Regione Toscana

Direzione Sistemi informativi, infrastrutture tecnologiche e innovazione

Indicazioni operative per l'adozione di soluzioni di Intelligenza Artificiale in riferimento all'AI Act in Toscana

Regione Toscana - Giunta Regionale

Aprile 2025

Data versione	Modifiche apportate
16/04/2025	Prima emissione per consultazione su partecipa.toscana.it

Indice generale

1 Introduzione

Regione Toscana è in linea con la sensibilità europea che, nel corso degli anni, ha sviluppato un percorso per un'IA responsabile, volta al benessere del genere umano e dell'ambiente, nel rispetto dei valori e dei diritti dei cittadini europei[1]. Questo percorso ha portato all'approvazione dell'AI Act[2] , il regolamento europeo sull'Intelligenza Artificiale (IA).

Con questo documento Regione Toscana intende quindi recepire e richiamare quanto prodotto da livello europeo e nazionale e fornire – sulla base del contesto di norme e linee guida attuali in materia - indicazioni pratiche su come accompagnare il recepimento dell'AI Act, che diventerà operativo pienamente nel corso dei prossimi anni. Questo documento implementa quanto previsto all'art.8 comma 2 della Legge Regionale Toscana 57 – 2024 "Disciplina dell'innovazione digitale nel territorio regionale e tutela dei diritti di cittadinanza digitale. Modifiche alla l.r. 54/2009."

In particolare il presente documento intende fornire indicazioni di supporto per l'adozione dell'IA e mette a disposizione alcuni modelli di base relativi alla documentazione tecnica e alla valutazione dell'impatto (Impact Assessment). L'obiettivo è accompagnare l'adozione delle tecnologie di IA nei settori chiave, garantendo trasparenza, accessibilità e partecipazione, oltre ad aiutare la Pubblica Amministrazione toscana a implementare l'IA in modo consapevole, accorto e il più sicuro possibile, in linea con la normativa vigente.

I destinatari del presente documento sono coloro i quali, all'interno della PA Toscana, siano chiamati ad acquisire sviluppare, implementare un sistema di IA.

Il documento rappresenta in particolare uno strumento di supporto per la Giunta Regionale, il Consiglio Regionale, il sistema sanitario e gli enti locali della Toscana.

---

1 *Ethics Guidelines for Trustworthy AI (2019), The Assessment List for Trustworthy Artificial Intelligence (2020), White Paper on Artificial Intelligence (2020), The Ethics of Artificial Intelligence: Issues and Initiatives (2020), Robustness and Explainability of Artificial Intelligence (2020), Review of the Coordinated Plan on Artificial Intelligence (2021).*

2 Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale) .

2 Principi, orientamenti e obiettivi

Le indicazioni fornite da Regione Toscana si basano sull'AI Act dell'Unione Europea, sulla Strategia Nazionale Italiana per l'Intelligenza Artificiale 2024-2026 e su principi etici ampiamente condivisi a livello internazionale.

Regione Toscana intende favorire lo sviluppo e l'utilizzo responsabile dell'IA in Toscana, con particolare attenzione a:

- **Un approccio antropocentrico nell'utilizzo dei sistemi di IA**, basato sul principio *human in the loop* e sull'art. 22 del GDPR, secondo cui il soggetto non deve mai essere sottoposto ad una decisione automatizzata, in grado di incidere sulla sua persona, senza che vi sia stato un preventivo controllo da parte dell'uomo stesso. Conseguentemente, nel rapporto tra uomo e macchina, il primo deve avere la possibilità di controllare e indirizzare sempre l'attività della seconda senza mai subirne passivamente i risultati[3].
- **Promuovere la ricerca e l'innovazione in IA**, privilegiando l'approccio Open e lo sviluppo di software a codice sorgente aperto (Open Source).
- **Sviluppare** competenze e formazione in IA per tutti gli attori coinvolti.
- **Garantire** un uso etico e responsabile dell'IA, basato su principi di proporzionalità e divieto di nuocere, diritto alla privacy, governance, accountability, responsabilità, equità e non discriminazione, trasparenza, spiegabilità, sicurezza e protezione del dato, supervisione e determinazione umana, sostenibilità, consapevolezza e alfabetizzazione.
- **Promuovere** la collaborazione e la partecipazione di tutti gli stakeholder nello sviluppo e nella governance dell'IA.

Regione Toscana è consapevole che, per agevolare il cambiamento in atto in conseguenza delle enormi potenzialità dei sistemi di AI, è necessario realizzare azioni di alfabetizzazione e formazione rivolte a tutti gli *stakeholders* coinvolti, in particolare al mondo dell'educazione, alla cittadinanza, alle Pubbliche Amministrazioni e alle imprese, al fine di rimuovere e prevenire gli ostacoli che impediscono la piena parità di accesso alle informazioni e alle tecnologie dell'informazione e della comunicazione, di ridurre il divario digitale esistente tra coloro che conoscono e abitualmente utilizzano efficacemente gli strumenti informatici (oggi arricchita dalla IA) e coloro che, per le ragioni più diverse, ne risultano esclusi o limitati. Favorendo una massimizzazione delle competenze digitali nella popolazione si persegue dunque un miglioramento della

[3] In particolare l'art. 14 (Human oversight) Regolamento (UE) 2024/1689.

qualità della vita dei cittadini nel rapporto con le pubbliche amministrazioni e gli enti del territorio toscano, favorendo anche forme di cittadinanza attiva.

Tra le iniziative avviate a livello regionale si segnalano i c.d. “centri di facilitazione digitale” attivi in maniera capillare sul territorio toscano (deliberazione di Giunta n. 295 del 20 marzo 2023 "Approvazione criteri dell’avviso per gli enti locali del territorio per l’attivazione di centri di facilitazione digitale previsti dalla misura 1.7.2 Missione 1 Componente 1 PNRR e assegnazione dei relativi finanziamenti a Sviluppo Toscana come organismo intermedio”), ovvero forme di coordinamento stabili e strutturate, mirate a supportare lo sviluppo di competenze di base della cittadinanza, e a contribuire all’inclusione digitale della popolazione non adeguatamente formata all’utilizzo dei servizi online e in particolare della Pubblica Amministrazione, individuando le esigenze dei singoli cittadini e fornendo loro aiuto e orientamento e a fruire efficacemente dei servizi pubblici offerti in modalità telematica dalla Pubblica Amministrazione e dagli enti eroganti servizi pubblici[4].

In riferimento alla recente pubblicazione del *Committee of the Regions* europeo sul livello di adozione della *AI in Local and Regional Authorities* del 07.01.2025 (“***AI and GenAI adoption by local and regional administrations”***)[5], si ritiene rilevante evidenziare anche per la Toscana le seguenti caratteristiche di un’efficace adozione di AI da parte di Comuni e Regioni:

- Considerata la carenza di personale qualificato in materia nella PA, risulta strategico incentivare forme di partnership pubblico-privato e collaborazioni con le Università. In questo senso, acquisire il know-how di fornitori esterni mediante procurement sul mercato, e promuovere iniziative di diffusione di competenze sul tema AI mediante un dialogo continuo che coinvolga PA e privati risulta strategico[6].
- L’obiettivo è far sì che i cittadini possano comprendere il funzionamento dei sistemi di IA, i loro limiti e vantaggi. Ciò aiuterebbe sia a promuovere la trasparenza degli stessi (Explainable AI) sia nel fidarsi dei loro risultati. Una possibile indicazione che si potrebbe adottare è dare modo – dentro i servizi basati su AI – di raccogliere in modo facile il feedback dell’utente in modo da permettere la segnalazione di allucinazioni o comportamenti errati.

2.1 Contesto europeo, nazionale e regionale di riferimento

4 https://competenzedigitali.toscana.it/

5 https://op.europa.eu/en/publication-detail/-/publication/40363d58-bdc8-11ef-91ed-01aa75ed71a1/language-en

6 Esempi ne sono in Toscana: a) il *Centro di Competenza Big Data e AI* (CBDAI) al quale aderiscono tutte le Università pubbliche toscane eche ha collaborato anche alla stesura del presente documento; b) collaborazioni con privati per l’organizzazione di iniziative (come gli *hackathon*); c) eventi di disseminazione sulla AI in Toscana; d) l’indizione di gare da parte di Regione Toscana quale soggetto aggregatore per accompagnare il percorso di implementazione portando ulteriore know-how dalle imprese.

Le presenti linee guida si inseriscono in un contesto di "*actification*" a livello nazionale ed europeo in tema di digitalizzazione, utilizzo dell'IA e garanzia della sicurezza cibernetica.

In particolare l'AI Act è il primo regolamento organico sulla Intelligenza Artificiale, approvato il 21 maggio 2024 dal Consiglio dell'UE e mira a garantire sistemi di IA sicuri e rispettosi dei diritti fondamentali nel mercato europeo, favorendo un'innovazione sicura in tale ambito attraverso regole armonizzate tra gli Stati membri dell'Unione europea per lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di sistemi di IA nell'Unione.

