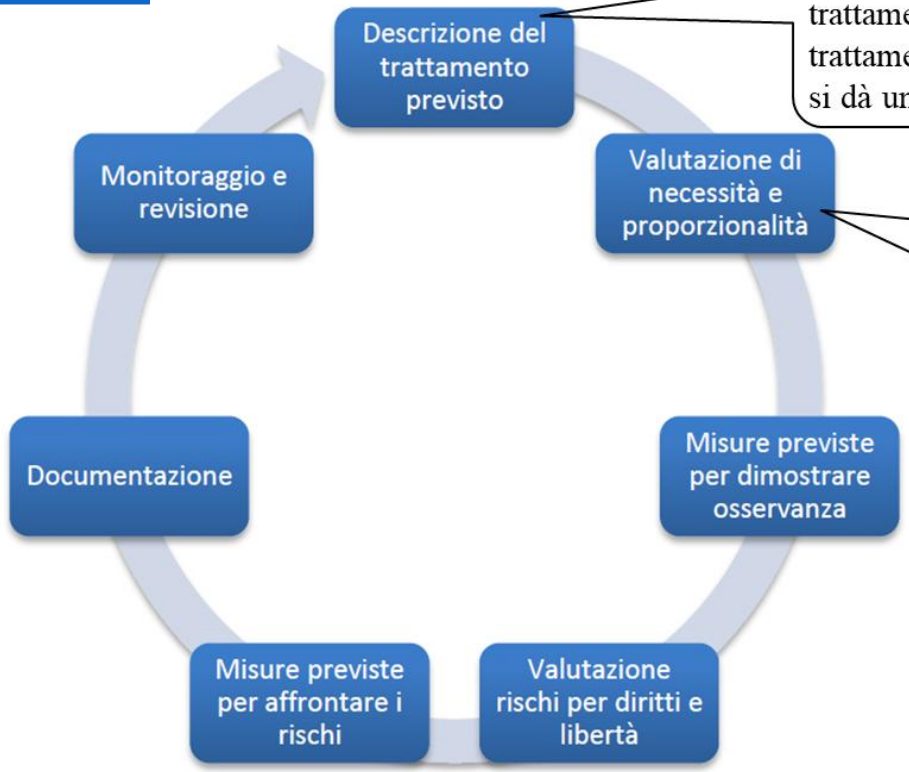
- Mantenere una documentazione tecnica completa con informazioni precise, tra cui specifiche di progettazione del sistema, capacità, limitazioni e impegni di conformità normativa.

## Sistemi ad alto rischio – misure di sicurezza

I sistemi IA ad alto rischio devono essere sviluppati e progettati in modo da garantire:

- Conservazione delle registrazioni, ad es. registrazione automatica degli eventi ("log") durante il funzionamento
- Trasparenza sufficiente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente
- Sorveglianza umana, strumenti di interfaccia uomo-macchina adeguati in modo da poter essere efficacemente supervisionati da persone fisiche
- Adeguato livello di accuratezza, robustezza e cibersicurezza
- L'AI Act definisce requisiti rigorosi per la valutazione d'impatto dei sistemi IA "ad alto rischio". Quest'ultima DEVE includere la valutazione d'impatto sui diritti fondamentali e la valutazione d'impatto sulla protezione dei dati personali (DPIA)(cfr. cap. 10).

# DPIA




si tiene conto della natura, dell'ambito, del contesto e delle finalità del trattamento (considerando 90); sono indicati i dati personali oggetto del trattamento, i destinatari e il periodo previsto di conservazione dei dati stessi; si dà una descrizione funzionale del trattamento;

Valutazione di necessità e proporzionalità del trattamento (art. 35, paragrafo 7, lettera b) ):
- liceità del trattamento (art. 6);
- dati adeguati, pertinenti e limitati a quanto necessario (art. 5(1)c) );
- periodo limitato di conservazione (art. 5(1), lettera e) );

- informazioni fornite agli interessati (artt. 12, 13, 14);
- diritto di accesso e portabilità dei dati (artt. 15 e 20);
- diritto di rettifica e cancellazione (artt. 16, 17, 19); diritto di opposizione e limitazione del trattamento (artt. 18,19, 21);
- rapporti con responsabili del trattamento (art. 28);
- garanzie per i trasferimenti internazionali di dati (Capo V);
- consultazione preventiva (art. 36);

