Bozza di linee guida per l’adozione di IA nella pubblica amministrazione

Ai sensi del D.P.C.M. 12 gennaio 2024, recante “Piano triennale per l’informatica nella pubblica amministrazione 2024-2026”

Versione 1.0 del 14.02.2025 – Consultazione pubblica

Gruppo di lavoro

Il documento è stato redatto nell'ambito del Tavolo di concertazione del piano triennale per l'informatica nella pubblica amministrazione. Ai lavori hanno partecipato:

- AgID - Agenzia per l'Italia Digitale;
- ACN - Agenzia per la Cybersicurezza Nazionale
- ANAC - Autorità Nazionale Anticorruzione
- ANCI - Associazione Nazionale Comuni Italiani
- Conferenza delle Regioni e delle Province autonome
- Consip
- Dipartimento della Funzione Pubblica della Presidenza del Consiglio dei Ministri
- Dipartimento per la Trasformazione Digitale della Presidenza del Consiglio dei Ministri
- INAIL - Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro
- INPS - Istituto Nazionale della Previdenza Sociale
- ISTAT - Istituto Nazionale di Statistica
- IPZS - Istituto Poligrafico e Zecca dello Stato
- Ministero dell'Economia e delle Finanze
- Ministero delle Imprese e del Made in Italy
- PagoPA
- Unione Provincie d'Italia

Indice

Introduzione 6
1. Ambito di applicazione 7
1.1. Ambito soggettivo 7
1.2. Ambito oggettivo 7
2. Riferimenti e sigle 8
2.1. Note di lettura del documento 8
2.2. Struttura 8
2.3. Riferimenti normativi 8
2.4. Linee Guida di riferimento 10
2.5. Acronimi 12
3. L'Intelligenza Artificiale 13
3.1. Ciclo di vita 15
3.2. I ruoli nella catena del valore dell'IA 16
3.3. Classificazione dei sistemi di IA sulla base del rischio 17
3.4. Principi per l'adozione dell'IA nella pubblica amministrazione 19
4. Modello di adozione dell'IA 22
4.1. Strategia per l'IA 22
4.2. Analisi del contesto e delle caratteristiche della PA 23
4.3. Obiettivi e ambiti prioritari di applicazione 24
4.4. Norme tecniche 26
4.5. Casi d'uso 26
4.5.1. Funzionalità dell'IA 27
4.5.2. Requisiti 27
4.5.3. Indicatori di prestazione 28
4.6. Governance 28
4.7. Gestione del rischio 29
4.8. Valutazione d'impatto 29
4.9. Piano operativo 30
4.10. Risorse, competenze, comunicazione 30
4.11. Implementazione 31
4.12. Monitoraggio e valutazione 31
4.13. Miglioramento continuo 31
5. Conformità delle soluzioni di IA 32
5.1. Monitoraggio del ciclo di vita delle soluzioni di IA 32

5.2. Misure di sorveglianza 34
5.3. Auditing e controlli nei confronti dei fornitori esterni 34
5.4. Conservazione della documentazione 35
6. Governance etica dell'IA 36
7. Comunicazione 39
7.1. Misure di trasparenza 39
7.2. Obblighi di informativa 40
7.3. Adottare l'IA nella comunicazione istituzionale 41
8. Formazione e sviluppo delle competenze 43
8.1. Indicazioni operative per le PA 47
9. Gestione e qualità dei dati 51
9.1. Tipologie di dati 52
9.2. Caratteristiche dei dati 54
9.3. Processi e governance dei dati 57
10. Protezione dei dati personali 64
11. Sicurezza cibernetica 67
11.1. Tassonomie di attacco 67
11.1.1. Evasion attacks 68
11.1.2. Poisoning attacks 68
11.1.3. Privacy attacks 69
11.1.4. Abuse attacks 69
11.2. Gestione del rischio cibernetico 69
11.3. Asset 71
11.4. Minacce 72
11.5. Obiettivi di sicurezza 73
11.6. Gestione integrata della sicurezza dei sistemi di IA 75
Allegati 78
A. Valutazione del livello di maturità nell'adozione di IA 79
B. Valutazione del rischio 85
C. Valutazione d'impatto 91
D. Modello di codice etico 92
E. Norme tecniche in ambito IA 100
F. Casi d'uso 102
G. Funzionalità dell'IA 105
H. Procedure di governance 110
I. Indicatori di prestazione 115

Introduzione

Nell'ambito dell'utilizzo da parte delle Pubbliche Amministrazioni (PA) di sistemi di Intelligenza Artificiale (IA), anche ai fini dell'articolo 2, comma 1 del D. Lgs. 7 marzo 2005, n. 82, recante "Codice dell'amministrazione digitale" (di seguito CAD), le Linee Guida per l'adozione, l'acquisto, lo sviluppo di sistemi di intelligenza artificiale nella Pubblica Amministrazione sono previste dal D.P.C.M. 12 gennaio 2024, recante "Piano triennale per l'informatica nella pubblica amministrazione 2024-2026".

Le Linee Guida per l'adozione, l'acquisto e lo sviluppo di sistemi di intelligenza artificiale nella Pubblica Amministrazione sono emanate seguendo l'iter previsto all'articolo 71 del CAD e nel rispetto della Determinazione dell'Agenzia per l'Italia Digitale (di seguito AgID) n. 160 del 17 maggio 2018 recante "Regolamento per l'adozione di linee guida per l'attuazione del Codice dell'Amministrazione Digitale".

1. Ambito di applicazione

1.1. Ambito soggettivo

Le presenti Linee Guida per l'Adozione di Intelligenza Artificiale nella Pubblica Amministrazione (di seguito Linee Guida) sono rivolte ai soggetti di cui all'articolo 2, comma 2, del CAD. Tali soggetti sono indicati nel documento con l'acronimo PA.

1.2. Ambito oggettivo

Le presenti Linee Guida concernono le modalità di adozione dei sistemi di Intelligenza Artificiale con particolare riferimento agli aspetti di conformità normativa e di impatto organizzativo.

2. Riferimenti e sigle

2.1. Note di lettura del documento

Le presenti Linee Guida adottano i termini chiave: «DEVE», «DEVONO», «NON DEVE», «NON DEVONO», «DOVREBBE», «NON DOVREBBE», «PUÒ», «POSSONO» e «OPZIONALE», definiti dalle norme ISO/IEC Directives, Part 2 *"Principles and rules for the structure and drafting of ISO and IEC documents"* per la stesura dei documenti tecnici. L'interpretazione di tali termini nell'ambito delle Linee Guida è descritta di seguito.

- DEVE o DEVONO indicano un requisito obbligatorio;
- NON DEVE o NON DEVONO o NON PUÒ o NON POSSONO indicano un assoluto divieto;
- DOVREBBE o NON DOVREBBE indicano che le implicazioni devono essere comprese e attentamente pesate prima di scegliere approcci alternativi;
- PUÒ o POSSONO o l'aggettivo OPZIONALE indica che il lettore può scegliere di applicare o meno la specifica.

2.2. Struttura

Considerata la velocità dell'innovazione, le Linee Guida devono garantire un adattamento costante ai cambiamenti imposti dall'incessante rivoluzione digitale. Di qui la scelta di corredare le Linee Guida di allegati chiamati "strumenti" i cui contenuti potranno essere adeguati agevolmente all'evoluzione tecnologica.

Le presenti linee guida includono i seguenti allegati:

- Allegato A: Valutazione del livello di maturità nell'adozione di IA
- Allegato B: Valutazione del rischio
- Allegato C: Valutazione d'impatto
- Allegato D: Modello di codice etico
- Allegato E: Norme tecniche rilevanti per l'adozione dell'IA
- Allegato F: Casi d'uso
- Allegato G: Funzionalità dell'IA
- Allegato H: Procedure di governance
- Allegato I: Indicatori di prestazione.

I contenuti dei suddetti allegati potranno essere adeguati a seguito di evoluzioni tecnologiche. Il processo di costante adeguamento degli allegati è realizzato in coerenza con il quadro normativo in materia di digitalizzazione e, nello specifico, ai sensi dell'articolo 14-bis, comma 2, lettera a) del CAD.

2.3. Riferimenti normativi

Sono riportati di seguito i principali atti normativi di riferimento per le presenti linee guida.

[CDFUE]	Carta dei diritti fondamentali dell'Unione Europea 2012/C 326/02
[AI Act]	Regolamento europeo (UE) 1689/2024 del 13 giugno 2024 del Parlamento Europeo e del Consiglio, 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).
[Data Act]	Regolamento europeo (UE) 2023/2854 del 13 dicembre 2023 del Parlamento Europeo e del Consiglio, riguardante norme armonizzate sull'accesso equo ai dati e sul loro utilizzo e che modifica il regolamento (UE) 2017/2394 e la direttiva (UE) 2020/1828 (regolamento sui dati).
[DGA]	Regolamento europeo (EU) 2022/868 del 30 maggio 2022 del Parlamento Europeo e del Consiglio, relativo alla governance europea dei dati e che modifica il regolamento (UE) 2018/1724 (Regolamento sulla governance dei dati).
[Direttiva Open Data]	Direttiva (UE) 2019/1024 del Parlamento europeo e del Consiglio, del 20 giugno 2019, relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico (rifusione).
[NIS 2]	Direttiva (UE) 2022/2555 del Parlamento europeo e del Consiglio del 14 dicembre 2022, relativa a misure per un livello comune elevato di cybersicurezza nell'Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva (UE) 2018/1972 e che abroga la direttiva (UE) 2016/1148.
[GDPR]	Regolamento (UE) 2016/679 del 27 aprile 2016 del Parlamento europeo e del Consiglio, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).
[Reg.UE 2018/1725]	Regolamento (UE) 2018/1725 del Parlamento europeo e del Consiglio, del 23 ottobre 2018, sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il regolamento (CE) n. 45/2001 e la decisione n. 1247/2002/CE.
[Dir. Prodotti]	Direttiva (UE) 2024/2853 del Parlamento europeo e del Consiglio del 23 ottobre 2024 sulla responsabilità per danno da prodotti difettosi, che abroga la direttiva 85/374/CEE del Consiglio.
[Codice privacy]	Decreto legislativo 30 giugno 2003, n. 196 e s.m.i. recante "Codice in materia di protezione dei dati personali".