Come spiega l'Art. 3 AI Act per Sistema di intelligenza artificiale si intende un sistema basato su macchina progettato per funzionare con diversi livelli di autonomia e che può mostrare adattabilità dopo il funzionamento, e che, per obiettivi espressi o impliciti, deduce, dagli input che riceve, come produrre output quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare le contesti fisici o virtuali.

L'AI Act è fondato su un approccio "*risk-based*" al pari del GDPR: all'aumentare del rischio insito nell'utilizzo di un determinato sistema intelligenza artificiale, aumenteranno le regole e quindi le forme di responsabilità di chi sviluppa e usa l'IA, con la previsione anche di divieti di utilizzo per quei sistemi che manifestano un rischio inaccettabile. In ambito nazionale, invece, il DDL n. 1146/2024[7], in materia di intelligenza artificiale e la "Bozza di Linee Guida per l'adozione della IA nella Pubblica Amministrazione"[8], elaborata da AGID, rappresentano un passo cruciale verso una regolamentazione consapevole e responsabile delle tecnologie che si dotano di IA nel settore pubblico. Una volta tradotti in testi normativi definitivi, questi strumenti potranno favorire l'adozione etica e trasparente dell'IA, garantendo sicurezza, tutela dei diritti fondamentali e affidabilità dei sistemi.

A livello territoriale, infine, la legge regionale n. 57/2024[9] tramite la previsione di articoli ad hoc in materia di intelligenza artificiale, mira a garantire un adeguamento al contesto normativo in continua evoluzione nell'ambito della digitalizzazione, dell'IA e della cybersecurity per favorire lo sviluppo di una pubblica amministrazione più accessibile ed efficiente. In quest'ottica Regione Toscana, nel perimetro di quanto

---

7 Il DDL n. 1146/2024 accompagna il quadro regolatorio delineato dall'AI Act in quegli spazi propri del diritto interno, tenendo debitamente conto dell'approccio risk based del regolamento, per cui maggiore è il rischio e maggiori sono le responsabilità e i divieti per chi sviluppa o utilizza sistemi di IA. Obiettivo di tale atto normativo è quello di individuare criteri regolatori in ambito IA, prevedendo principi e disposizioni che siano in grado di promuovere l'utilizzo di nuove tecnologie e di fornire soluzioni per una corretta gestione dei rischi; tutto ciò nel pieno rispetto dei diritti fondamentali di ogni individuo.

8 Il documento (https://www.agid.gov.it/it/ambiti-intervento/intelligenza-artificiale) mira a definire principi, criteri e raccomandazioni operative per accompagnare le amministrazioni nell'adozione dell'IA, garantendo il rispetto dei diritti fondamentali, la qualità dei dati e la (cyber) sicurezza dei sistemi. L'obiettivo è favorire l'innovazione tecnologica al fine di migliorare l'efficienza e l'efficacia dell'azione amministrativa, nel rispetto dei valori e principi fondamentali del nostro ordinamento.

9 https://raccoltanormativa.consiglio.regione.toscana.it/articolo?urndoc=urn:nir:regione.toscana:legge:2024-12-09;57&pr=idx,0;artic,1;articparziale,0

previsto dagli artt. 6, 8 e 25 della presente legge, ha avviato un processo di redazione di "Linee Guida Generali per l'utilizzo dei sistemi di IA in Regione Toscana"[10].

In attesa della completa entrata in vigore dell'AI Act e dell'eventuale adozione della normativa nazionale di riferimento, le indicazioni del presente documento sono finalizzate ad orientare i "*deployers*" ad un uso consapevole e responsabile delle soluzioni di intelligenza artificiale adottate dall'Ente e conformi all'AI Act.

3 Indicazioni sul recepimento dell'AI ACT[11]

Il 12 luglio del 2024 l'EU AI Act è stato pubblicato nel Gazzetta Ufficiale dell'UE, ed è entrato in vigore il 1° agosto 2024. Tuttavia, l'applicazione delle sue disposizioni sarà graduale e prevede in sintesi[12]:

- il **2 febbraio 2025** l'applicazione dei Capitoli I (disposizioni generali) e Capo II (sistemi di IA proibiti);
- il **2 agosto 2025** il Capo V ("*General Purpose AI models*");
- il **2 agosto 2026** tutto l'AI Act eccetto l'art. 6 – Comma 1;
- il **2 agosto 2027** saranno incluse anche tutte le norme relative ai sistemi di IA ad Alto Rischio (Art. 6 comma 1 dell'AI Act).

In questo periodo di transizione, la Regione Toscana intende fornire indicazioni concrete per supportare l'efficace recepimento della norma nelle diverse fasi della sua attuazione. Inoltre, si propone di affiancare la normativa con ulteriori indicazioni riguardanti i sistemi di IA che saranno utilizzati. L'obiettivo principale è tutelare la cittadinanza e supportare i vari *stakeholders*, garantendo la massima trasparenza, spiegabilità, affidabilità e robustezza nell'uso di questa tecnologia.

Le indicazioni si basano sull'approccio orientato al rischio, come indicato nell'AI Act, e includono il supporto alla valutazione dei sistemi di IA, la fornitura di modelli per l'*AI Impact Assessment* e la redazione di documentazione descrittiva dei sistemi di IA.

In ogni caso, per ogni sistema di IA, in quanto software, occorre valutare la sua compatibilità con la normativa vincolante in vigore quale il Regolamento UE 2016/679 (GDPR), Direttiva (UE) 2022/2555 (NIS II), d.lgs n. 138/2024 di recepimento della direttiva (UE) 2022/2555, Legge n. 90/2024 "Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici", il d.lgs. n. 82/2005 e

[10] IN CORSO DI PUBBLICAZIONE - Linee Guida Generali per l'utilizzo dei sistemi di IA in Regione Toscana

[11] Per approfondimenti vedi Allegato I – Approfondimenti: Intelligenza Artificiale e AI Act

[12] Al seguente riferimento è possibile trovare la tempistica dettagliata dell'entrata in vigore dell'AI Act: https://artificialintelligenceact.eu/implementation-timeline/

successive modifiche (CAD), legge regione Toscana n. 57/2024, delibera Giunta Regione Toscana n. 521 del 23 aprile 2019, decreto dirigenziale n. 7677 del 17 maggio 2019 e successive modifiche e integrazioni, delibera Giunta Regione Toscana n. 810 del 2 agosto 2021 che ha approvato l'allegato A "Data protection policy, Addendum alle linee guida", decreto dirigenziale n. 14917 del 10 agosto 2021 Indicazioni tecnico-operative, delibera Giunta Regione Toscana n. 145 del 20 febbraio 2023 che ha approvato il modello di IT Governance dei sistemi e servizi IT (Allegato A di cui sopra).

A titolo esemplificativo per ogni sistema di IA occorre valutare la necessità di redigere una DPIA (Data Protection Impact Assessment), secondo quanto indicato nell'art. 35 GDPR. Le modalità di redazione della DPIA non sono oggetto del presente documento.

In allegato F "Esempi di applicazione delle checklist indicate" del presente documento sono riportati alcuni casi di studio esemplificativi, come supporto all'utilizzo del presente documento.

3.1 Indicazioni operative

Il presente paragrafo fornisce alcune indicazioni operative per l'adozione di sistemi di soluzioni di IA. Prima di tutto è opportuno individuare il proprio ruolo riguardo il sistema di IA in esame. I ruoli possibili, previsti dallo AI Act sono i seguenti:

- **Fornitore** (Provider). Una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un "modello di IA per scopi generali" (o che fa sviluppare un sistema di IA o un modello di IA per scopi generali) e li immette sul mercato o mette in servizio il sistema con il proprio nome o marchio, a titolo oneroso o gratuito (Art. III, AI Act).
- **Utilizzatore** (Deployer). Qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale (Art. III, AI Act).
- **Distributore**. Una persona fisica o giuridica nella catena di approvvigionamento, diversa dal fornitore o dall'importatore, che mette a disposizione un sistema di IA sul mercato dell'Unione (Art. III, AI Act).
- **Importatore**. Una persona fisica o giuridica ubicata o stabilita nell'Unione che immette sul mercato un sistema di IA recante il nome o il marchio di una persona fisica o giuridica stabilita in un paese terzo (Art. III, AI Act).

Per ciascun sistema di IA preso in considerazione, ne va poi valutato il livello di rischio, classificandolo in una delle seguenti categorie:

- **Proibito.** Che include sistemi con rischi inaccettabili per i diritti fondamentali o la sicurezza (Capo II, AI Act).
- **Alto Rischio.** Comprende sistemi operanti in settori critici o con impatti significativi sulla sicurezza o sui diritti degli utenti (Capo III, AI Act).
- **Determinati sistemi di IA:** Si tratta di specifici sistemi che, pur non essendo classificati ad alto rischio, sono soggetti ad obblighi di trasparenza, ad esempio "*chat bot"* o strumenti di manipolazione di contenuti come "*deepfake"* (Capo IV, AI Act).
- **GPAI (General Purpose AI)[13] models con rischio sistemico:** sono modelli di intelligenza artificiale con finalità generali caratterizzati da un'elevata complessità tecnica e sono quindi soggetti a requisiti rafforzati (Capo V, AI Act).
- **GPAI models senza rischio sistemico: s**ono modelli di intelligenza artificiale con finalità generali che non presentano caratteristiche tali da comportare un rischio sistemico, e che sono soggetti a obblighi più leggeri (Capo V, AI Act).
- **Sistemi a rischio minimo:** sono sistemi che non rientrano nelle categorie sopra indicate e per i quali l'AI Act non prevede obblighi specifici.