## DPIA: quando è obbligatoria?

Il Gruppo Art. 29 individua nove criteri specifici a questo proposito:
Il Gruppo Art. 29 individua nove criteri specifici a questo proposito:
1) valutazione o assegnazione di un punteggio, compresa la profilazione;
2) processo decisionale automatizzato che produce significativi effetti giuridici (es: assunzioni, concessione di prestiti, stipula di assicurazioni);
3) monitoraggio sistematico (es: videosorveglianza);
4) trattamento di dati sensibili, giudiziari o aventi carattere altamente personale;
5) trattamenti di dati personali su larga scala;
6) combinazione o raffronto di insiemi di dati derivanti da due o più trattamenti svolti per diverse finalità e/o da titolari distinti, secondo modalità che esulano dal consenso iniziale;
7) dati relativi a soggetti vulnerabili (minori, soggetti con patologie psichiatriche, anziani, ecc.) che possono non essere in grado di esercitare diritti, acconsentire o opporsi al trattamento;
8) utilizzi innovativi o applicazione di nuove soluzioni tecnologiche o organizzative (es:riconoscimento facciale, device IoT, ecc.);
9) trattamenti che, di per sé, potrebbero impedire agli interessati di esercitare un diritto o di avvalersi di un servizio o di un contratto4

# Metodologia: ISO 29134

INTERNATIONAL STANDARD

ISO/IEC 29134

First edition
2017-06

Information technology — Security techniques — Guidelines for privacy impact assessment

*Technologies de l'information — Techniques de sécurité — Lignes directrices pour l'évaluation d'impacts sur la vie privée*

**1 –Determinare se la PIA necessaria:** si fa un'attenta analisi del progetto.

**2 –Descrizione dei flussi di informazioni e coinvolgimento dei partecipanti**
individuare con precisione quali dati saranno utilizzati, come e dove serviranno, quali finalità, da chi si ottengono, a chi saranno comunicati, chi ne avrà l'accesso.

**3 –Identificazione dei rischi privacy e di quelli correlati**
individuare i rischi di privacy e valutare il rischio in termini di coefficienti di probabilità e di gravità.

**4 –Identificazione delle soluzioni e delle misure di sicurezza**
individuare le possibili soluzioni per i vari rischi censiti e valutare i costi e i benefici.

**5 –Approvazione delle decisioni e registrazione dei risultati**
sarebbe opportuno verbalizzare i vari passi seguiti nel processo decisionale e chi li ha approvati.

**6 –Integrazione dei risultati del PIA nel piano di progetto**
la DPIA va allegata e integrata al piano di progetto

## Sistemi ad alto rischio – misure di sicurezza

Esistono ulteriori obblighi di trasparenza per tipi specifici di AI. Per esempio:

- I sistemi AI destinati a interagire direttamente con le persone devono essere progettati per informare gli utenti che stanno interagendo con un sistema AI, a meno che ciò non sia ovvio considerando il contesto. Un chatbot, ad esempio, deve essere progettato per avvisare gli utenti di essere un chatbot.

- I sistemi Ai che generano testo, immagini o altri contenuti devono utilizzare formati leggibili a macchina per contrassegnare gli output come generati o manipolati dall'AI. Ciò include, ad esempio, l'AI che genera deepfake, immagini o video alterati per mostrare qualcuno che fa o dice qualcosa che non ha fatto o detto.

# Evidenza degli obblighi

## Sistemi di IA ad alto rischio: valutazione della conformità (ex ante)

- Il fornitore dei sistemi di IA ad alto rischio garantisce che il sistema sia sottoposto alla procedura di valutazione della conformità
- Gli organismi notificati verificano la conformità del sistema
- Gli organismi notificati rilasciano certificati con validità non superiore a 5 anni
- Il fornitore redige una dichiarazione di conformità UE e appone la marcatura CE

## Evidenza degli obblighi

I fornitori di sistemi di intelligenza artificiale ad alto rischio dovranno implementare sistemi di gestione della qualità e del rischio per garantire la conformità ai nuovi requisiti e ridurre al minimo i rischi per gli utenti e le persone interessate, anche dopo l'immissione del prodotto sul mercato;

le autorità pubbliche e gli enti che agiscono per loro conto che utilizzano sistemi di intelligenza artificiale ad alto rischio dovranno registrarli in una banca dati pubblica dell'UE;

le autorità di sorveglianza del mercato, per garantire la conformità durante l'intero ciclo di vita del sistema di IA, devono condurre audit regolari e facilitare il monitoraggio successivo all'immissione sul mercato, consentendo ai fornitori di segnalare volontariamente qualsiasi incidente grave o violazione degli obblighi in materia di diritti fondamentali di cui vengano a conoscenza.