[CAD]	Decreto legislativo 7 marzo 2005, n. 82 e s.m.i. recante "Codice dell'amministrazione digitale".
[Piano Triennale]	D.P.C.M. 12 gennaio 2024, recante "Piano triennale per l'informatica nella pubblica amministrazione 2024-2026"
[PT agg. 2025]	D.P.C.M. 3 dicembre 2024 recante "Aggiornamento 2025 del Piano triennale 2024-2026"
[D.Lgs. 36/2006]	Decreto Legislativo 24 gennaio 2006, n. 36 recante "Attuazione della direttiva (UE) 2019/1024 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico che ha abrogato la direttiva 2003/98/CE".
[D.L. 135/2018]	Decreto-legge 14 dicembre 2018, n. 135 recante "Disposizioni urgenti in materia di sostegno e semplificazione per le imprese e per la Pubblica Amministrazione", convertito in legge, con modificazioni, dall'art. 1, comma 1 della legge 11 febbraio 2019, n. 12.
[Codice dei contratti]	Decreto legislativo 31 marzo 2023, n.36 e s.m.i recante "Codice dei contratti pubblici in attuazione dell'art. 1 della legge 21 giugno 2022, n.78, recante delega al Governo in materia di contratti pubblici".
[D.P.R. 81/2022]	Decreto del Presidente della Repubblica 24 giugno 2022, n. 81. Regolamento recante individuazione degli adempimenti relativi ai piani assorbiti dal piano integrato di attività e organizzazione.
[L. 90/2024]	Legge 28 giugno 2024, n. 90. Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici.
[D.Lgs. 134/2024]	Decreto Legislativo 4 settembre 2024, n. 134. Attuazione della direttiva (UE) 2022/2557 (NIS2) del Parlamento europeo e del Consiglio, del 14 dicembre 2022, relativa alla resilienza dei soggetti critici e che abroga la direttiva 2008/114/CE del Consiglio.
[Digital Decade]	Dichiarazione europea sui diritti e i principi digitali per il decennio digitale (2023/C 23/01)
[Strategia IA]	Strategia Italiana per l'Intelligenza Artificiale 2024-2026.

2.4. Linee Guida di riferimento

Di seguito sono elencate le linee guida emesse da AgID ai sensi dell'art. 71 del CAD e altra documentazione regolamentare, che sono richiamate, anche indirettamente, nel presente documento. Le linee guida AgID sono disponibili tramite il sito istituzionale al seguente indirizzo: https://www.agid.gov.it/it/linee-

guida, dove sono pubblicati anche i relativi aggiornamenti in conseguenza dell'evoluzione tecnologica o della necessità di adeguamento alla normativa di riferimento.

[LLGG_IA]	Linee guida per l'adozione, l'acquisto, lo sviluppo di sistemi di intelligenza artificiale nella Pubblica Amministrazione
[LLGG_DES]	Linee guida di design per i siti internet e i servizi digitali della Pubblica Amministrazione
[LG_DOC_INF]	Linee guida sulla formazione, gestione e conservazione dei documenti informatici
[LG_PDND_INTER]	Linee guida sull'infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l'interoperabilità dei sistemi informativi delle basi dati
[LG_SIC_INTER]	Linee guida Tecnologie e standard per la sicurezza dell'interoperabilità tramite API dei sistemi informatici
[LG_INTER_TEC]	Linee guida sull'interoperabilità tecnica delle Pubbliche Amministrazioni
[ST_eDGUE]	Specifiche tecniche per la definizione del DGUE elettronico italiano "eDGUE-IT"
[FICEP]	Progetto FICEP - nodo eIDAS italiano – Avviso n. 1-2018 – Note per il dispiegamento del LOGIN eIDAS presso le Pubbliche Amministrazioni
[DET_CLOUD]	Decreto direttoriale ACN prot. n. 5489 del 08 febbraio 2023 per la transizione di infrastrutture e servizi digitali
[QUAL_CLOUD]	Decreto direttoriale ACN prot. n. 29 del 02 gennaio 2023 sulla qualificazione del cloud nella PA
[REG_CLOUD]	Regolamento recante i livelli minimi di sicurezza, capacità elaborativa, risparmio energetico e affidabilità delle infrastrutture digitali per la Pubblica Amministrazione e le caratteristiche di qualità, sicurezza, performance e scalabilità, portabilità dei servizi cloud per la Pubblica Amministrazione, le modalità di migrazione nonché le modalità di qualificazione dei servizi cloud per la pubblica amministrazione
[LG_OPENDATA]	Linee guida Open Data. Linee guida recanti regole tecniche per l'attuazione del decreto legislativo 24 gennaio 2006, n. 36 e s.m.i. relativo all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico
[LG_ACCESS]	Linee guida sull'accessibilità degli strumenti informatici
[LG_SITI]	Linee guida di design per i siti internet e i servizi digitali della PA
[LG_GPDP]	Linee guida in materia di trattamento di dati personali, contenuti anche in atti e documenti amministrativi, effettuato per finalità di pubblicità e trasparenza sul web da soggetti pubblici e da altri enti obbligati, adottate con provvedimento del Garante per la protezione dei dati personali n. 243 del 15 maggio 2014
[LG-RIUSO]	Linee guida su acquisizione e riuso di software per le Pubbliche Amministrazioni.
[HLEG]	Orientamenti etici per un'IA affidabile.

[AI_ACT_PROHIB]	C(2025) 884 Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act).

2.5. Acronimi

Di seguito si riportano gli acronimi utilizzati nelle presenti Linee Guida.

[PA]	Tutti soggetti di cui all'articolo 2, comma 2, del CAD.
[DPIA]	Valutazione d'impatto sulla protezione dei dati.
[FRIA]	Valutazione d'impatto sui diritti fondamentali.
[GPAI]	General-Purpose AI System.
[KPI]	Key Performance Indicator, indicatore chiave di prestazione.
[PIAO]	Piano integrato di attività e organizzazione di cui al D.P.R. 81/2022.
[PNRR]	Piano nazionale di ripresa e resilienza.
[POC]	Proof of Concept, prova di concetto sperimentale.
[QoS]	Quality of Service, parametri usati per caratterizzare la qualità degli e-service.
[RPD]	Responsabile della protezione dei dati.
[RTD]	Responsabile per la transizione al digitale.
[SLA]	Service Level Agreement, accordo sul livello di servizio frutto della contrattazione tra erogatore e fruitore.
[SLI]	Service-Level Indicator, metrica atta a misurare l'efficienza dei servizi individuati dall'erogatore.
[SLO]	Service-Level Objective, obiettivi degli SLI per i servizi definiti dall'erogatore.
[UTD]	Ufficio per la transizione al digitale.

3. L'Intelligenza Artificiale

L'Intelligenza Artificiale (IA) è un insieme di tecnologie in grado di trasformare e potenziare attività economiche e sociali, migliorando i processi decisionali, l'efficienza operativa e la qualità dei servizi offerti alle organizzazioni e agli individui.

Con "sistema di IA" si intende *"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"*[1].

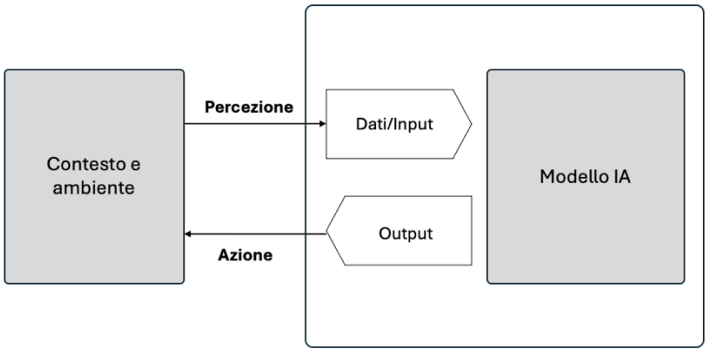

Figura 1 – Sistema di IA[2]

Una caratteristica distintiva dei sistemi di IA è la capacità inferenziale[3] della sua componente principale, il modello di IA[4]; tale componente, addestrata su grandi quantità di dati, consente al sistema di generare output quali previsioni, contenuti, raccomandazioni o decisioni a partire dai dati ricevuti in input. Questa capacità va oltre l'elaborazione semplice o meccanica dei dati, permettendo al sistema di:

- apprendere dai dati: utilizzare tecniche di apprendimento automatico (*machine learning*) per identificare modelli nei dati, acquisire conoscenze e raggiungere obiettivi specifici;
- ragionare e dedurre: applicare approcci basati sulla logica e sulla conoscenza, che permettono al sistema di trarre inferenze utilizzando regole predefinite, rappresentazioni simboliche o conoscenze codificate;
- modellizzare: creare rappresentazioni astratte del problema o del contesto per supportare processi decisionali complessi o fornire soluzioni innovative.

[1] AI Act art. 3 definizione 1).
[2] L'immagine è tratta da OECD AI Principles overview https://oecd.ai/en/ai-principles.
[3] L'inferenza è il processo di derivare nuove informazioni, conclusioni o previsioni a partire da dati, regole o conoscenze preesistenti. Nel contesto dell'IA, si riferisce alla capacità di un modello o sistema di elaborare dati in ingresso per generare risultati significativi, come classificazioni, raccomandazioni, decisioni o previsioni. Vedi: ISO/IEC 22989:2022 - Artificial Intelligence - Concepts and terminology https://www.iso.org/standard/74296.html.
[4] La ISO/IEC 22989:2022 definisce modello come "rappresentazione fisica, matematica o altrimenti logica di un sistema, entità, fenomeno, processo o dato". I modelli di IA includono, tra gli altri, modelli statistici e vari tipi di funzioni di input-output (come alberi decisionali e reti neurali).