Per i dettagli del **livello di rischio di un sistema di IA** come definito nell'AI Act è possibile fare riferimento all'**allegato A "Individuazione del livello di rischio del sistema di IA"** del presente documento

Di seguito vengono illustrate le azioni suggerite in base al ruolo ricoperto e al livello di rischio dei sistemi di IA considerati, con particolare riferimento all'"*Impact Assessment"* (Valutazione d'Impatto) e alla documentazione da redigere.

Per i sistemi proibiti non sono fornite indicazioni, in quanto tali sistemi sono vietati dall'AI Act.

Uno strumento utile per il supporto alla valutazione dell'impatto dell'AI Act è fornito dalla Future of Life Institute[14] che mette a disposizione un sito per il "**Controllo della conformità all'AI Act**"[15]. Per alcuni spunti di riflessione sull'IA e sui sistemi di IA, in riferimento AI ACT, è possibile consultare l'**Allegato H "Domande e FAQ"** del presente documento.

3.1.1. Valutazione dell'Impact Assessment e documentazione dei sistemi di IA

L'"*Impact Assessment"* per i sistemi di IA è una procedura volta ad analizzare i rischi potenziali associati all'uso di un sistema di IA. Nel contesto dell'AI Act, questa valutazione si concentra principalmente sui rischi dei sistemi di IA classificati come ad **Alto Rischio**, ma risulta rilevante anche per altri sistemi,

[13] Modelli di IA per finalità generali.
[14] https://futureoflife.org/.
[15] https://artificialintelligenceact.eu/it/assessment/controllo-della-conformita-dell-ue-alla-legge-ai/.

specialmente quelli che possono avere un impatto significativo sulle persone o sulla società.***a) Valutazione dell'Impatto per "Sistemi di AI proibiti"***

Le norme dell'AI Act in riferimento ai sistemi di IA proibiti (Art. 5, AI Act) sono entrate in vigore il **2 febbraio 2025**.

In quanto sistemi proibiti non sono fornite indicazioni né riguardo l'Impact Assessment né in riferimento alla documentazione. Se ne suggerisce comunque un'analisi delle caratteristiche al fine di essere certi di non ricadere in quella tipologia di sistemi di IA[16].

b) Valutazione dell'Impatto per Sistemi di IA di tipo ad Alto Rischio

Le norme dell'AI Act in riferimento ai sistemi di IA ad Alto Rischio (Art. 6 – Comma 1, AI Act) entreranno in vigore dal **2 agosto 2027**.

Entro il **2 febbraio 2026** la Commissione dovrà fornire le linee guida che specifichino l'attuazione pratica dell'articolo 6.

Per quanto concerne l'"*Impact Assessment"* si forniscono indicazioni che si ispirano a quelle riportate dalla normativa dell'AI Act riguardo il *"**Fundamental Rights Impact Assessment"*** (FRIA) per i sistemi di IA ad Alto Rischio (Art. 27, AI Act).

Nell'**allegato B** del presente documento è riportato il **modello di "*Impact Assessment"*** suggerito per i sistemi ad Alto Rischio.

Per la **documentazione** si suggerisce di seguire le indicazioni riportate nell'AI Act (Art.11, AI Act).

Di seguito sono riportate le indicazioni previste nell'AI Act per i sistemi di IA ad alto rischio (Capo III – Sezione 2, AI Act):

- la conformità ai requisiti (Art. 8, AI Act);
- deve essere istituito, attuato e documentato un **Sistema di Gestione dei Rischi** (Art. 9, AI Act) e sono date indicazioni specifiche riguardo ai dati e alla loro governance (Art. 10, AI Act);
- deve essere disponibile una documentazione tecnica che deve essere redatta prima dell'immissione sul mercato o della messa in servizio di tale sistema e deve essere tenuta aggiornata (Art. 11, AI Act) [17];

[16] Per approfondimenti v. https://ec.europa.eu/newsroom/dae/redirection/document/112367.
[17] Per le PMI e le start up è previsto di fornire una documentazione semplificata (Art. 11 comma 1).

- deve essere eseguita la registrazione automatica degli eventi («log») per la durata del ciclo di vita del sistema secondo le indicazioni riportate nell'Art. 12 dell'AI Act;
- devono essere fornite **indicazioni di trasparenza** (Art. 13, AI Act). E' inoltre richiesta la sorveglianza umana secondo specifiche modalità (Art. 14, AI Act);
- devono essere fornite indicazioni riguardo l'accuratezza, robustezza e cybersicurezza (Art. 15, AI Act);
- per quanto attiene agli ulteriori obblighi e comportamenti previsti per i singoli soggetti, si rimanda a quanto previsto dalle disposizioni dell'AI Act e in particolare:
    - per i ***providers*** agli artt. 16-22 dell'AI Act;
    - per gli **importatori** l'art. 23 dell'AI Act;
    - per i **distributori** l'art. 24 dell'AI Act;
    - per i ***deployers*** l'art. 26; dell'AI Act.

In aggiunta l'art. 25 dell'AI Act definisce alcuni casi in cui **importatori**, **distributori** o ***deployers*** sono considerati ***providers***.

c) Valutazione dell'Impatto per sistemi di IA di tipo "Determinati sistemi di IA

Le norme riferite a tali sistemi previste nell'AI Act entreranno in vigore dal **2 agosto 2026.** In attesa dell'entrata in vigore della normativa Regione Toscana suggerisce, di fornire le indicazioni richieste dall'AI Act e di redigere i modelli generali dell'*Impact Assessment* (Allegato C "Modello di *Impact Assessment* generale") e della documentazione (Allegato D "Modello di documentazione generale").

Riguardo ai determinati sistemi di IA, nell'AI Act per i ***providers*** e per i ***deployers*** si richiedono sostanzialmente solo **obblighi di trasparenza** riportati nel Capo IV, art. 50 dell'AI Act.

d) Valutazione dell'impatto per Sistemi di IA di tipo "GPAI models senza rischio sistemico"

Le norme riferite a tali sistemi previste nell'AI Act entreranno in vigore dal **2 agosto 2025**. Per quanto riguarda i GPAI *models* nell'AI Act sono dati obblighi solo per i ***providers*** (Art. 53, AI Act). Indicazioni specifiche sono fornite per i rappresentanti autorizzati dei fornitori di modelli di IA per finalità generali (Art. 54, AI Act).

In attesa dell'entrata in vigore della normativa Regione Toscana suggerisce per i ***providers*** di redigere il **modello generale per la valutazione dell'*Impact Assessment*** (Allegato C "Modello di *Impact Assessment* generale") e il **modello di documentazione specifico per i GPAI *models*** (Allegato E "Modello di documentazione per i GPAI *Models*).

In base alle indicazioni dell'art. 53, comma 3, dell'AI Act, la documentazione non è richiesta per i provider di GPAI models non a rischio sistemico che sono rilasciati in licenza ***free and open-source***.

e) Valutazione dell'impatto per Sistemi di IA di tipo "GPAI models con rischio sistemico"

Le norme riferite a tali sistemi previste nell'AI Act entreranno in vigore dal **2 agosto 2025**.

Per quanto riguarda i GPAI Models nell'AI Act sono dati obblighi solo per i ***providers*** (Art. 55, AI Act). E' necessario adempiere a quanto indicato per i GPAI *models* senza rischio sistemico (Artt. 53 e 54, AI Act), ai quali si aggiunge l'art. 55 dell'AI Act.

In attesa dell'entrata in vigore della normativa, Regione Toscana suggerisce per i ***providers*** di redigere il **modello generale dell'*Impact Assessment*** (Allegato C "Modello di *Impact Assessment* generale") e il **modello di documentazione specifico per i GPAI models** (Allegato E "Modello di documentazione peri GPAI models").

f) Valutazione dell'impatto per Sistemi di IA di tipo "Rischio minimo"

Per questi sistemi Regione Toscana suggerisce ai ***providers***, di fornire le indicazioni riguardo alla **trasparenza** richieste dall'AI Act per i determinati sistemi di IA (Art. 50, AI Act) e di redigere i **modelli generali dell'*Impact Assessment*** (Allegato C "Modello di *Impact Assessment* generale") e della documentazione (Allegato D "Modello di documentazione generale").

In conclusione, con le presenti indicazioni si intende fornire orientamenti pratici per adottare soluzioni di IA in Regione Toscana in linea con l'AI Act. Assieme alle indicazioni, negli Allegati al presente documento vengono proposti alcuni modelli di documentazione tecnica e di *Impact Assessment* utili, ai quali è possibile ricorrere come esempi di applicazione pratica.