In caso di violazione, i requisiti consentiranno altresì, alle autorità nazionali, di avere accesso alle informazioni necessarie per indagare se l'uso del sistema di IA è conforme alla legge.

## Sanzioni

In caso di mancata conformità alle pratiche proibite AI, le organizzazioni possono essere multate fino a 35.000.000 € o il 7% del fatturato annuo globale, a seconda di quale sia l'importo più elevato.

Per la maggior parte delle altre violazioni, inclusa la mancata conformità ai requisiti per i sistemi di AI ad alto rischio, le organizzazioni possono essere multate fino a 15.000.000 € o al 3% del fatturato annuo globale, a seconda di quale sia il valore più alto.

Fornire informazioni errate, incomplete o fuorvianti alle autorità può comportare multe per le organizzazioni fino a 7.500.000 € o l'1% del fatturato annuo globale, a seconda di quale sia l'importo più elevato.

In particolare, l'EU AI Act prevede regole diverse per multare le startup e altre piccole o medie organizzazioni. Per queste aziende, la multa è il più basso dei due importi possibili specificati sopra.

## Sanzioni

In caso di mancata conformità alle pratiche proibite AI, le organizzazioni possono essere multate fino a 35.000.000 € o il 7% del fatturato annuo globale, a seconda di quale sia l'importo più elevato.

Per la maggior parte delle altre violazioni, inclusa la mancata conformità ai requisiti per i sistemi di AI ad alto rischio, le organizzazioni possono essere multate fino a 15.000.000 € o al 3% del fatturato annuo globale, a seconda di quale sia il valore più alto.

Fornire informazioni errate, incomplete o fuorvianti alle autorità può comportare multe per le organizzazioni fino a 7.500.000 € o l'1% del fatturato annuo globale, a seconda di quale sia l'importo più elevato.

In particolare, l'EU AI Act prevede regole diverse per multare le startup e altre piccole o medie organizzazioni. Per queste aziende, la multa è il più basso dei due importi possibili specificati sopra.

## Indice degli argomenti:

- **Conoscere le normative europee e le best practice per la sicurezza.**
- Comprendere i rischi e le minacce legate all'IA e alla sicurezza dei dati
- Valutare le misure di sicurezza da implementare (Framework NIST e la Iso 42001)

# DDL Intelligenza artificiale - Atto Senato 1146

Senato della Repubblica

| | |
|---|---|
| Documento completo | **Legislatura 19ª - Disegno di legge n. 1146** |
| FRONTESPIZIO | Copia questo link |
| Relazione | |
| Relazione tecnica | |
| Altro | |

## DDL Intelligenza artificiale - Atto Senato 1146

### Articolo 3 punto 5

Al fine di garantire il rispetto dei diritti e dei principi di cui al presente articolo deve essere assicurata, quale precondizione essenziale, la cybersicurezza lungo tutto il ciclo di vita dei sistemi e dei modelli di intelligenza artificiale, secondo un approccio proporzionale e basato sul rischio, nonché l'adozione di specifici controlli di sicurezza, anche al fine di assicurarne la resilienza contro tentativi di alterarne l'utilizzo, il comportamento previsto, le prestazioni o le impostazioni di sicurezza.

### Articolo 4 punto 2

L'utilizzo di sistemi di intelligenza artificiale garantisce il trattamento lecito, corretto e trasparente dei dati personali e la compatibilità con le finalità per le quali sono stati raccolti, in conformità col diritto dell'Unione europea in materia di dati personali e di tutela della riservatezza.

## DDL Intelligenza artificiale - Atto Senato 1146

CAPO V DISPOSIZIONI PENALI ART. 2

Al codice penale sono apportate le seguenti modificazioni:

**all'articolo 61,** al primo comma, dopo il numero 11 novies ), è aggiunto il seguente: «11 decies )

l'avere commesso il fatto mediante l'impiego di **sistemi di intelligenza artificiale**, quando gli stessi, per la loro natura o per le modalità di utilizzo, abbiano costituito **mezzo insidioso**, ovvero quando il loro impiego abbia comunque ostacolato la pubblica o la privata difesa, ovvero aggravato le conseguenze del reato.»;

all'articolo 494, dopo il primo comma è aggiunto il seguente: «La pena è della reclusione da uno a tre anni se il fatto è commesso mediante l'impiego di sistemi di intelligenza artificiale.»;

## DDL Intelligenza artificiale - Atto Senato 1146

«Art. 612 quater

(Illecita diffusione di contenuti generati o manipolati con sistemi di intelligenza artificiale)

Chiunque, al fine di arrecare nocumento a una persona e senza il suo consenso, ne invia, consegna, cede, pubblica o comunque diffonde l'immagine, un video o la voce, falsificati o alterati mediante l'impiego di sistemi di intelligenza artificiale e idonei a indurre in inganno sulla loro genuini è punito con la reclusione da sei mesi a tre anni.