La capacità inferenziale consente ai sistemi di IA di andare oltre l'elaborazione statica dei dati (come i sistemi software tradizionali), adattandosi e migliorandosi in base ai risultati ottenuti e alle condizioni operative. In altre parole, i sistemi di IA non si limitano a eseguire istruzioni predefinite, ma imparano e migliorano nel tempo, ampliando la loro utilità e applicazione in contesti complessi.

Gli obiettivi di un sistema di IA possono differire dalla finalità specifica per cui viene utilizzato in un determinato contesto; pertanto, un sistema di IA può essere applicato in modo diverso a seconda del caso d'uso o del contesto operativo.

I sistemi di IA possono operare con livelli di autonomia variabili, che spaziano da un controllo umano diretto a un funzionamento indipendente.

L'adattabilità è una caratteristica chiave dei sistemi di IA: grazie all'autoapprendimento, i sistemi di IA possono modificare il proprio comportamento durante l'uso, migliorando le loro prestazioni nel tempo.

Un sistema di GPAI si basa su un modello di IA per finalità generali e possiede la capacità di essere utilizzato per molteplici scopi. Un modello di IA per finalità generali possiede un'elevata capacità di generalizzazione, ed è addestrato su grandi volumi di dati spesso utilizzando tecniche di auto-apprendimento su larga scala. Questo tipo di modello è in grado di eseguire una vasta gamma di compiti distinti e può essere integrato in diversi sistemi o applicazioni.

I sistemi di IA posso essere soggetti a *bias*[5], cioè distorsioni o pregiudizi che possono manifestarsi nei risultati o nei comportamenti di un sistema di IA a causa di diversi fattori, come:

- *bias* nei dati: quando i dati utilizzati per addestrare o testare il modello di IA utilizzato riflettono pregiudizi, incompletezze o rappresentazioni distorte della realtà, influenzando negativamente l'equità e l'accuratezza del sistema;
- *bias* algoritmico: quando le scelte tecniche o le ipotesi implementate nei modelli IA o negli algoritmi utilizzati introducono distorsioni nei risultati;
- *bias* umano: quando pregiudizi impliciti o espliciti dei progettisti o degli stakeholder si riflettono nella progettazione, nell'addestramento o nell'applicazione del sistema di IA.

Il *bias* può portare a trattamenti ingiusti, risultati inaffidabili o discriminazioni nei confronti di specifici individui o gruppi. Riconoscere, mitigare e monitorare i *bias* è essenziale per garantire che i sistemi di IA operino in modo trasparente ed equo.

L'adozione dell'IA consiste nell'integrazione strategica, organizzativa e operativa di tecnologie, sistemi e applicazioni basati su IA all'interno dei processi, dei servizi e delle attività di un'organizzazione.

[5] ISO/IEC TR 24027 Artificial intelligence - Bias in AI systems and AI aided decision making https://www.iso.org/standard/77607.html

3.1. Ciclo di vita

Le presenti linee guida adottano come modello di riferimento il ciclo di vita di un sistema di IA definito dall'OECD, che descrive le fasi dei sistemi dalla pianificazione iniziale alla dismissione.

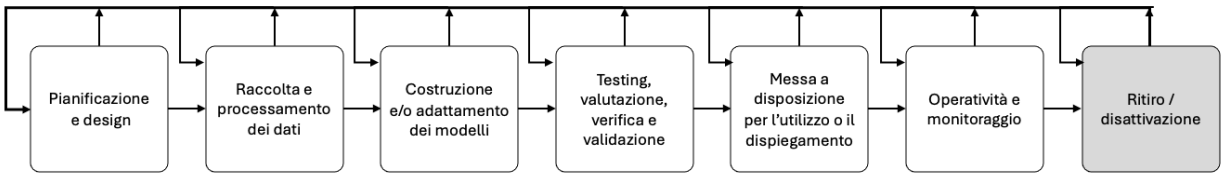

Figura 2 – Ciclo di vita di un sistema di IA[6]

- **Pianificazione e design:** l'organizzazione individua l'obiettivo che il sistema di IA si prefigge di raggiungere comprendendo il contesto nel quale il sistema dovrà operare e i dati richiesti. La fase comprende la definizione dei bisogni e requisiti degli stakeholder, la traduzione di questi ultimi in requisiti tecnici misurabili, la progettazione dell'architettura e del design del sistema.
- **Raccolta e processamento dei dati:** i dati (in forma strutturata e non) e i corrispondenti metadati sono acquisiti dalle varie sorgenti, analizzati in modo da verificare se si adattano a distribuzioni statistiche al fine di stimarne i parametri corrispondenti e ripuliti, integrati e trasformati secondo le esigenze. Tali operazioni devono seguire le indicazioni incluse nel paragrafo sulla gestione e la qualità dei dati. Le caratteristiche più rilevanti dell'insieme di dati (*dataset*) sono identificate e selezionate, scartando quelle non di interesse, con l'obiettivo di ridurre le dimensioni del dataset.
- **Costruzione e/o addestramento dei modelli:** l'organizzazione costruisce o seleziona un modello di IA adeguato al raggiungimento dell'obiettivo del sistema e ai dati disponibili. Scegliere il modello implica anche la scelta degli algoritmi di apprendimento[7] con il quale sarà effettuato l'addestramento. Successivamente il modello di IA viene addestrato sulla base dei dati e dell'algoritmo di apprendimento individuato. Segue la fase di *tuning* del modello, durante la quale il modello viene messo a punto regolando gli iperparametri[8] sulla base di un *dataset* di convalida. In alternativa l'organizzazione può scegliere l'approccio del "trasferimento dell'apprendimento", acquisendo un modello di IA già addestrato e configurato, da utilizzare come punto di partenza per l'ulteriore addestramento. Questo approccio può essere scelto quando l'organizzazione non dispone di dati sufficienti per l'addestramento.
- **Testing valutazione, verifica e validazione**: il sistema di IA è sottoposto a una serie di test rigorosi per garantire che soddisfi i requisiti e le aspettative degli stakeholder. Durante questa fase vengono

[6] OECD AI Principles overview, disponibile al seguente link: https://oecd.ai/en/ai-principles.
[7] Le tre categorie più comuni di algoritmi sono l'apprendimento supervisionato, non supervisionato e con rinforzo.
[8] Parametri che consentono di controllare il processo di addestramento del modello.

identificate e risolte eventuali anomalie, mentre si verifica che il modello operi correttamente nel contesto previsto e rispetti i criteri di conformità e qualità.

- **Messa a disposizione per l'utilizzo o il dispiegamento**: il sistema di IA addestrato viene reso disponibile agli utenti e integrato nei processi operativi, garantendo la compatibilità con altri sistemi e l'accessibilità per tutti gli utenti previsti.
- **Operatività e monitoraggio:** il sistema di IA e i dati in ingresso sono monitorati continuamente per rilevare eventuali variazioni nei pattern operativi, anomalie nei dati o insorgenza di bias che possano compromettere l'affidabilità delle decisioni. L'organizzazione effettua attività di manutenzione, riaddestramento e ottimizzazione del modello, valutando costantemente il raggiungimento degli obiettivi prefissati e l'efficacia operativa.
- **Ritiro/Disattivazione:** quando il sistema di IA non è più necessario, viene ritirato o disattivato in modo controllato. Questa fase include la gestione sicura dei dati, la documentazione delle operazioni finali e la pianificazione di eventuali sostituzioni o alternative.

3.2. I ruoli nella catena del valore dell'IA

La catena del valore dell'IA coinvolge diversi soggetti che ricoprono ruoli fondamentali nelle fasi di sviluppo, distribuzione e utilizzo delle tecnologie di IA. Le presenti linee guida adottano le definizioni di fornitore *(provider)* e *deployer* fornite dall'AI Act[9] nel contesto generale dei sistemi di IA. Per una descrizione dettagliata degli obblighi specifici applicabili a fornitore e *deployer* in conformità con l'AI Act, si faccia riferimento al capitolo 5.

- Fornitore: soggetto che sviluppa un sistema di IA o un modello GPAI o che fa sviluppare un sistema di IA o un modello GPAI e immette tale sistema o modello sul mercato o mette in servizio il sistema di IA con il proprio nome o marchio;
- *Deployer*: soggetto che utilizza un sistema di IA, anche integrandolo nei propri sistemi, senza modificarne in modo significativo il funzionamento. Se un *deployer* modifica in modo significativo il sistema o lo utilizza sotto il proprio nome o marchio, assume le responsabilità del fornitore.

[9] AI Act art. 3 definizioni 3) e 4).

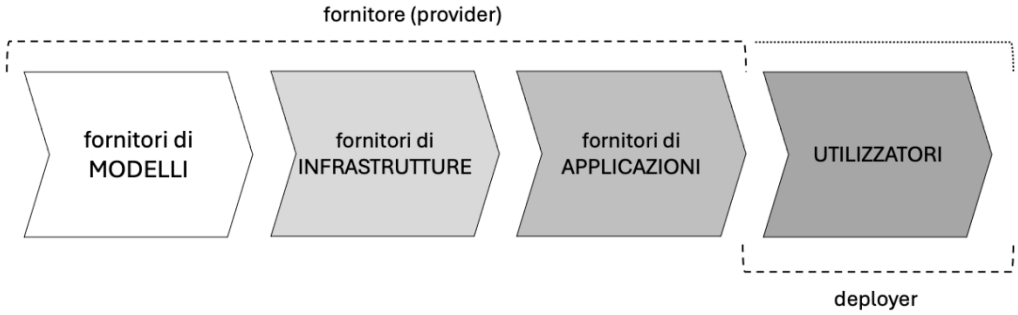

Figura 3 – Catena del valore dell'IA

La catena del valore dell'IA rappresentata in Figura 3 evidenzia quattro livelli principali:

- **Fornitori di modelli**

  Soggetti che sviluppano e/o forniscono modelli di IA, modelli di IA per finalità generali (GPAI).
- **Fornitori di infrastrutture**

  Soggetti che sviluppano e/o forniscono componenti infrastrutturali per l'IA quali gestione dei dati, networking, hardware, servizi cloud, piattaforme MLOps. Questi soggetti rappresentano il cuore operativo che supporta l'intero ciclo di vita dell'IA.
- **Fornitori di applicazioni**

  Soggetti che sviluppano e/o forniscono applicazioni per scopi specifici adattando i modelli di IA e GPAI. Questi attori possiedono una conoscenza approfondita dei settori e delle esigenze delle organizzazioni utilizzatrici, combinando le capacità dell'IA con competenze di dominio per favorire la personalizzazione delle soluzioni.
- **Utilizzatori**

  Soggetti che adottano e integrano le tecnologie di IA nei propri sistemi per soddisfare esigenze operative e strategiche. Questi soggetti possono effettuare personalizzazioni o fine-tuning dei modelli utilizzando i propri dati al fine di ottimizzare le soluzioni per il contesto operativo specifico.