Allegato A – Individuazione del livello di rischio del sistema di IA

Il presente allegato intende fornire un valido supporto per valutare il livello di rischio del sistema IA, che è oggetto di interesse e aiutare il *deployer* a classificarlo in una delle seguenti categorie:

- **Alto rischio** (Capo III, AI Act);
- **Determinati sistemi di IA** (Capo IV, AI Act);
- **GPAI** (General Purpose AI)[18] models con rischio sistemico (Capo V, AI Act);
- **GPAI** models senza rischio sistemico (Capo V, AI Act);
- **Sistemi a rischio minimo**

Individuazione del livello di rischio per sistemi di IA ad Alto Rischio

L'art. 6 dell'Act definisce i sistemi ad Alto Rischio come quei sistemi utilizzati "come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato I[19]", inoltre "sono considerati ad alto rischio anche i sistemi di IA di cui all'allegato III[20]". Ed infine un sistema di IA di cui all'allegato III dell'AI Act è sempre considerato ad alto rischio qualora esso effettui profilazione di persone fisiche.

Questa sezione fornisce indicazioni per esaminare quando il sistema di IA può dirsi ad Alto Rischio ai sensi dell'art. 11 comma II dell'AI Act.
Lo schema di domande qui proposto serve ad aiutare l'operatore a valutare se il sistema di IA può essere considerato ad Alto Rischio:

- in presenza di un simbolo (*), il livello di attenzione deve essere elevato; è fortemente indicato di valutare con estrema attenzione se il sistema è da classificare come sistema ad Alto Rischio;
- in presenza di un simbolo (°), il livello di attenzione è meno elevato; è comunque indicato di valutare con attenzione se il sistema è da classificare come sistema ad Alto Rischio.

1. Il sistema ha un impatto sui diritti fondamentali delle persone?

La portata dell'impatto negativo del sistema di AI sui diritti fondamentali protetti dalla Carta è di particolare rilevanza ai fini della classificazione di un sistema di AI tra quelli ad alto rischio. Tali diritti comprendono il diritto alla dignità umana, il rispetto della vita privata e della vita familiare, la protezione dei dati personali, la libertà di espressione e di informazione, la libertà di riunione e di associazione e la non discriminazione, la protezione dei consumatori, i diritti dei lavoratori, i diritti delle persone con

[18] Modelli di IA per finalità generali.
[19] Allegato I dell'AI Act.
[20] Allegato III dell'AI Act.

disabilità, il diritto a un ricorso effettivo e a un giudice imparziale, i diritti della difesa e la presunzione di innocenza e il diritto a una buona amministrazione. Oltre a tali diritti, è importante sottolineare che i minori godono di diritti specifici sanciti dall'articolo 24 della Carta dell'UE e dalla Convenzione delle Nazioni Unite sui diritti del fanciullo (considerando 28).

Quesito	**SI**	**NO**
L'impiego del sistema di IA è in grado di influenzare il diritto alla privacy, alla non discriminazione, alla libertà di espressione o altri diritti garantiti e libertà della personalità[21]?	*	
Il sistema tratta dati personali o biometrici in modo che potrebbe mettere a rischio la privacy o l'identità di una persona?	*	
Il sistema potrebbe essere impiegato per la sorveglianza di massa o il controllo delle attività quotidiane delle persone?	*	
Esistono rischi di violazione della libertà di movimento o di assemblea?	*	
Motivazioni:		

2. In quali settori viene utilizzato il sistema di AI?

Ai sensi dell'allegato III dell'AI Act, del regolamento sono indicati otto settori che possono determinare la qualifica di alto rischio del sistema AI (Identificazione e categorizzazione biometrica delle persone fisiche; Gestione e funzionamento delle infrastrutture critiche; Istruzione e formazione professionale; Occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso a prestazioni e servizi pubblici e a servizi privati essenziali e fruizione degli stessi; Attività di contrasto; Gestione della migrazione, dell'asilo e del controllo delle frontiere; Amministrazione della giustizia e processi democratici).

Quesito	**SI**	**NO**
Il sistema è impiegato in un settore critico o sensibile di cui Allegato III dell'AI Act, dove errori o malfunzionamenti potrebbero avere conseguenze gravi?	*	
Un errore nel sistema potrebbe influenzare negativamente la vita, la salute o la sicurezza di una persona?	*	
Motivazioni:		

[21] I **diritti della personalità** sono diritti soggettivi assoluti che spettano all'essere persona in quanto tale, così funzionalmente diretti ad affermare e garantire esigenze di carattere esistenziale.

3. Il sistema tratta dati particolari?

I dati particolari sono quelli rientranti nelle categorie individuate dall'art. 9 del Regolamento UE 2016/679, cioè quelli che rivelano l'origine razziale od etnica, le convinzioni religiose, filosofiche, le opinioni politiche, l'appartenenza sindacale, relativi alla salute, i dati genetici, i dati biometrici e quelli relativi all'orientamento sessuale o alla vita sessuale.

Quesito	**SI**	**NO**
L'operazione coinvolge il trattamento di dati particolari?	*	
Il sistema viene utilizzato per identificare[22] o profilare[23] persone tramite dati biometrici (ad esempio, riconoscimento facciale, impronte digitali)?	*	
Motivazioni:		

4. Il sistema prende decisioni autonome o sostituisce il giudizio umano? Sono messe in atto misure di sorveglianza umana?

La sorveglianza umana aiuta a garantire che un sistema di IA non comprometta l'autonomia umana o provochi altri effetti negativi. La sorveglianza può avvenire mediante meccanismi di governance che consentano un approccio con intervento umano (human-in-the-loop - HITL), con supervisione umana (human-on-the-loop - HOTL) o con controllo umano (human-in-command – HIC). La sorveglianza umana deve essere garantita con misure individuate e integrate nel sistema di AI ad alto rischio dal fornitore prima della sua immissione sul mercato o messa in servizio, ove tecnicamente possibile; oppure con misure individuate dal fornitore prima dell'immissione sul mercato o della messa in servizio del sistema di IA ad alto rischio, adatte ad essere attuate dall'utente.

Quesito	**SI**	**NO**
Il sistema di IA interagisce con il processo decisionale degli utenti finali umani (ad esempio azioni raccomandate o decisioni da prendere, presentazione di opzioni)?	°	
Il sistema di AI prende decisioni senza una supervisione umana diretta o limita il controllo umano sulle decisioni finali?	°	
Le decisioni dell'AI possono influenzare in modo irreversibile gli individui o i loro diritti?	°	
È possibile eventualmente descrivere il livello di controllo o di coinvolgimento umano?		°
Gli output del sistema di IA vengono esaminati da un essere umano prima di essere applicati?		°
Motivazioni:		

[22] Vedi vocabolario.
[23] Vedi vocabolario.

5. C'è un concreto rischio di discriminazione o distorsione delle decisioni operata dall'AI?

Il rischio di tali risultati distorti ed effetti discriminatori è particolarmente importante per quanto riguarda l'età, l'etnia, la razza, il sesso o le disabilità.

Quesito	**SI**	**NO**
Il sistema potrebbe perpetuare discriminazioni basate su fattori come razza, etnia, genere o età?	*	
Esiste una concreta probabilità che si verifichi una variabilità delle decisioni a parità di condizioni?	°	
È stata prevista una strategia o una serie di procedure per evitare di creare o rafforzare distorsioni inique (unfair bias), sia per quanto riguarda l'uso dei dati di input che per la progettazione dell'algoritmo?		°
I dati usati per addestrare il sistema sono stati verificati in merito al fatto che non contengano pregiudizi che potrebbero perpetuare o amplificare comportamenti discriminatori?		*
Motivazioni:		

7. Qual è l'impatto potenziale in caso di errore o malfunzionamento?

I tipi di rischi e minacce dovrebbero essere basati su un approccio per singolo settore e per singolo caso. I rischi dovrebbero inoltre essere calcolati tenendo conto del loro impatto su diritti e sicurezza.

Quesito	**SI**	**NO**
Il sistema di IA può arrecare danni o nuocere agli utenti o a terze parti?	*	
I problemi di sicurezza o di rete (ad esempio i rischi di sicurezza informatica) comportano rischi per la sicurezza o arrecano danni a causa di comportamenti non intenzionali del sistema di IA?	*	
Un errore nel sistema può causare danni fisici, economici o psicologici alle persone coinvolte?	*	
Motivazioni:		

8. Il sistema è soggetto a manipolazioni esterne?

I sistemi di IA, come tutti i sistemi software, dovrebbero essere protetti contro le vulnerabilità che li espongono allo sfruttamento da parte degli avversari, ad esempio l'hacking. Gli attacchi possono colpire i dati (avvelenamento dei dati), il modello (model leakage) o l'infrastruttura sottostante, sia software che hardware.