Se dal fatto deriva un danno ingiusto, la pena è della reclusione da uno a cinque anni.

**VERSIONE ALTERNATIVA**

Chiunque cagiona ad altri un danno ingiusto, mediante invio, consegna, cessione, pubblicazione o comunque diffusione di immagini o video di persone o di cose ovvero di voci o suoni in tutto o in parte falsi, generati o alterati mediante l'impiego di sistemi di intelligenza artificiale, atti a indurre in inganno sulla loro genuinità, è punito con la reclusione da uno a cinque anni.

## Indice degli argomenti:

- Conoscere le normative europee e le best practice per la sicurezza.
- Comprendere i rischi e le minacce legate all'IA e alla sicurezza dei dati
- **Valutare le misure di sicurezza da implementare (Framework NIST e la Iso 42001)**

# Artificial Intelligence Risk Management Framework (AI RMF 1.0)

# Artificial Intelligence Risk Management Framework (AI RMF 1.0)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Key Dimensions | Application Context | Data & Input | AI Model | AI Model | Task & Output | Application Context | People & Planet |
| Lifecycle Stage | Plan and Design | Collect and Process Data | Build and Use Model | Verify and Validate | Deploy and Use | Operate and Monitor | Use or Impacted by |
| TEVV | TEVV includes audit & impact assessment | TEVV includes internal & external validation | TEVV includes model testing | TEVV includes model testing | TEVV includes integration, compliance testing & validation | TEVV includes audit & impact assessment | TEVV includes audit & impact assessment |
| Activities | Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations. | Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations. | Create or select algorithms; train models. | Verify & validate, calibrate, and interpret model output. | Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience. | Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations. | Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights. |
| Representative Actors | System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators. | Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts. | Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts. | | System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts, | System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators. | End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers. |

# Artificial Intelligence Risk Management Framework (AI RMF 1.0)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Key Dimensions | Application Context | Data & Input | AI Model | AI Model | Task & Output | Application Context | People & Planet |
| Lifecycle Stage | Plan and Design | Collect and Process Data | Build and Use Model | Verify and Validate | Deploy and Use | Operate and Monitor | Use or Impacted by |
| TEVV | TEVV includes audit & impact assessment | TEVV includes internal & external validation | TEVV includes model testing | TEVV includes model testing | TEVV includes integration, compliance testing & validation | TEVV includes audit & impact assessment | TEVV includes audit & impact assessment |
| Activities | Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations. | Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations. | Create or select algorithms; train models. | Verify & validate, calibrate, and interpret model output. | Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience. | Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations. | Use system/ technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights. |
| Representative Actors | System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/ communities; evaluators. | Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts. | Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts. | | System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts, | System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impact-ed individuals/commu-nities; evaluators. | End users, operators, and practitioners; impacted individu-als/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers. |

# Artificial Intelligence Risk Management Framework (AI RMF 1.0)

# Artificial Intelligence Risk Management Framework (AI RMF 1.0)

| | Baseline Activities to Establish Trustworthy AI | Activities to Enhance AI Trustworthiness |
|---|---|---|
| **Govern** | • Design structures to align AI risk management with organizational principles, policies and strategy<br>• Document accountability structures for AI systems and related processes | • Establish mechanisms for the team(s) that develop and deploy AI systems to receive and incorporate feedback from AI stakeholders into system updates |
| **Map** | • Determine and document organizational risk tolerances<br>• Assess AI capabilities, targeted usage and goals | • Map risks across components of AI systems and the data supply chain<br>• Create processes for refreshing risk perspectives and understanding changes in potential drift in the model over time |
| **Measure** | • Define metrics for measuring AI risk and control effectiveness<br>• Establish monitoring of AI systems and components in production | • Monitor AI tools to identify and track existing and emerging AI risks<br>• Create reporting and feedback mechanisms to measure AI trustworthy characteristics |
| **Manage** | • Document and prioritize AI risks<br>• Align risk strategies for prioritized AI risks with broader organizational strategy | • Enact robust, tech-enabled incident and issue management and communication processes<br>• Define evaluations and monitoring third-party AI resources |

# ISO/IEC 42001:2023

## Clause 4: Context of the Organization

The organization must identify the purpose of its AI system, internal and external requirements, stakeholders, business objectives, and other dependencies to set the scope of the AIMS.