3.3. Classificazione dei sistemi di IA sulla base del rischio

L'IA offre benefici in settori chiave come sanità, educazione, trasporti, energia e pubblica amministrazione, ma può comportare rischi per gli interessi pubblici e i diritti fondamentali tutelati dal diritto dell'Unione. Le presenti Linee Guida adottano la classificazione del rischio definita dall'AI Act (cfr. cap. 5). In Figura 4 sono sinteticamente riportati, per ciascun livello di rischio, alcuni esempi di applicazione e gli obblighi definiti dall'AI Act.

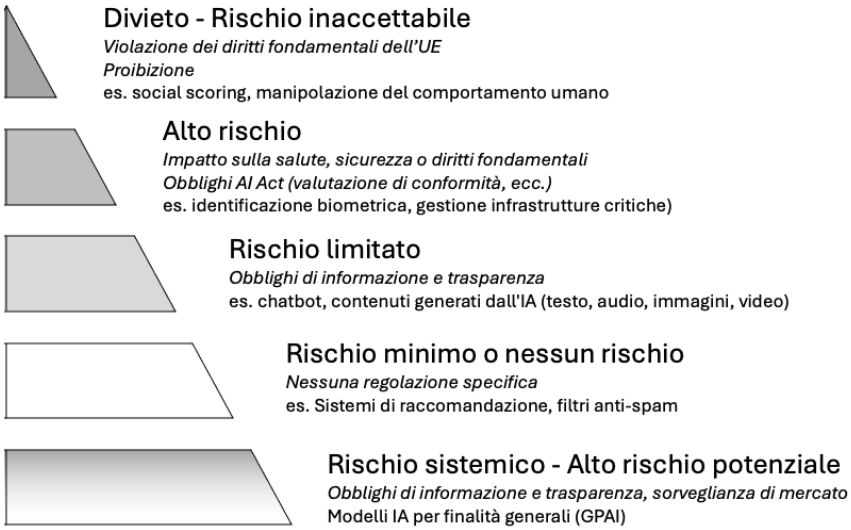

Figura 4 – Classi di rischio dei sistemi secondo l'AI Act

Le PA DEVONO tenere in considerazione la classificazione dei sistemi di IA in base ai livelli di rischio delineati dall'AI Act:

- **Sistemi di IA vietati**[10]*:* sono i sistemi dal rischio inaccettabile e, pertanto, vietati. L'art. 5 dell'AI Act fornisce un'elencazione delle IA vietate (per esempio, IA che utilizzano tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli volte a distorcere il comportamento delle persone; meccanismi di sfruttamento delle persone vulnerabili; sistemi di IA che introducano meccanismi di s*coring sociale).*
- **Sistemi di IA ad alto rischio:** sono i sistemi che pongono elevati rischi. In caso di malfunzionamento, tali sistemi rappresentano una potenziale rilevante dannosità, con elevata compromissione di interessi pubblici e diritti fondamentali. Per qualificare un sistema come ad alto rischio, le PA DEVONO fare riferimento al sistema di classificazione delineato dall'AI Act (si veda, in particolare, l'art. 6 e l'Allegato III). Le PA DOVREBBERO, inoltre, considerare che alla Commissione europea è conferito il potere di adottare atti delegati aggiungendo o modificando i casi d'uso dei sistemi di IA ad alto rischio di cui all'Allegato III all'AI Act.
- **Sistemi di IA a rischio limitato**: sono sistemi dal rischio minore. In caso di malfunzionamento, si contraddistinguono per una potenziale dannosità limitata (per esempio, *chatbot* generici), con un impatto limitato su interessi pubblici e diritti fondamentali.

[10] Gli obblighi relativi agli usi vietati dell'IA di cui all'art. 5 dell'AI Act sono entrati in vigore il 2 febbraio 2025. La Commissione europea ha pubblicato apposite Linee guida relative alla definizione dei sistemi di IA e all'attuazione delle pratiche di IA che comportano rischi inaccettabili ai sensi dell'AI Act *"Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)"* https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

- **Sistemi di IA a rischio minimo o nullo***:* sono sistemi dal rischio irrilevante. In caso di malfunzionamento, la relativa dannosità è molto limitata, potendo avere un impatto del tutto trascurabile su interessi pubblici e diritti fondamentali.

Per i sistemi IA ad alto rischio, oltre agli obblighi disposti dall'AI Act, al fine di garantire un livello costante ed elevato di tutela degli interessi pubblici in materia di salute, sicurezza e diritti fondamentali, le PA DEVONO:

- rispettare i principi della **Carta dei diritti fondamentali dell'UE**, valutando l'impatto dei sistemi di IA sui diritti tutelati;
- considerare la **Dichiarazione europea sui diritti e i principi digitale**[11] per il *Digital Decade*, assicurando l'allineamento delle applicazioni IA ai valori europei;
- seguire gli **Orientamenti etici per un'IA affidabile** dell'AI HLEG[12], promuovendo trasparenza, equità e responsabilità nell'uso dell'IA.

La PA DEVE tenere in considerazione che il Capo V dell'AI Act definisce una regolamentazione specifica per i **modelli di IA per finalità generali con rischio sistemico**, che possono avere effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso.

3.4. Principi per l'adozione dell'IA nella pubblica amministrazione

Le PA adottano l'IA al fine di:

- automatizzare attività semplici e ripetitive di ricerca e analisi delle informazioni, liberando tempo di lavoro per attività a maggior valore;
- aumentare le capacità predittive, migliorando il processo decisionale basato sui dati;
- supportare la personalizzazione dei servizi incentrata sull'utente, aumentando l'efficacia, l'efficienza e la tempestività dei servizi pubblici anche attraverso meccanismi di proattività;
- promuovere l'innovazione dei servizi pubblici e dei processi amministrativi.

Le PA adottano i sistemi di IA applicando i seguenti principi:

Conformità e governance

P.1 **Conformità normativa.** Le PA adottano i sistemi di IA nel pieno rispetto delle normative nazionali e dell'Unione Europea, assicurando l'aderenza al quadro legislativo vigente. Le PA monitorano l'evoluzione normativa e aggiornano costantemente i propri sistemi per assicurarne la conformità alle disposizioni vigenti.

[11] Dichiarazione europea sui diritti e principi digitali https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:JOC_2023_023_R_0001.
[12] Orientamenti etici per un'IA affidabile https://data.europa.eu/doi/10.2759/640340

P.2 **Rispetto dei valori fondamentali dell'UE.** Le PA adottano i sistemi di IA garantendo i principi definiti dalla Carta dei diritti fondamentali dell'Unione Europea.

P.3 **Gestione del rischio.** Le PA adottano adeguate politiche di gestione del rischio conducendo un'analisi approfondita dei rischi associati all'impiego di sistemi di IA, al fine di prevenire violazioni dei diritti fondamentali o altri danni rilevanti.

P.4 **Protezione dei dati personali.** Le PA adottano i sistemi di IA nel rispetto delle norme unionali e nazionali vigenti in materia di protezione dei dati personali, garantendo elevata qualità e integrità dei dati stessi (cfr. cap. 10).

Etica e inclusione

P.5 **Responsabilità.** Le PA adottano l'IA come strumento di supporto all'attività umana consapevoli che la responsabilità ultima delle decisioni adottate, in modo automatico o supervisionato, dai sistemi di IA rimane in capo alla PA. Le PA identificano chiaramente le responsabilità di tutti gli attori coinvolti nel ciclo di vita dei sistemi di IA.

P.6 **Accessibilità, inclusività, non discriminazione.** Le PA assicurano un trattamento equo per tutti i soggetti e gruppi coinvolti nell'adozione dell'IA, promuovendo la parità di accesso, l'uguaglianza di genere e la diversità culturale. Le PA adottano misure preventive per evitare la riproduzione o l'amplificazione di *bias* presenti nella società.

P.7 **Trasparenza.** le PA adottano sistemi di AI garantendo la trasparenza e comprensibilità delle loro decisioni e del funzionamento. Le PA garantiscono una adeguata spiegabilità dei risultati, rendendo comprensibili le motivazioni che supportano le decisioni e le azioni intraprese dai sistemi di IA.

P.8 **Informazione.** Le PA informano gli utenti sull'interazione con sistemi di IA, rendendoli consapevoli delle capacità e dei limiti di tali sistemi.

Qualità e affidabilità dei sistemi di IA

P.9 **Qualità dei dati.** Le PA adottano sistemi di IA garantendo una gestione etica, trasparente e conforme dei dati. Questo include l'utilizzo di dati di qualità elevata, affidabili e aggiornati, la definizione di politiche chiare per l'accesso, l'uso e la conservazione dei dati, l'adozione di misure tecniche e organizzative per tutelare l'integrità e la sicurezza dei dati istituzionali, assicurando che i dati siano gestiti in modo sostenibile e responsabile.