Quesito	**SI**	**NO**
Gli effetti di un guasto del sistema di IA (che comportano risultati errati o l'indisponibilità del sistema) dà impatti socialmente inaccettabili?	*	

Motivazioni:		

9. Il sistema richiede un'autorizzazione preventiva da norma nazionale per l'uso?

Quesito	**SI**	**NO**
È necessario ottenere un'autorizzazione speciale per utilizzare questo sistema?	*	
Ci sono regolamenti o leggi che richiedono una valutazione d'impatto sui diritti fondamentali prima dell'implementazione?	*	
Motivazioni:		

Individuazione del livello di rischio per sistemi di IA di tipo "Determinati sistemi di IA"

I determinati sistemi di IA sono specificati nel Capo IV - Art. 50 dell'AI Act, che si suggerisce di consultare per comprenderne in dettaglio le caratteristiche.

Di seguito è riportata una sintesi di queste tipologie di sistemi[24]:

- Un sistema di IA che interagisce direttamente con persone fisiche[25].
- Un sistema di IA che genera contenuti sintetici/artificiali, *Deep Fake* o che genera/manipola testi di interesse pubblico;
- Un sistema di IA per il riconoscimento delle emozioni o per la categorizzazione biometrica.

Individuazione del livello di rischio per sistemi di IA di tipo "*General PurposeAI model*"

Al fine di individuare i sistemi di IA di tipo *General Purpose AI models* (modelli di IA per finalità generali), l'AI Act ha cercato di definire i *Foundation Models*, ovvero l'IA generativa, nel seguente modo: "*Un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato da una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere*

[24] Per una versione più approfondita: www.euaiact.com/key-issue/5.

[25] Si richiama lo standard internazionale ISO 9241-210: "*Interactive system: combination of hardware, software and/or services that receives input from, and communicates output to, users*." (P. 2, ISO 9241-210).

immessi sul mercato" (Art. III, AI Act). Tale definizione deve essere adottata per la classificazione di questi sistemi di IA.

Individuazione del livello di rischio per sistemi di IA di tipo "General Purpose AI model con rischio sistemico"

Al fine di classificare un "GPAI model con rischio sistemico" è necessario valutare se il GPAI model soddisfa una delle condizioni seguenti (Art. 51, AI Act):

- presenta capacità di impatto elevato valutate sulla base di strumenti tecnici e metodologie adeguati, compresi indicatori e parametri di riferimento;
- sulla base di una decisione della Commissione, ex officio o a seguito di una segnalazione qualificata del gruppo di esperti scientifici, presenta capacità o un impatto equivalenti a quelli di cui alla lettera a), tenendo conto dei criteri di cui all'allegato XIII dell'AI Act.

Si presume che un GPAI model presenti un **rischio sistemico** quando la **quantità cumulativa di calcolo utilizzata per il suo addestramento misurata in operazioni in virgola mobile (FLOPs) è superiore a 10^{25}**.

Individuazione del livello di rischio per sistemi di IA di tipo "Sistemi di IA con rischio minimo"

Nel caso in cui il sistema di IA preso in considerazione non rientri in nessuna delle categorie sopra indicate, esso è da classificare come sistema di IA a rischio minimo. Tale sistema non è soggetto ad alcun obbligo secondo l'AI Act.

Allegato B – Modello di Impact Assessment per i sistemi ad Alto Rischio[26]

Per l'utilizzo di sistemi di IA ad Alto Rischio è necessario avere piena consapevolezza delle loro caratteristiche e dei rischi che comportano. Per tale ragione è suggerito di redigere il seguente modello di documento di *Impact Assessment.* Ciò consente di realizzare una valutazione approfondita del sistema di IA che si andrà ad adottare, al fine di comprenderne in dettaglio le proprietà, il funzionamento, le eventuali interazioni con altri sistemi, le misure di sorveglianza adottate e le strategie di prevenzione dei rischi implementate.

Domanda	Descrizione
Nome del Modello di Impact Assessment	*Attribuire un titolo al trattamento oggetto di Impact Assessment*
Autore	*Inserire cognome e nome del soggetto che materialmente compila la FRIA*
Revisore	*Inserire cognome e nome del soggetto che revisiona la FRIA*
Validatore	*Inserire cognome e nome del soggetto che valida la FRIA*

[26] Il presente modello è definito sulla base del *Fundamental Rights Impact Assessment* (FRIA) dell'AI ACT.

1 Disamina del sistema di IA

Rispondendo alle seguenti domande si chiede di individuare e presentare le caratteristiche del sistema di IA oggetto dell'analisi. Si richiede, in particolare, di descrivere il sistema di IA, sia nella sua struttura che nei suoi rapporti con altri sistemi esterni e le eventuali implicazioni che derivino da tale interazione.

Domanda	Descrizione
Qual è il sistema di IA in considerazione?	*Presentare sinteticamente il sistema di IA: gli obiettivi, i soggetti che la sviluppano, la data e la versione del sistema, le forme in cui il sistema è immesso nel mercato, l'hardware su cui è destinato a operare.*
In che modo il sistema di IA interagisce con sistemi esterni?	*Presentare come il sistema interagisce o può essere utilizzato con harware o software che non fanno parte del sistema di IA stesso (ove applicabile).*

2. Descrizione del sistema di IA e del suo sviluppo

Rispondendo alle seguenti domande si chiede di definire e descrivere alcuni dettagli del sistema di IA in oggetto, in particolare per quanto riguarda i suoi elementi caratteristici, i destinatari dell'attività che il servizio svolge e il trattamento dei dati raccolti nel ciclo di vita del sistema. Così si potrà garantire maggiore trasparenza e tutela nella gestione dei dati raccolti.

Domanda	Descrizione
Quali sono gli elementi del sistema di IA e del processo relativo al suo sviluppo?	*Descrivere sinteticamente le specifiche di progettazione del sistema, le principali scelte di progettazione, i metodi applicati e le azioni eseguite per lo sviluppo del sistema (compresi – ove opportuno – il ricorso a sistemi preaddestrati forniti da terzi).*

Nei confronti di chi opera il sistema?	*Indicare le persone o i gruppi di persone sui quali il sistema è destinato a essere utilizzato.*
Quali sono gli aspetti che il sistema è progettato per ottimizzare?	*Descrivere le principali scelte di classificazione e gli aspetti che il sistema è chiamato a ottimizzare e la pertinenza dei diversi parametri.*
Quali sono e come vengono trattati i dati utilizzati dal sistema per le metodologie di addestramento?	*Indicare le metodologie e le tecniche di addestramento e i set di dati di addestramento utilizzati, comprese le informazioni sull'origine, sull'ambito di applicazione e sulle loro principali caratteristiche.*
Come vengono acquisiti i dati necessari ai fini dell'addestramento?	*Descrivere le modalità di ottenimento e di selezione dei dati, le procedure di etichettatura e le metodologie di pulizia dei dati.*
Quali misure di sorveglianza umana sono necessarie?	*Indicare la valutazione delle misure di sorveglianza umana necessarie in conformità dell'art. 14 dell'AI Act*

Quali sono le soluzioni tecniche adottate per garantire la conformità costante del sistema di IA?	*Descrivere le modifiche predeterminate del sistema di IA e delle sue prestazioni, unitamente a tutte le informazioni pertinenti relative alle soluzioni tecniche adottate per garantire la conformità costante del sistema di IA*
Quali sono le procedure di convalida e di prova utilizzate?	*Individuare le informazioni sui dati di convalida e di prova utilizzati e sulle loro principali caratteristiche; le metriche utilizzate per misurare l'accuratezza, la robustezza, la cibersicurezza e la conformità, nonché gli impatti potenzialmente discriminatori.*
Durante la fase di addestramento, il modello viene /è stato sottoposto a dati bilanciati e controlli per evitare bias e pregiudizi che potrebbero portare a allucinazioni?	*Descrivere sinteticamente le procedure adottate per evitare possibili pregiudizi.*

3. Capacità e grado di accuratezza del sistema di IA

Rispondendo alle seguenti domande si chiede di definire le modalità di applicazione e il grado di accuratezza del sistema di IA.

Domanda	Descrizione
Quali sono le capacità del sistema di IA e a quali limitazioni è sottoposto?	*Presentare le capacità del sistema di IA e le limitazioni in termini di prestazioni.*

Qual è il grado di accuratezza del sistema di IA?	*Indicare i gradi di accuratezza relativi a determinate persone o determinati gruppi di persone sui quali il sistema è destinato a essere utilizzato e il livello di accuratezza complessivo atteso in relazione alla finalità prevista del sistema.*
Gli errori prodotti dall'IA possono essere ritrattati o corretti prima che abbiano un impatto sul mondo reale?	*Individuare le tecniche utilizzate per evitare possibili rischi sugli utenti.*
Il sistema di IA informa gli utenti di eventuali incertezze o limiti dell'output?	*Indicare le modalità e le informazioni che vengono comunicate agli utenti in merito a incertezze o limiti dell'output del sistema.*
Esistono protocolli chiari che definiscono chi è responsabile nel caso in cui un errore o che causi danni?	*Indicare quali protocolli sono stati adottati per definire il responsabile di eventuali bias o errore che causino danni agli utenti nell'utilizzo del sistema di AI.*

4. Misure di prevenzione dei rischi

Rispondendo alle seguenti domande si chiede di definire i rischi a cui il sistema di IA si espone e come questi rischi vengono gestiti e prevenuti. È necessario illustrare quali siano i rischi che l'utilizzo del sistema di IA oggetto di interesse possa comportare. In particolare, sarà necessario conoscere i rischi per la salute e i diritti fondamentali e indicare come si cercherà, nell'utilizzo del sistema, di evitare il loro verificarsi. Conoscere tali rischi a monte, infatti, è particolarmente importante per poter portare avanti una attività di continuo monitoraggio durante l'utilizzo, con maggiore consapevolezza dei rischi che possono verificarsi e,

pertanto, agire ex ante affinché l'operato di tali sistemi non possa mai sfociare in una violazione dei diritti fondamentali.