## Clause 5: Leadership

The organization must secure the support of top management in all AIMS-related activities, from AI policy creation to ongoing monitoring and periodic reviews.

## Clause 6: Planning

The organization must determine the risks, opportunities, and impacts associated with its AI system, plan actions for addressing them, and set the objectives for the AIMS.

## Clause 7: Support

The organization must support the effective implementation of the AIMS through adequate resource allocation, documentation and communication of information, and training and awareness.

# ISO/IEC 42001:2023

## Clause 8: Operation

The organization must implement processes and controls for the development, operation, and maintenance of the AIMS.

## Clause 9: Performance Evaluation

The organization must monitor, measure, analyze, and evaluate the performance of the AIMS through internal audits and management review.

## Clause 10: Improvement

The organization must correct nonconformities and put in place processes for the continuous improvement of the AIMS.

# ISO/IEC 42001:2023

## Data privacy

The ability of the AI system to access, process, or disclose sensitive or confidential information

## Algorithmic transparency

The traceability and explicability of the logic and processes behind the decision-making capabilities of the AI system

## Bias and manipulation

The quality and integrity of the training data used for the AI system to ensure accurate and fair outcomes

# ISO/IEC 42001:2023 Annex

| | |
|---|---|
| **Control A.2:** Policies Related to AI | The establishment and documentation of an AI policy that aligns with business objectives and responsible AI principles to guide the use, development, or deployment of the AI system<br>Integration of the AI policy into existing organizational processes<br>Maintaining the relevance and effectiveness of the AI policy through regular reviews by the organization's top management |
| **Control A.3:** Internal Organization | Definition and allocation of roles and responsibilities throughout the development, deployment, operation, and maintenance of the AIMS to ensure traceability and foster a culture of accountability across the organization<br>Establishment of reporting mechanisms for AI concerns |
| **Control A.4:** Resources for AI Systems | Identification and documentation of resources critical to the AI system, such as data resources, tooling resources, AI system components, computing resources, and human resources at each stage of the AI system life cycle |
| **Control A.5:** Assessing Impacts of AI Systems | Identification, evaluation, and management of the impacts including potential benefits of the AI system on individuals, communities, and societies throughout its life cycle<br>Mitigation of potential harms<br>Documentation of the intended use of the AI system and the process and results for risk and impact assessments |

# ISO/IEC 42001:2023 Annex

| | |
|---|---|
| **Control A.6:** AI System Life Cycle | Documentation of the organization's objectives for the development of the AI system<br>Documentation of design and development processes for the AI system<br>Identification and documentation of AI system requirements and specifications<br>Documentation of the AI system's architecture, interface, and other design and development components<br>Maintaining the integrity of the AI system through verification and validation processes<br>Documentation of a deployment plan for the AI system<br>Documentation of the performance of the AI system throughout its life cycle to facilitate management and continuous improvement: from development, to deployment, operation, and monitoring<br>Technical documentation of the AI system including information on functionality, usage, and limitations<br>Recording of event logs throughout the AI system's life cycle |
| **Control A.7:** Data for AI Systems | Definition and documentation of requirements for ensuring the quality of data used in the AI system to maintain reliable and fair outputs and prevent errors and biases<br>Documentation of the origin, transformation, and usage of data throughout the AI system's life cycle<br>Documentation of data acquisition, selection, and preparation methods to ensure suitability for use in the AI system |

# ISO/IEC 42001:2023 Annex

| | |
|---|---|
| **Control A.8:** Information for Interested Parties of AI Systems | Provision of documentation and essential information about the AI system by the organization to users and other interested parties, including its purpose, instructions for use, and technical limitations<br>Establishment of mechanisms for external reporting of AI-related issues by users and other interested parties<br>Timely communication of AI-related incidents to internal and external stakeholders to build trust and reinforce the organization's commitment to ethical AI implementation<br>Identification and documentation of the organization's obligations for communicating information about the AI system to users and other interested parties |
| **Control A.9:** Use of AI Systems | Definition and documentation of processes to guide other organizations in the responsible use of the AI system<br>Documenting and monitoring the use of the AI system to ensure it is operated and deployed according to its intended purpose<br>Incorporating human oversight into the development and operation of the AI system |
| **Control A.10:** Third-Party & Customer Relationships | Clear allocation of responsibilities between the organization and its partners, suppliers, customers, and other third parties regarding the use of the AI system<br>Establishing assessment processes for suppliers to ensure that the AI system your organization will use or provide meets ethical standards and compliance requirements<br>Aligning the development of the AI system with customer needs and expectations |