P.10 **Affidabilità**. Le PA garantiscono che i risultati e le decisioni prodotte dai sistemi di IA utilizzati siano affidabili e coerenti con gli obiettivi prefissati. Questo include il monitoraggio continuo delle prestazioni per identificare e correggere eventuali errori o deviazioni dai parametri stabiliti.

P.11 **Robustezza**. Le PA assicurano la robustezza dei sistemi di IA, garantendo che essi siano in grado di operare correttamente anche in condizioni avverse o in presenza di guasti. La robustezza include la capacità di mantenere le prestazioni desiderate nonostante imprevisti, errori nei dati o problemi tecnici.

P.12 **Sicurezza cibernetica.** Le PA garantiscono la sicurezza cibernetica dei sistemi di IA, prevenendo tentativi di alterazione, compromissione o uso illecito da parte di terzi. Questo comprende l'adozione

di misure di protezione adeguate a salvaguardare l'integrità, la disponibilità e la riservatezza dei dati e dei processi gestiti dai sistemi di IA.

P.13 **Supervisione umana.** Le PA garantiscono un livello adeguato di supervisione umana dei sistemi di IA. Ai fini della supervisione umana, le PA assicurano che i sistemi di IA siano progettati e implementati in modo tale da consentirne la verifica, correzione o sostituzione da parte di personale umano.

P.14 **Registrazioni (logging).** le PA adottano sistemi di IA dotati di adeguati meccanismi di registrazione necessari a tracciare e conservare nel tempo le operazioni svolte.

P.15 **Adozione di standard tecnici.** Le PA tengono in debita considerazione le norme tecniche definite a livello nazionale, europeo e internazionale al fine di garantire l'interoperabilità, la manutenibilità, la sicurezza e la conformità dei sistemi di IA alla normativa vigente.

Innovazione e sostenibilità

P.16 **Efficienza e qualità dei servizi.** Le PA adottano l'IA per incrementare l'efficienza operativa e migliorare la qualità dei servizi destinati a cittadini e imprese. Le PA utilizzano l'IA per automatizzare i compiti ripetitivi connessi ai servizi istituzionali e al funzionamento dell'apparato amministrativo, favorendo l'implementazione di soluzioni innovative e proattive che semplifichino l'accesso ai servizi e ne migliorino l'efficienza complessiva.

P.17 **Innovazione e miglioramento continuo.** Le PA adottano un approccio all'IA orientato all'innovazione e al miglioramento continuo, al fine di ottimizzare i processi amministrativi e la qualità dei servizi offerti a cittadini e imprese. Le PA mantengono rapporti di collaborazione con altre PA, enti di ricerca, università e imprese per sperimentare e integrare soluzioni tecnologiche basate sull'IA, creando un ecosistema favorevole all'innovazione e all'adozione di tecnologie digitali.

P.18 **Sostenibilità ambientale.** Le PA adottano sistemi di IA secondo un approccio sostenibile, attento alla tutela ambientale e in linea con i principi di sostenibilità energetica.

Formazione e organizzazione

P.19 **Formazione e sviluppo delle competenze.** Le PA investono nella formazione del proprio personale per garantire le competenze necessarie a utilizzare, gestire e sviluppare sistemi di IA in modo efficace e responsabile. Le PA promuovono iniziative di alfabetizzazione digitale rivolte a cittadini e imprese, al fine di favorire una comprensione diffusa delle opportunità e delle implicazioni dell'IA, assicurando che l'adozione dell'IA nei servizi pubblici avvenga in modo inclusivo e consapevole.

P.20 **Rafforzamento dell'organizzazione e delle infrastrutture.** Le PA, nell'adozione di sistemi di IA, ottimizzano il proprio assetto organizzativo e le proprie infrastrutture tecnologiche per agevolare il processo di trasformazione digitale e migliorare i livelli di sicurezza e resilienza.

Per i sistemi IA classificati ad alto rischio, i principi sopra enunciati sono rafforzati da specifici requisiti definiti dall'AI Act (cfr. cap. 5).

4. Modello di adozione dell'IA

Le PA DOVREBBERO adottare un insieme strutturato di processi, politiche, risorse e strumenti per governare, implementare, monitorare e migliorare l'utilizzo dei sistemi di IA durante il loro ciclo di vita.

Le PA DOVREBBERO implementare un modello di adozione dell'IA in grado di rispondere prontamente ai cambiamenti del contesto normativo e tecnologico.

Le presenti Linee guida propongono un modello di adozione dell'IA basato sul ciclo per il miglioramento continuo Plan-Do-Check-Act (PDCA) e su alcune pratiche gestionali definite dallo standard ISO/IEC 42001:2023[13]. La Figura 5 riporta una rappresentazione sintetica del modello, approfondito nei successivi paragrafi.

Le PA POSSONO adottare tale modello adattandolo alle proprie caratteristiche, esigenze e responsabilità, con particolare riferimento al ruolo di fornitore o *deployer* di sistemi di IA. In particolare, le PA POSSONO, per l'adozione di sistemi di IA di rischio limitato, minimo o nullo, accorpare le fasi del modello, adeguandole alla bassa complessità del progetto di adozione.

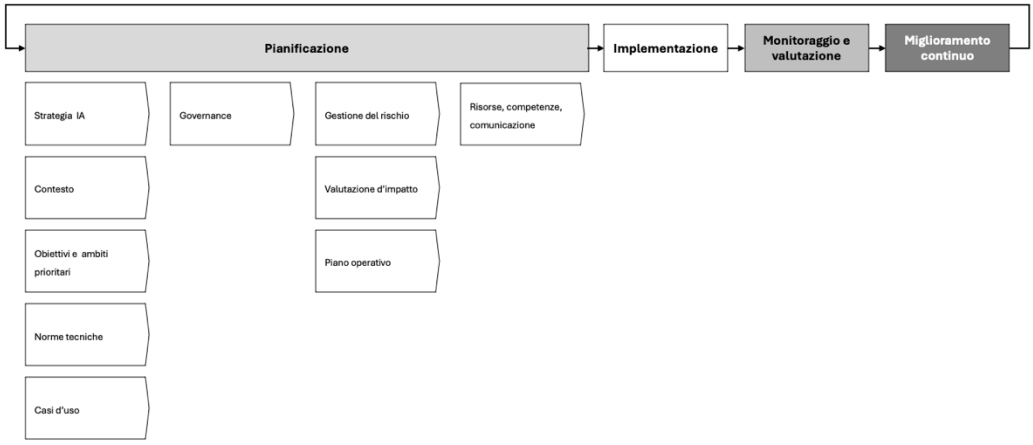

Figura 5. Rappresentazione delle fasi del modello di adozione dell'IA.

4.1. Strategia per l'IA

La PA DEVE sviluppare una strategia per l'IA coerente con il proprio contesto (cfr. par. 4.2) e allineata alla propria missione in termini di funzione amministrativa, giurisdizione e ambiti operativi specifici.

Le PA POSSONO definire una strategia comune in base alle proprie caratteristiche e tipologia (ad esempio comuni o università).

La strategia DEVE contenere gli obiettivi dell'utilizzo dell'IA (cfr. par. 4.3). e le azioni per conseguirli, promuovendo un approccio condiviso e collaborativo che coinvolga tutte le funzioni della PA.

Le PA includono nella strategia per l'IA le azioni per:

[13] ISO/IEC 42001:2023 Information technology - Artificial intelligence - Management system https://www.iso.org/standard/81230.html

- migliorare la qualità dei dati, ivi comprese le basi documentali (cfr. cap. 9);
- migliorare le competenze del personale interno in materia di IA e di protezione dei dati personali coinvolti (cfr. cap. 8);
- individuare i casi d'uso che possono apportare maggiore beneficio (cfr. 4.5);
- valutare le esperienze e sperimentazioni già effettuate da altre PA;
- avviare sperimentazioni, anche in forma associata con altre PA, partendo dai casi d'uso di minore complessità (*quick win*);
- avviare collaborazioni con altre PA per la sperimentazione e successivamente l'adozione, acquisto e sviluppo dei sistemi di IA in forma associata.

La strategia per l'IA DEVE essere coerente con la strategia definita nel PIAO, la strategia sui dati e il Piano triennale per l'informatica nella Pubblica Amministrazione. Ciò in modo da delineare una strategia completa e coerente con la politica complessiva sulle tecnologie della PA, che coinvolga anche l'eventuale territorio di riferimento.

La PA DEVE implementare la propria strategia dell'IA affidandone il coordinamento, unitamente alla strategia sui dati, al Responsabile per la transizione al digitale (RTD) e al suo ufficio (UTD), come previsto dal piano triennale per l'informatica nella PA 2024-2026, coinvolgendo altresì il Responsabile della protezione dei dati (RPD).

4.2. Analisi del contesto e delle caratteristiche della PA

Le PA DEVONO analizzare i fattori esterni e interni che influenzano le capacità dell'ente di conseguire i risultati attesi dall'uso dei sistemi di IA, con l'obiettivo di individuare una strategia e un modello di adozione adeguati. Le PA POSSONO definire strategia e modello in collaborazione con altre amministrazioni o adottando strategie e modelli definiti da enti sovraordinati o della stessa tipologia.

I fattori esterni includono i requisiti normativi definiti da CAD, AI Act, DGA, GDPR e NIS2, oltre agli eventuali requisiti espressi dagli stakeholder esterni (cittadinanza, imprese, altre PA).

I fattori interni comprendono i requisiti relativi a: struttura organizzativa (incluse le dimensioni dell'ente), stakeholder interni (es. personale dell'ente), ambiti operativi specifici, contesto territoriale e capacità tecnologiche. Queste ultime comprendono la disponibilità di dati, infrastrutture e competenze necessarie per l'implementazione e la gestione efficace dell'IA.

Sulla base delle analisi condotte a livello internazionale da organismi quali EU JRC[14] e OECD[15] è stato definito uno strumento pratico (cfr. Allegato A) per determinare il livello di maturità organizzativa e tecnologica delle PA.