Domanda	Descrizione
Come è strutturato il sistema di gestione dei rischi?	*Descrivere sinteticamente il sistema di gestione dei rischi predisposto in conformità dell'art. 9 dell'AI Act.*
Quali sono i rischi ragionevolmente prevedibili che il sistema di IA può porre?	*Indicare i rischi noti e ragionevolmente prevedibili che il sistema di IA ad alto rischio può porre per la salute, la sicurezza e i diritti fondamentali quando il sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista.*
Quali ulteriori rischi eventuali possono porsi a seguito dell'immissione del sistema nel mercato?	*Indicare altri eventuali rischi derivanti dall'analisi dei dati raccolti dal sistema di monitoraggio successivo all'immissione sul mercato di cui all'art. 72 dell'AI Act.*

5. Ciclo vitale del sistema di IA

Rispondendo alle seguenti domande, si chiede di descrivere il ciclo di vita del sistema di IA, con particolare riferimento alle modifiche apportate nel corso del suo sviluppo.

Domanda	Descrizione
Quale è il ciclo di vita del sistema di IA?	*Descrivere il ciclo di vita dei dati fornendo una dettagliata descrizione di ciascun processo effettuato.*

Quali modifiche sono apportare al sistema di IA durante il suo ciclo di vita?	*Individuare modifiche sostanziali che possono richiedere una nuova procedura di valutazione della conformità.*

6. Base giuridica del sistema di IA

Rispondendo alla seguente domanda si chiede di individuare le norme armonizzate a cui si è fatto riferimento e che sono state applicate.

Domanda	Descrizione
Quali norme armonizzate sono state applicate?	*Fornire un elenco delle norme armonizzate applicate integralmente o in parte; nei casi in cui tali norme armonizzate non sono state applicate, una descrizione dettagliata delle soluzioni adottate per soddisfare i requisiti di cui al Capo III, Sezione 1, Art. 6 dell'AI Act, compreso un elenco delle altre norme e specifiche tecniche pertinenti applicate.*

7. Valutazione delle prestazioni nella fase di immissione nel mercato

Rispondendo alla seguente domanda si chiede di descrivere il sistema utilizzato per valutare il sistema di IA nella fase successiva all'immissione sul mercato ai sensi dell'art. 72 dell'AI Act.

Qual è il sistema predisposto per valutare le prestazioni del sistema di IA nella fase successiva all'immissione nel mercato?	*Indicare il sistema di monitoraggio successivo all'immissione sul mercato che sia proporzionato alla natura delle tecnologie di IA e ai rischi del sistema di IA ad alto rischio.*

8. Trasparenza e comprensibilità del sistema di IA

I sistemi di IA ad Alto Rischio devono essere progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e

utilizzarlo adeguatamente (Art. 13, AI Act). Si chiede di inserire le informazioni richieste nella tabella seguente.

Descrivere le modalità con cui gli utenti sono adeguatamente informati sulle modalità di funzionamento del sistema di IA e sulle sue decisioni	
Descrivere le caratteristiche di interpretabilità e comprensibilità del sistema di IA. Ovvero come gli utenti possono interpretare e comprendere le ragioni e i criteri che determinano i risultati del sistema di IA.	

9. Caratteristiche del sistema rispetto a manipolazioni esterne

I sistemi di IA, come tutti i sistemi software, dovrebbero essere protetti contro le vulnerabilità che li espongono allo sfruttamento da parte degli "avversari", ad esempio l'hacking. Gli attacchi possono colpire i dati (avvelenamento dei dati), il modello (model leakage) o l'infrastruttura sottostante, sia software che hardware. Si chiede di inserire le informazioni richieste nella seguente tabella:

Descrivere come il sistema prevede di comportarsi in caso di adversarial attack o di altre situazioni analoghe impreviste	

Allegato C – Modello di *Impact Assessment* generale

Per i sistemi di IA non ad Alto Rischio si suggerisce il seguente modello di documento di Impact Assessment semplificato rispetto a quello riportato sopra.

Impact Assessment per il sistema di IA: "Nome del sistema"

Informazioni

Nome del sistema o identificativo	
Fase di vita	
Autore	
Data ultima modifica	
Revisionato da	
Data della revisione	
Approvato da	
Data approvazione	

- Descrizione del sistema

- Caratteristiche del sistema

N.	Caratteristica	Nella versione corrente?	In una versione futura? Data stimata per il rilascio

- Interazioni con altri sistemi

- Obiettivi del sistema

- Uso previsto

Informazioni sui dati e la qualità

- Informazioni sui dati

- Documentazione sulla qualità dei dati

Informazioni sugli algoritmi e sui modelli

Informazioni sull'ambiente di implementazione

- Informazioni generali

- Informazioni su vincoli e complessità

Parti interessate

Parti potenzialmente interessate dal sistema di AI e dal suo utilizzo

Benefici e rischi attuali e potenziali

Rischio	Probabilità	Impatto	Misure di Mitigazione

Fallimenti del sistema, uso improprio e abuso

- Fallimenti

- Uso improprio e abuso

Allegato D – Modello di documentazione generale

Il seguente modello di documentazione suggerisce di inserire le informazioni ritenute essenziali per descrivere un sistema di IA. In particolare si indica di riportare le seguenti informazioni, che sono state riprese e adattate dalle specifiche di documentazione indicate per i modelli GPAI (Capo V – Sezione 2 – Art. 53 e Allegato XI dell'AI Act):[27]

- una descrizione del sistema di IA che includa:
  - i compiti che il sistema è destinato a eseguire;
  - l'architettura e il numero di parametri;
  - la modalità (ad esempio testo, immagine) e il formato degli input e degli output;
  - la licenza.
- una descrizione dettagliata degli elementi del sistema di cui al punto sopra e informazioni pertinenti sul processo di sviluppo, compresi gli elementi seguenti:
  - le specifiche di progettazione del sistema e del processo di addestramento, comprese le metodologie e le tecniche di addestramento, le principali scelte progettuali, comprese le motivazioni e le ipotesi formulate;
  - informazioni sui dati utilizzati per l'addestramento, la prova e la convalida, se del caso, compresi il tipo e la provenienza dei dati e le metodologie di organizzazione (ad esempio pulizia, filtraggio, ecc.); il modo in cui i dati sono stati ottenuti e selezionati e tutte le altre misure per rilevare l'inadeguatezza delle fonti di dati e i metodi per rilevare distorsioni identificabili, se del caso;
  - le risorse computazionali utilizzate per addestrare il modello[28] (ad esempio il numero di operazioni in virgola mobile), il tempo di addestramento e altri dettagli pertinenti relativi all'addestramento;
  - il consumo energetico noto o stimato del sistema. Se il consumo energetico del sistema non è noto, il consumo energetico può basarsi su informazioni relative alle risorse computazionali utilizzate.

[27] Per quanto concerne i sistemi che adottano soluzioni Open Source, è sufficiente riferire al codice sorgente e alla documentazione online.

[28] Nel caso il sistema in esame preveda l'utilizzo di un modello preaddestrato si riportino i riferimenti relativi a quest'ultimo. Qualora invece il sistema preveda una fase di addestramento/tuning del modello i riferimenti da riportare sono riferiti anche a questa fase.

Allegato E – Modello di documentazione per i *GPAI models*

Per quanto riguarda la documentazione relativa ai G*PAI models* si propone di utilizzare quella indicata nell'AI Act (Capo V, Sezioni 2 e 3, AI Act).

Documentazione per GPAI models senza rischio sistemico

Si suggerisce di inserire le seguenti informazioni (Art. 53 e Allegato XI, AI Act):

- una descrizione del modello di IA che includa:
  - i compiti che il modello è destinato a eseguire e il tipo e la natura dei sistemi di IA in cui può essere integrato (se previsto e possibile);
  - le politiche di utilizzo accettabili applicabili;
  - la data di pubblicazione e i metodi di distribuzione;
  - l'architettura e il numero di parametri;
  - la modalità (ad esempio testo, immagine) e il formato degli input e degli output;
  - la licenza.
- una descrizione dettagliata degli elementi del modello di cui al punto sopra e informazioni pertinenti sul processo di sviluppo, compresi gli elementi seguenti:
  - mezzi tecnici (ad esempio istruzioni per l'uso, infrastruttura, strumenti) necessari per integrare il sistema di IA in altri sistemi di IA (se previsto e possibile);
  - le specifiche di progettazione del sistema e del processo di addestramento, comprese le metodologie e le tecniche di addestramento, le principali scelte progettuali, comprese le motivazioni e le ipotesi formulate; gli aspetti che il sistema è progettato per ottimizzare e la pertinenza dei diversi parametri, se del caso;
  - informazioni sui dati utilizzati per l'addestramento, la prova e la convalida, se del caso, compresi il tipo e la provenienza dei dati e le metodologie di organizzazione (ad esempio pulizia, filtraggio, ecc.), il numero di punti di dati, la loro portata e le principali caratteristiche; il modo in cui i dati sono stati ottenuti e selezionati e tutte le altre misure per rilevare l'inadeguatezza delle fonti di dati e i metodi per rilevare distorsioni identificabili, se del caso;

- le risorse computazionali utilizzate per addestrare il modello (ad esempio il numero di operazioni in virgola mobile), il tempo di addestramento e altri dettagli pertinenti relativi all'addestramento;
- il consumo energetico noto o stimato del sistema. Se il consumo energetico del sistema non è noto, il consumo energetico può basarsi su informazioni relative alle risorse computazionali utilizzate.