Dall'analisi del contesto e delle caratteristiche della PA derivano le azioni della strategia per l'IA (cfr. par. 4.1) con particolare riferimento al:

- miglioramento della qualità del patrimonio della PA in termini di dati e documenti digitali;
- rafforzamento delle competenze;
- definizione di collaborazioni con altre PA per la sperimentazione e la gestione associata dell'IA.

L'analisi del contesto fornisce gli elementi per valutare la fattibilità degli obiettivi e individuare i casi d'uso in cui l'IA può essere impiegata in modo efficace, definendone le modalità di implementazione.

4.3. Obiettivi e ambiti prioritari di applicazione

Le PA DEVONO adottare le tecnologie di IA indentificando in via preliminare gli obiettivi e gli ambiti prioritari di applicazione sulla base del proprio contesto, pur potendo le PA condurre sperimentazioni di uso dell'IA in ambiti non prioritari e meno soggetti a rischi.

Le PA DEVONO verificare le proprie esigenze in modo dettagliato e documentato, con l'obiettivo di identificare i casi d'uso in cui l'IA offre il massimo beneficio, in termini di miglioramento dell'efficienza operativa e dell'erogazione dei servizi.

La verifica delle esigenze specifiche contribuisce all'eventuale studio di fattibilità e alla valutazione comparativa necessaria per la scelta delle soluzioni di IA, come previsto dall'art. 68 del CAD.

Gli ambiti prioritari per l'utilizzo dell'IA da parte della PA deriva da un lavoro di ricognizione e analisi condotto su progetti e sperimentazioni in corso a livello nazionale, europeo e internazionale[16]. Gli ambiti prioritari individuati in tale ricognizione sono:

1. **Miglioramento dell'efficienza operativa**: le PA POSSONO utilizzare l'IA per aumentare la propria capacità di analisi e gestione dei dati e di automatizzazione dei processi ripetitivi, al fine di semplificare i processi interni, ridurre i tempi operativi e migliorare l'efficienza complessiva. In particolare, le aree in cui si individuano i maggiori benefici sono:

[14] EU JRC AI Watch, AI Watch "Road to the Adoption of Artificial Intelligence by the Public Sector", AI Watch "European landscape on the use of artificial intelligence by the public sector. Annex II, Case studies description"
[15] OECD Observatory of Public Sector Innovation
[16] L'Agenzia per l'Italia digitale ha avviato a settembre 2024 la ricognizione dei progetti di IA nella PA e delle Banche dati strategiche ai fini dell'IA come previsto dal Piano Triennale per la PA 2024-2026 (cfr. RA5.4.4 - Realizzazione di applicazioni di IA a valenza nazionale, RA5.5.1 - Basi di dati nazionali strategiche). Le indicazioni del presente e dei successivi paragrafi sono basate sui risultati della suddetta ricognizione e delle analisi svolte dall'Osservatorio Agenda Digitale del Politecnico di Milano e dal JRC AI Watch sui progetti di utilizzo dell'IA nella PA nel contesto italiano, europeo ed internazionale nell'anno 2024.

a. **Supporto alle decisioni**: le PA impiegano l'IA per sviluppare modelli predittivi che consentano di adottare decisioni consapevoli e basate sui dati reali, aumentando l'affidabilità e la tempestività delle decisioni.
b. **Ottimizzazione dell'allocazione delle risorse**: le PA utilizzano l'IA per distribuire le risorse in modo più efficiente, individuando le priorità e focalizzandosi sulle aree di maggiore necessità, ottimizzando così l'uso delle risorse pubbliche.
c. **Miglioramento della gestione documentale**: le PA utilizzano l'IA per automatizzare la classificazione, l'archiviazione e il recupero dei documenti, facilitando la ricerca e riducendo i tempi di gestione.
d. **Miglioramento del supporto giuridico:** l'IA supporta le PA nell'analisi normativa e giurisprudenziale, consentendo di elaborare pareri legali più accurati e tempestivi e di monitorare aggiornamenti legislativi rilevanti.
e. **Miglioramento delle procedure di acquisto**: le PA adottano l'IA per ottimizzare le procedure di procurement, migliorando sia l'efficienza che la trasparenza del processo di acquisto.

2. **Miglioramento dei servizi ai cittadini e alle imprese:** le PA POSSONO utilizzare l'IA per aumentare la propria capacità di analisi e gestione dei dati al fine di personalizzare i servizi digitali in base alle specifiche esigenze degli utenti, anche in logica proattiva. In particolare, le aree in cui si individuano i maggiori benefici sono:
   a. **Personalizzazione:** le PA impiegano l'IA per adattare i servizi pubblici alle esigenze specifiche di cittadini e imprese, migliorando l'interazione digitale e l'efficienza nella risposta.
   b. **Proattività:** l'IA consente alle PA di anticipare le esigenze degli utenti, fornendo servizi o informazioni pertinenti prima che siano richiesti, semplificando così l'accesso e riducendo i tempi di attesa.
   c. **Trasparenza:** le PA utilizzano l'IA per migliorare la trasparenza, fornire a cittadini e imprese informazioni chiare e immediatamente fruibili sui propri adempimenti e sullo stato di avanzamento dei procedimenti amministrativi avviati presso la PA stessa.
   d. **Accessibilità:** le PA adottano soluzioni di IA per rendere i servizi pubblici accessibili e conformi all'art. 53 del CAD, garantendo l'usabilità delle piattaforme digitali anche a persone con disabilità o con limitate competenze digitali. In particolare, l'IA deve essere utilizzata come strumento di assistenza per la creazione e la gestione di contenuti nativamente accessibili.
   e. **Inclusione**: le PA adottano l'IA per analizzare le esigenze dei cittadini al fine di promuovere servizi destinati alle fasce deboli della popolazione, promuovendo l'inclusione sociale.
3. **Sicurezza e protezione dei dati**: le PA POSSONO utilizzare l'IA per migliorare la sicurezza dei dati e delle infrastrutture, identificando potenziali minacce e garantendo una protezione avanzata.

4.4. Norme tecniche

Nell'acquisizione, sviluppo e adozione dei sistemi di IA, le PA DEVONO tenere in considerazione le norme tecniche a livello nazionale, europeo e internazionale.

L'AI Act definisce i requisiti che devono essere soddisfatti dai sistemi di IA ad alto rischio (cfr. cap. 5). La Commissione Europea ha richiesto[17] agli organismi europei di standardizzazione CEN e CENELEC di sviluppare norme tecniche armonizzate[18] che definiscono approcci concreti per soddisfare tali requisiti. A partire dalla data di applicabilità dell'AI Act (2 agosto 2026[19]), la conformità dei sistemi di IA ad alto rischio dovrà essere garantita mediante un sistema di gestione della qualità e una valutazione di conformità prima dell'immissione sul mercato, entrambi basati sulle suddette norme tecniche.

A livello internazionale, le attività di normazione tecnica sull'IA sono coordinate in ambito ISO/IEC dal JTC 1/SC 42 "Artificial Intelligence". Pur considerando le specificità normative richieste per i sistemi di IA ad alto rischio, le norme tecniche sviluppate in ambito ISO costituiscono attualmente un valido riferimento per aspetti rilevanti dell'IA.

In Italia la UNI CT 533 "Intelligenza artificiale" presso UNINFO è la commissione tecnica dell'ente di normazione nazionale UNI deputata alla normazione in ambito IA ed è *mirror committee* del ISO/IEC JTC 1/SC 42 e CEN/CENELEC JTC 21.

Si rimanda all'allegato **Errore. L'origine riferimento non è stata trovata.** per un elenco più completo d elle principali norme tecniche sull'IA e per i riferimenti per la consultazione dello stato di avanzamento della normazione tecnica in ambito IA a livello nazionale, europeo e internazionale[20].

4.5. Casi d'uso

Considerato quanto detto ai paragrafi precedenti, il modello proposto per l'adozione dell'IA prevede che le PA individuino i casi d'uso in cui l'IA offre il massimo beneficio in termini di miglioramento dell'efficienza operativa e dell'erogazione dei servizi, sulla base della propria strategia per l'IA (cfr. par. 4.1), dell'analisi del contesto e delle caratteristiche dell'organizzazione (cfr. par. 4.2), tenuto conto degli obiettivi e gli ambiti prioritari di applicazione dell'IA (cfr. par. 4.3).

Le PA POSSONO comunque identificare, operando una semplificazione del modello, i casi d'uso specifici per sperimentazioni, progetti pilota e iniziative circoscritte a basso rischio.

[17] La Commissione Europea ha adottato il 25 maggio 2023 la Decisione C(2023)3215 - Standardisation request M/5932 con la quale ha affidato agli Enti di normazione europei CEN e CENELEC la redazione di norme tecniche europee per i sistemi di intelligenza artificiale conformi all'AI Act.
[18] L'attività di normazione tecnica condotta dal CEN/CENELEC Joint Technical Committee (JTC) 21 "Artificial Intelligence", prevede lo sviluppo di standard tecnici per i requisiti di: gestione del rischio, qualità e governance dei dati, logging e tracciabilità, documentazione tecnica, trasparenza, supervisione umana, accuratezza, robustezza e sicurezza cibernetica.
[19] Si veda l'art. 113 dell'AI Act. In ogni caso, lo stato di avanzamento dell'attività di normazione tecnica a supporto dell'AI Act è consultabile alla pagina web del CEN/CENELEC JTC 21 "Artificial Intelligence".
[20] L'elenco esaustivo delle norme tecniche pubblicate a livello nazionale, europeo ed internazionale e lo stato di avanzamento delle attività di normazione tecnica in corso è disponibile alla pagina della UNI/CT 533.

Una volta identificati i casi d'uso, le PA DEVONO raccogliere, documentare e aggiornare le informazioni fondamentali sugli stessi e sulle soluzioni di IA che li implementano, durante l'intero ciclo di vita di queste ultime, dall'ideazione alla prova di concetto sperimentale (POC), al rilascio in ambiente operativo, fino alla dismissione[21]. L'Allegato F riporta, a titolo di esempio, un modello di documentazione di un caso d'uso[22].

I casi d'uso POSSONO essere implementati in forma associata da più PA, anche attraverso gli spazi di sperimentazione e sviluppo previsti dal piano triennale per l'informatica nella PA.

Nel caso di sviluppo in forma associata, la gestione della documentazione è affidata alla PA capofila o a una delegata.

Le PA DEVONO trasmettere periodicamente all'AgID un estratto della documentazione descrittiva dei casi d'uso secondo le modalità definite nel Piano triennale per l'informatica nella pubblica amministrazione.

4.5.1. Funzionalità dell'IA

A partire dai casi d'uso individuati, le PA definiscono le funzionalità dei sistemi di IA che implementano i casi d'uso stessi.

La ricognizione sui progetti e le sperimentazioni a livello nazionale, europeo e internazionale già citata al precedente paragrafo 4.3 ha prodotto anche una classificazione non esaustiva di funzionalità di sistemi di IA. Detta classificazione è riportata nell'Allegato G.

4.5.2. Requisiti

Nelle fasi di sviluppo e gestione dei sistemi di IA, le PA DEVONO identificare i requisiti a partire dalle esigenze normative, tecniche, etiche e operative che devono essere soddisfatte, tenendo presente le aspettative delle parti interessate. Le PA DEVONO tenere in considerazione, oltre ai *Principi* indicati nelle presenti Linee guida (cfr. par 0), le seguenti categorie di requisiti:

- conformità all'AI Act (cfr. cap. 5)
- gestione e qualità dei dati (cfr. cap. 9)
- protezione dei dati personali (cfr. cap. 10)
- sicurezza cibernetica (cfr. cap. 10).

Nell'identificazione dei requisiti, le PA DEVONO tenere in considerazione anche il livello di rischio del sistema di IA da implementare: nel caso di sistemi ad alto rischio, si applicano i requisiti obbligatori previsti dell'AI Act.

[21] Per individuare il livello di maturità tecnologica del caso d'uso e delle soluzioni che lo implementano si suggerisce di utilizzare i livelli di maturità tecnologica secondo la ISO 16290:2013 Space system - Definition of the Technology Readiness Levels (TRLs) and their criteria of assessment

[22] Il modello per la raccolta delle informazioni sui casi d'uso è una versione adattata del template impiegato nella norma ISO/IEC TR 24030:2024, Artificial Intelligence (AI) - Use cases https://www.iso.org/obp/ui/en/#iso:std:iso-iec:tr:24030:ed-2:v1:en.

4.5.3. Indicatori di prestazione

Le PA DEVONO definire, per ciascun caso d'uso, indicatori di prestazione (Key Performance Indicator, KPI) che misurino l'efficacia del sistema di IA nel raggiungimento degli obiettivi prefissati.

I KPI POSSONO essere differenti a seconda del livello di maturità del sistema di IA.

I KPI DEVONO essere monitorati nel corso dell'evoluzione del caso d'uso il ciclo di vita consentendo di determinare l'evoluzione delle prestazioni del sistema di IA nelle sue diverse fasi di maturità.

I KPI POSSONO misurare sia prestazioni tecniche sia l'impatto sui processi aziendali e il valore aggiunto per la PA. I KPI possono riguardare:

- prestazioni del modello;
- qualità dei dati;
- robustezza e affidabilità;
- efficienza computazionale;
- usabilità e accessibilità;
- impatto etico e conformità normativa;
- costo e sostenibilità.

L'allegato I delle linee guida fornisce uno strumento operativo per l'individuazione dei KPI di un sistema di IA.

4.6. Governance

Gli organi direttivi della PA approvano la strategia per l'IA dell'ente.

Secondo quanto previsto dal Piano triennale per l'informatica nella pubblica amministrazione, la responsabilità della gestione dell'IA nelle PA deve essere affidata al RTD e al suo ufficio (UTD). Gli organi direttivi DEVONO garantire questa assegnazione e la devono comunicare all'interno della PA.

Nello specifico, al RTD è assegnata la responsabilità e l'autorità per:

- assicurare che i sistemi di IA siano conformi alle indicazioni delle presenti Linee guida;
- riferire agli organi direttivi sulle prestazioni e sulle attività di controllo, monitoraggio e evoluzione dei sistemi di IA.

La PA DOVREBBE definire procedure di governance per lo sviluppo e l'utilizzo dei sistemi di IA. Le eventuali procedure DEVONO essere coerenti con la strategia per IA dell'ente e con la strategia definita nel PIAO.

A titolo di esempio e per facilitare il compito delle PA, un elenco di possibili procedure di governance per l'IA è riportato nell'allegato G[23].

[23] JRC - Commissione Europea "Competences and governance practices for AI in the public sector" https://publications.jrc.ec.europa.eu/repository/handle/JRC138702.

Fermi restando gli obblighi di conformità all'AI Act, le PA DEVONO adottare un codice etico per l'IA. Tale codice DEVE divenire uno strumento di governance vincolante, allineato con il quadro normativo vigente, integrato nei processi decisionali e operativi della PA, finalizzato a un uso responsabile, equo e trasparente dell'IA (cfr. cap. 6).

4.7. Gestione del rischio

L'AI Act definisce requisiti rigorosi per l'identificazione, valutazione e mitigazione dei rischi associati ai sistemi di IA "ad alto rischio", in particolare per garantire la tutela della salute, sicurezza e diritti fondamentali degli individui.

Le PA DEVONO adottare un approccio alla gestione del rischio conforme all'AI Act. Nell'attesa della definizione delle specifiche norme tecniche armonizzate (cfr. par. 4.4), la PA PUÒ fare riferimento alle norme tecniche UNI ISO 31000[24] e alla ISO/IEC 23894[25]. Queste ultime vanno applicate tenendo comunque conto delle prescrizioni dell'AI Act (cfr. par. 3.3).

L'allegato B propone uno strumento operativo utile alle attività di gestione del rischio per i sistemi IA "ad alto rischio".

Per i sistemi IA "di rischio minimo o nessun rischio", viceversa, le PA DOVREBBERO identificare rischi e contromisure legate all'uso dei sistemi stessi. L'analisi può essere svolta in forma semplificata rispetto ai sistemi "ad alto rischio".

Si noti che l'analisi del rischio non è un'attività raccomandata solo nelle realizzazioni di sistemi di PA; al contrario, si tratta di un passo metodologico previsto nella generalità dei progetti, informatici o meno. Pertanto, le indicazioni del presente paragrafo non vanno intese come ulteriori adempimenti che aggiungono complessità alle realizzazioni di sistemi di IA, già di per sé onerose. Al contrario, si tratta di declinare un'attività comune a tutte le tipologie di progetto (l'analisi del rischio), in modo che la suddetta analisi aiuti a conseguire gli obiettivi previsti minimizzando effetti indesiderati.

4.8. Valutazione d'impatto

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 (FRIA)(cfr. cap. 5) e la valutazione d'impatto sulla protezione dei dati personali (DPIA)(cfr. cap. 10).

I risultati della valutazione d'impatto DEVONO essere documentati e condivisi con le parti interessate.

Anche in questo caso, fermi restando gli obblighi di conformità all'AI Act, le PA DOVREBBERO definire un processo per valutare, anche se in forma semplificata, le potenziali conseguenze sui diritti fondamentali, derivanti dallo sviluppo, utilizzo o eventuale uso improprio, anche per i sistemi IA di "rischio

[24] UNI ISO 31000:2018 Gestione del rischio - Linee guida
[25] ISO/IEC 23894:2023 Information technology -Artificial intelligence -Guidance on risk management

minimo o nessun rischio". Resta ferma la necessità di svolgere la valutazione del rischio sulla protezione dei dati personali nelle attività istituzionali e progettuali condotte mediante strumenti di IA, ai sensi della normativa vigente e dei provvedimenti del Garante per la protezione dei dati personali (cfr. Cap. 10)

4.9. Piano operativo

Le PA DEVONO definire obiettivi operativi chiari per l'adozione e l'uso dell'IA (cfr. par. 4.3) assicurandone la coerenza con la propria strategia per l'IA e il rispetto dei requisiti normativi. Gli obiettivi DEVONO essere misurabili, monitorati e aggiornati periodicamente.

Per il raggiungimento degli obiettivi, la PA DEVE adottare un approccio sistematico e documentato di project management definendo un piano operativo, nel quale DEVE:

- definire le attività operative che devono essere effettuate;
- identificare le risorse richieste;
- definire i tempi entro i quali le attività devono essere completate;
- assegnare responsabilità specifiche;
- stabilire criteri di valutazione e monitoraggio dei risultati.

Il piano operativo DEVE includere tutte le attività necessarie per il conseguimento degli obiettivi riguardanti la gestione (cfr. presenti Linee guida), il procurement (cfr. Linee guida per il procurement dell'IA), lo sviluppo (cfr. Linee guida per lo sviluppo dell'IA) dei sistemi di IA adottati o adottandi dalla PA.

4.10. Risorse, competenze, comunicazione

Le PA DEVONO determinare e allocare le risorse finanziarie, tecnologiche e umane adeguate a sviluppare, mantenere e migliorare con continuità i propri sistemi di IA.

Le risorse possono includere:

- dati utilizzati nelle varie fasi del ciclo di vita del sistema di IA;
- algoritmi e modelli di IA;
- infrastrutture IT (es.: cloud computing, edge computing, risorse di elaborazione);
- risorse umane con le competenze necessarie a gestire il ciclo di vita del sistema di IA.

Le PA DEVONO determinare le competenze necessarie per il personale che svolge attività relative ai sistemi di IA. Le PA DEVONO garantire che tali persone acquisiscano le competenze necessarie tramite formazione, istruzione o esperienza e che conservino documentazione a comprova delle competenze acquisite.