Informazioni sulla trasparenza per i provider dei GPAI models verso i fornitori a valle che integrano il modello nel loro sistema di IA (Art. 53, par 1, lettera b), AI Act)

Fatta salva la necessità di rispettare e proteggere i diritti di proprietà intellettuale e le informazioni commerciali riservate o i segreti commerciali conformemente al diritto dell'Unione e nazionale, si suggerisce ai **provider** di elaborare, mantenere aggiornate e mettere a disposizione le seguenti ulteriori informazioni e documentazione per i **provider di sistemi di IA che intendono integrare il proprio sistema di IA** con un GPAI model (Art. 53 e Allegato XII, AI Act).

Le informazioni e la documentazione dovrebbero:

- consentire ai fornitori di sistemi di IA di avere una buona comprensione delle capacità e dei limiti del modello di IA per finalità generali e di adempiere ai loro obblighi a norma del presente regolamento; nonché
- contenere:
  - una descrizione generale del modello di IA comprendente:
    - i compiti che il modello è destinato a eseguire e il tipo e la natura dei sistemi di IA in cui può essere integrato;
    - le politiche di utilizzo accettabili applicabili;
    - la data di pubblicazione e i metodi di distribuzione;
    - il modo in cui il modello interagisce o può essere utilizzato per interagire con hardware o software che non fanno parte del modello stesso, ove applicabile;
    - le versioni del software pertinente relative all'uso del modello di IA per finalità generali, se del caso;
    - l'architettura e il numero di parametri;

        - la modalità (ad esempio testo, immagine) e il formato degli input e degli output;
        - la licenza per il modello.
- una descrizione degli elementi del modello e del processo relativo al suo sviluppo, compresi:
    - i mezzi tecnici (ad esempio istruzioni per l'uso, infrastruttura, strumenti) necessari per integrare il sistema di IA in altri sistemi di IA;
    - la modalità (ad esempio testo, immagine, ecc.) e il formato degli input e degli output e la loro dimensione massima (ad esempio, lunghezza della finestra contestuale, ecc.);
- informazioni sui dati utilizzati per l'addestramento, la prova e la convalida, se del caso, compresi il tipo e la provenienza dei dati e le metodologie di organizzazione.

Documentazione per GPAI models con rischio sistemico

In aggiunta agli obblighi di cui agli articoli 53 e 54, riferiti ai GPAI models senza rischio sistemico, l'art 55 dell'IA Act prevede una serie di obblighi per i fornitori di GPAI models con rischio sistemico. L'Allegato XI dell'AI Act nella sezione 2 indica, le informazioni aggiuntive che gli stessi devono inserire nella documentazione, e che sono:

- una descrizione dettagliata delle strategie di valutazione, compresi i risultati della valutazione, sulla base dei protocolli e degli strumenti di valutazione pubblici disponibili o di altri metodi di valutazione;
- se del caso, una descrizione dettagliata delle misure messe in atto al fine di effettuare il test contraddittorio (adversarial testing) interno e/o esterno (ad esempio, red teaming), adeguamenti dei modelli, compresi l'allineamento e la messa a punto;
- se del caso, una descrizione dettagliata dell'architettura del sistema che spiega in che modo i componenti software si basano l'uno sull'altro o si alimentano reciprocamente e si integrano nel processo complessivo

Allegato F – Glossario[29]

***Artificial Neural Network* (ANN) - Rete Neurale Artificiale.** Una rete neurale artificiale è un modello computazionale ispirato al funzionamento del cervello umano, composto da strati di nodi (neuroni) interconnessi che elaborano dati per apprendere modelli complessi tramite aggiustamenti iterativi dei pesi delle connessioni.

***Autoencoder* (AE).** Un *autoencoder* è un tipo di rete neurale artificiale utilizzata principalmente per apprendere codifiche efficienti di dati non etichettati, con la finalità per esempio di riduzione della dimensionalità, e di rimozione del rumore. Si tratta di un'architettura di apprendimento non supervisionato progettata per imparare una rappresentazione compressa dei dati in ingresso.

***Convolutional Neural Network* (CNN).** Una *Convolutional Neural Network* è una rete neurale in cui il modello di connettività tra i neuroni è ispirato all'organizzazione della corteccia visiva degli animali. Le CNN sono reti neurali specializzate progettate per elaborare dati visivi, come immagini e video, ma funzionano bene anche con dati non visivi (ad esempio, nel *Natural Language Processing*).

Distributore. Una persona fisica o giuridica nella catena di approvvigionamento, diversa dal fornitore o dall'importatore, che mette a disposizione un sistema di IA sul mercato dell'Unione (AI Art. III, AI Act).

Dati biometrici. I dati personali ottenuti da un trattamento tecnico specifico relativi alle caratteristiche fisiche, fisiologiche o comportamentali di una persona fisica,quali le immagini facciali o i dati dattiloscopici (Art. III, AI Act).

***Deep Fake*.** Un'immagine o un contenuto audio o video generato o manipolato dall'IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona (Art. III, AI Act).

***Deep Learning*.** Il *Deep Learning* è un sottocampo del machine learning che utilizza reti neurali artificiali (*Artificial Neural Network* – ANN) con molti strati per analizzare grandi quantità di dati e apprendere rappresentazioni complesse.

Deployer. Qualsiasi persona fisica o giuridica, autorità pubblica, agenzia o altro organismo che utilizza un sistema di IA sotto la propria autorità , tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale (Art. III, AI Act).

Determinati Sistemi di IA. Vedi definizione dedotta dall'Art 50 dell'AI Act.

[29] Vedere anche AI Act – Capo I – Art. 3 – Definizioni.

Fornitore (Provider). Una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un "**modello di IA per scopi generali"** (o che fa sviluppare un sistema di IA o un modello di IA per scopi generali) e li immette sul mercato o mette in servizio il sistema con il proprio nome o marchio, a titolo oneroso o gratuito (Art. III, AI Act).

Fundamental Rights Impact Assessment **(FRIA):** è la Valutazione di impatto sui diritti fondamentali. Mediante essa, infatti, si è in grado di valutare l'impatto che l'utilizzo delle tecnologie ha sui diritti individuali.

Generative Adversarial Network **(GAN).** Una *Generative Adversarial Network* è un tipo di rete neurale utilizzata per generare nuovi dati simili a quelli di un *dataset* di riferimento. Le GAN sono utilizzate nella creazione di immagini, video, e testi.

Generative Pre-trained Transformer **(GPT).** Un Generative *Pre-trained Transformer* è un modello di intelligenza artificiale basato sull'architettura dei *Transformer*, progettato per generare testo in modo coerente e naturale. GPT è un modello di apprendimento non supervisionato che viene prima pre-addestrato (*pre-trained*) su grandi quantità di testo e poi affinato (*fine-tuned*) per compiti specifici come la traduzione, la generazione di testo o il riassunto automatico.

Importatore. Una persona fisica o giuridica ubicata o stabilita nell'Unione che immette sul mercato un sistema di IA recante il nome o il marchio di una persona fisica o giuridica stabilita in un paese terzo (Art. III, AI Act).

General Purpose **AI (GPAI) model - Modello di IA per finalità generali.** Un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato da una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato (Art. III, AI Act).

General Purpose **AI (GPAI) - Modello di IA per finalità generali (con rischio sistemico)** (Art. 51, AI Act). Vedere Allegato E.

General Purpose **AI (GPAI) system – Sistema di IA per finalità generali.** Per sistema GPAI si intende un sistema di IA basato su un modello GPAI, in grado di servire una varietà di scopi, sia per l'uso diretto che per l'integrazione in altri sistemi di IA.

***Generative AI* (Gen AI).** La Generative AI è un ramo dell'intelligenza artificiale che crea nuovi contenuti, come testi, immagini, audio o video, utilizzando modelli addestrati su grandi quantità di dati per generare risultati originali e simili a quelli creati dall'uomo.

Identificare. Riguarda la capacità di un sistema di IA di riconoscere o determinare l'identità di una persona o di un oggetto. Questo processo può avvenire attraverso diverse tecnologie, come: il riconoscimento facciale (*Face Recognition*), il riconoscimento biometrico (come impronte digitali, voce o andatura), analisi di pattern comportamentali (movimenti, clic online, ecc.). Nel contesto nell'AI Act i sistemi che effettuano l'identificazione biometrica a distanza in tempo reale o in differita sono classificati come ad alto rischio o addirittura vietati in determinati scenari.

Identificazione biometrica. Il riconoscimento automatizzato delle caratteristiche umane fisiche, fisiologiche, comportamentali o psicologiche allo scopo di determinare l'identità di una persona fisica confrontando i suoi dati biometrici con quelli di individui memorizzati in una banca dati (Art. III, AI Act).

***Large Language Model* (LLM).** Un LLM è un modello linguistico in grado di ottenere la comprensione e la generazione di linguaggio di ambito generale. I LLM acquisiscono questa capacità adoperando enormi quantità di dati per apprendere miliardi di parametri nell'addestramento e consumando grandi risorse di calcolo sia nell'addestramento che nell'operatività. L'aggettivo l*arge* presente nel nome si riferisce alla grande quantità di parametri del modello probabilistico, nell'ordine dei miliardi. I LLM sono in larga parte reti neurali artificiali e in particolare trasformer e sono (pre-)addestrati usando l'apprendimento auto-supervisionato o l'apprendimento semi-supervisionato.

***Machine Learning*.** Il *Machine Learning* è una branca dell'intelligenza artificiale in cui i computer apprendono dai dati per migliorare le proprie prestazioni su un compito senza essere esplicitamente programmati.

Profilare. Si riferisce alla capacità di un sistema di IA di analizzare dati personali e comportamentali di una persona per generare un profilo dettagliato che predica, deduce o descrive caratteristiche, preferenze o comportamenti di un individuo. Esempi includono: suggerimenti personalizzati basati sulla cronologia di navigazione, classificazione del rischio creditizio o assicurativo, analisi di tendenze politiche o preferenze di acquisto. Nel contesto dell'AI Act la profilazione è strettamente regolata quando utilizzata in contesti sensibili, come per esempio nel settore lavorativo per assunzioni o promozioni, nel settore bancario la valutazione di solvibilità, o nella Pubblica Amministrazione per l'allocazione di risorse o benefici.

Sistema di Intelligenza Artificiale. Un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione 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 (Art. III, AI Act).

***Reinforcement Learning* (RL).** L'Apprendimento per Rinforzo è una tecnica di apprendimento delle reti neurali artificiali che si aggiunge a quello di tipo supervisionato e non supervisionato. Non richiede dati di input etichettati ed è un paradigma basato sull'interazione con un ambiente e sull'apprendimento tramite feedback. Il suo obiettivo è massimizzare una ricompensa definita in base all'output della rete.

***Recurrent Neural Network* (RNN).** Una RNN è un tipo di rete neurale artificiale progettata per gestire dati sequenziali mantenendo una memoria degli input precedenti attraverso connessioni ricorrenti. A differenza delle reti neurali *feedforward*, le RNN hanno *loop* che permettono alle informazioni di persistere attraverso i vari passi temporali, rendendole adatte a compiti in cui il contesto e l'ordine sono importanti.

Supervised Learning (Apprendimento Supervisionato). Il *Supervised Learning* è un paradigma di apprendimento automatico in cui un modello di AI viene addestrato utilizzando un set di dati etichettato, ovvero con input associati a output corretti (etichette). L'obiettivo del modello è apprendere la relazione tra input e output per poter fare previsioni accurate su nuovi dati.

Training Set (Dati di addestramento). I dati utilizzati per addestrare un sistema di IA adattandone i parametri che può apprendere (Art. III, AI Act).

***Transformer*.** Il *Transformer* è un'architettura di rete neurale introdotta nel 2017 nell'articolo "*Attention Is All You Need*" (Vaswani et al.). Ha rivoluzionato il campo del *Natural Language Processing* (NLP) e viene utilizzato in modelli avanzati come GPT (*Generative Pre-trained* Transformer) e BERT (*Bidirectional Encoder Representations from Transformers*).

***Unsupervised Learning* (Apprendimento non Supervisionato).** L'*Unsupervised Learning* è un paradigma di apprendimento automatico in cui un modello viene addestrato su dati non etichettati, ovvero senza output predefiniti. L'obiettivo è scoprire strutture, pattern o raggruppamenti nascosti nei dati senza un supervisore esplicito.

Allegato G – Domande e FAQ

In questa sezione sono riportate alcune domande utili ad evidenziare informazioni sull'AI e sui sistemi di AI, in particolare in riferimento AI Act.

Domande

- ***Sono consapevole di cosa è un sistema di IA[30]?***
- ***La mia attività si configura come deployer, fornitore, o distributore di un sistema di IA?***
- ***I sistemi di IA che sto distribuendo o fornendo come si inquadrano nell'ambito dell'AI Act?*** (Proibito - Capo II, AI Act, Alto rischio - Capo III, AI Act, determinati sistemi di IA - Capo IV, AI Act, Modelli GPAI sistemici o non sistemici - Capo V, AI Act, sistemi a rischio minimo)
- ***Ho fatto un test con EU AI Act Compliance Checker[31]***?
- ***Ho fatto una prova di self-assessment (autovalutazione) rispetto ai sistemi di IA con la The Assessment List for Trustworthy Artificial Intelligence (Lista di Valutazione per un'Intelligenza Artificiale Affidabile)[32]?***
- ***Chi ha accesso al sistema di IA durante ogni fase del suo sistema di vita?***
- ***Ho impiegato delle misure corrette per poter proteggere i dati personali?***
- ***Ho definito propriamente le finalità del sistema fin dalla fase di progettazione?***
- ***Qual è la catena di approvvigionamento nel mio sistema di IA?***
- ***Il venditore del sistema di IA che ho acquistato ha rispettato gli standard richiesti in tema di cybersicurezza?***
- ***Può un malfunzionamento del sistema di IA impattare sull'organizzazione dell'ente?***

Frequently Asked Questions

- ***Sto sviluppando e mettendo in servizio sistemi di IA o modelli di IA, al solo scopo di ricerca e sviluppo scientifici, devo tenere di conto dell'AI Act?***

  No, in questi casi non si applica.

30 Vedi la definizione riportata sopra.
31 https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/
32 file:///C:/Users/GV20888/Desktop/Aperti/The%20Assessment%20List%20for%20Trustworthy%20Artificial%20Intelligence

- ***Sto realizzando attività di ricerca, prova o sviluppo relative a sistemi di IA o modelli di IA prima della loro immissione sul mercato o messa in servizio, devo tenere di conto dell'AI Act?***

  No, in questi casi non si applica, salvo se sto facendo prove in condizioni reali.

- ***L'AI Act si applica ai sistemi di IA rilasciati con licenza libera e open source?***

  No non si applica, a meno che non siano immessi sul mercato o messi in servizio come sistemi di IA ad alto rischio, o come sistemi di IA proibiti (Art. 5, AI Act) o come determinati sistemi di IA (Art. 50, AI Act).

- ***Sono una persona fisica deployer che utilizza sistemi di IA nel corso di un'attività non professionale puramente personale, devo tenere di conto dell'AI Act?***

  No, in questi casi non si applica.

Allegato H – Approfondimenti: IA e AI Act

In questo paragrafo sono riportati alcuni riferimenti riguardanti l'IA e le sue possibili applicazioni, oltre a riferimenti di supporto per l'interpretazione e l'applicazione dell'AI Act.

- "Intelligenza artificiale" AGID (Agenzia per l'Italia Digitale)[33].
- "L'AI generativa e i servizi pubblici digitali: scenario e possibili applicazioni"[34].
- Artificial Intelligence for Sustainable Development" (Italian Ministry of Foreign Affairs, 2021)[35] .
- "Robustness and explainability of artificial intelligence"[36].
- "AI and GenAI adoption by local and regional administrations"[37].
- "Artificial intelligence: a modern approach"[38].
- The Neural Network Zoo[39].
- The EU Artificial Intelligence Act. Up-to-date developments and analyses of the EU AI Act[40].
- Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)[41].

33 https://www.agid.gov.it/it/ambiti-intervento/intelligenza-artificiale.
34 https://www.forumpa.it/pa-digitale/lai-generativa-e-i-servizi-pubblici-digitali-scenario-e-possibili-applicazioni/.
35 Artificial Intelligence for Sustainable Development" (Italian Ministry of Foreign Affairs, 2021).
36 European Commission, "Robustness and explainability of artificial intelligence," European Commission, Tech. Rep., 2020. [Online]. Available: https://publications.jrc.ec.europa.eu/repository/handle/JRC119336.
37 https://interoperable-europe.ec.europa.eu/collection/portal/news/study-ai-and-genai-adoption-local-regional-administrations.
38 S. J. Russell and P. Norvig, Artificial intelligence: a modern approach. Pearson,
39 https://www.asimovinstitute.org/neural-network-zoo/.
40 https://artificialintelligenceact.eu/.
41 https://ec.europa.eu/newsroom/dae/redirection/document/112367.