Le PA DEVONO adottare le misure necessarie per acquisire le competenze e per valutarne l'efficacia (cfr. cap. 8).

Le PA DEVONO promuovere un uso responsabile ed efficace dell'IA, assicurando che il personale sia pienamente consapevole dei principi, degli obiettivi e della strategia per l'IA, nonché del codice etico e di comportamento.

Le PA DEVONO definire il piano di comunicazione interno ed esterno relativo ai sistemi di IA adottati (cfr. cap. 7).

Fermi restando gli obblighi di conformità all'AI Act, le PA DEVONO documentare e conservare le informazioni necessarie per assicurare la conformità normativa, trasparenza e tracciabilità dei sistemi di IA. È inclusa la documentazione operativa (manuali, procedure, valutazioni di impatto e report di monitoraggio) necessaria a garantire l'efficacia e il miglioramento continuo del sistema.

Le PA applicano a tale documentazione le Linee guida AgID sulla formazione, gestione e conservazione dei documenti informatici.

4.11. Implementazione

In questa fase le PA DEVONO attuare il piano operativo definito nella fase di pianificazione (cfr. par. 4.9).

Sempre in questa fase, le PA DEVONO condurre, a intervalli pianificati o in caso di modifiche significative, la valutazione dei rischi legati all'IA (cfr. par. 4.7). Le PA DEVONO mantenere, sui risultati delle valutazioni, una documentazione adeguata a garantire tracciabilità, conformità e azioni correttive tempestive.

Le PA DEVONO condurre la valutazione dell'impatto del sistema di IA, secondo la pianificazione definita (cfr. par 4.8) e in caso di modifiche significative. Le PA DEVONO mantenere una documentazione adeguata sui risultati di tutte le valutazioni effettuate.

4.12. Monitoraggio e valutazione

Le PA DEVONO definire - e poi monitorare regolarmente, conservando i risultati del monitoraggio - KPI utili:

- a misurare l'efficacia delle misure organizzative e tecniche pianificate dalla PA per la gestione dell'IA;
- a misurare le prestazioni tecniche dei sistemi di IA;
- a valutare l'impatto del sistema sui processi interni della PA e sui diritti fondamentali;
- valutare il valore aggiunto per la PA prodotto dal sistema di IA.

4.13. Miglioramento continuo

Le PA DEVONO garantire nel tempo l'idoneità, l'adeguatezza e l'efficacia dei sistemi di IA adottati e delle misure tecniche e organizzative per la gestione dell'IA tramite un processo di "miglioramento continuo".

5. Conformità delle soluzioni di IA

Le PA che adottano sistemi di IA DEVONO agire nel pieno rispetto della normativa nazionale ed unionale in materia di IA, con particolare riferimento all'AI Act.

Le PA DEVONO fare riferimento alla definizione di "sistema di IA" contenuta nell'art. 3, n. 1) dell'AI Act e richiamata nel cap. 3 delle presenti Linee guida, per comprendere se una tecnologia che la PA ha già adottato o intende adottare rientra o meno nell'ambito applicativo del suddetto regolamento. 3

Le PA DEVONO procedere alla classificazione del sistema di IA stabilendone il livello di rischio (cfr. par. 3.3).

Le PA DEVONO riconoscere il ruolo che ricoprono secondo l'AI Act, vale a dire: fornitore o *deployer* (cfr. cap. 3.2).

I casi in cui le PA ricoprono il ruolo di fornitore di sistemi di IA includono:

- l'erogazione di sistemi di IA mediante servizi in cloud ad altri soggetti pubblici o privati;
- lo sviluppo e la messa a disposizione mediante riuso di sistemi di IA ivi comprese le liste di prompt.

L'individuazione del livello di rischio a cui appartiene il sistema IA e il ruolo assunto dalla PA rispetto a tale sistema è funzionale all'individuazione delle disposizioni dell'AI Act applicabili.

5.1. Monitoraggio del ciclo di vita delle soluzioni di IA

Le PA, nel ruolo di fornitori o di *deployer,* DEVONO monitorare i propri sistemi di IA e il relativo impatto sui diritti fondamentali per tutto il ciclo di vita del sistema. Il tipo di monitoraggio dipende (come attività, frequenza, impegno) dal livello di rischio che emerge dall'attività di classificazione (vedi paragrafo precedente).

Per tutti i sistemi di IA, inclusi quelli classificati "a rischio limitato o minimo", le PA DEVONO:

- verificare che i dati siano trattati in conformità alle normative vigenti, con particolare attenzione alla protezione dei dati personali, al diritto d'autore e a eventuali ulteriori normative di settore;
- garantire la presenza e l'esaustività della documentazione tecnica a supporto del sistema di IA; la documentazione DEVE includere gli elementi utili a verificare i principi di trasparenza e spiegabilità dei risultati; la documentazione DEVE essere aggiornata durante tutto il ciclo di vita del sistema di IA.

Per i sistemi di IA classificati "a rischio limitato o minimo" le attività di monitoraggio possono consistere, ad esempio, in:

- **acquisizione e verifica di documentazione:** raccolta di report e relazioni che descrivono le caratteristiche dei sistemi IA, i suoi impatti sui servizi istituzionali della PA e sui diritti fondamentali;

- **verifiche e test operativi:** sperimentazione dei sistemi IA in condizioni reali, per valutarne prestazioni, affidabilità ed eventuali criticità;
- **scambio di informazioni e best practice:** creazione di un sistema di confronto periodico con il coinvolgimento di altre PA, fornitori, università e centri di ricerca per monitorare e migliorare l'esperienza di utilizzo.

Per i sistemi di IA ad alto rischio, le PA, nel ruolo di *deployer*, DEVONO effettuare, nei casi previsti dall'art. 27 dell'AI Act, una valutazione dell'impatto sui diritti fondamentali (FRIA), prima dell'utilizzo del sistema. Nell'effettuare tale valutazione, le PA DEVONO considerare i seguenti elementi:

- la **descrizione dei processi** della PA in cui il sistema di IA ad alto rischio sarà utilizzato;
- il **periodo e la frequenza temporale** con cui il sistema di IA ad alto rischio sarà utilizzato;
- le **categorie di persone fisiche e gruppi** verosimilmente interessati dal suo uso nel contesto specifico;
- i **dati oggetto di trattamento**, con particolare attenzione alla presenza di dati personali, attivando in tal caso ogni valutazione specifica;
- i **rischi specifici di danno** che possono incidere sulle categorie di persone fisiche o sui gruppi di persone interessati dal suo uso e sui relativi diritti fondamentali;
- le **misure da adottare** qualora tali rischi si concretizzino, comprese le disposizioni relative alla governance interna e ai meccanismi di reclamo;
- la descrizione **delle misure di sorveglianza umana**.

Se, invece, una PA ha il ruolo di fornitore di sistemi di IA ad alto rischio, DEVE:

- **documentare il sistema di monitoraggio** successivo all'immissione sul mercato, il quale deve essere proporzionato alla natura delle tecnologie di IA e ai rischi del sistema di IA ad alto rischio. Il sistema di monitoraggio successivo all'immissione sul mercato raccoglie, documenta e analizza i dati pertinenti raccolti per tutta la durata del loro ciclo di vita. Il monitoraggio successivo all'immissione sul mercato può includere analisi dell'interazione con altri sistemi di IA.
- **monitorare** circa la definizione e il rispetto del sistema di gestione dei rischi. Tale processo mira a individuare nonché attenuare i rischi pertinenti di tali sistemi di IA per la salute, la sicurezza e i diritti fondamentali. Il sistema di gestione dei rischi DEVE essere periodicamente riesaminato e aggiornato per garantirne l'efficacia costante, nonché la giustificazione e la documentazione delle eventuali decisioni e azioni significative adottate;
- **mettere a disposizione informazioni puntuali e comprensibili** in relazione alle modalità di sviluppo nonché di funzionamento di tali sistemi;
- assicurare, in linea con quanto disposto dall'art. 12 dell'AI Act, a livello tecnico, che la **registrazione degli eventi rilevanti** (i cosiddetti log) avvenga con modalità automatiche, per l'intera durata del ciclo di vita del sistema;

- **conservare le registrazioni** e disporre di documentazione tecnica contenente le informazioni necessarie per valutare la conformità del sistema ai requisiti dei sistemi di IA ad alto rischio.

5.2. Misure di sorveglianza

Le PA DEVONO adempiere gli obblighi riguardanti le misure di supervisione umana dei sistemi IA ad alto rischio previsti all'art. 14 dell'AI Act, in funzione del proprio ruolo:

- la PA fornitore DEVE progettare e sviluppare i sistemi di IA in modo da garantire un'efficace supervisione umana, integrando interfacce adeguate e strumenti di controllo;
- la PA *deployer* DEVE adottare procedure e strumenti per assicurare il monitoraggio, l'interpretazione e, se necessario, l'intervento umano sulle decisioni generate dall'IA.

A titolo meramente esemplificativo e in conformità con gli standard e le best practice elaborate dalle organizzazioni internazionali ed europee competenti, le PA DOVREBBERO adottare misure di sorveglianza umana che consentano, ove possibile, di intervenire sui parametri che determinano l'output del sistema IA («Human-On-The-Loop»), di influenzarne l'output («Human-In-The-Loop») oppure di determinare il modo in cui il medesimo viene utilizzato nei processi decisionali della PA («Human-In-Command»).

Le PA *deployer* affidano la sorveglianza umana dei sistemi di IA ad alto rischio al RTD e al suo Ufficio (UTD) e, come sopra già chiarito, DEVONO garantire che il personale incaricato possieda le competenze, l'esperienza e l'autorità necessarie per svolgere tale funzione in modo efficace.

5.3. Auditing e controlli nei confronti dei fornitori esterni

Le PA che si avvalgono di un sistema IA prodotto esternamente, che agiscono quindi in qualità di *deployer,* DEVONO preliminarmente verificare il rispetto delle norme previste dall'*AI Act* per i fornitori, a seconda della tipologia di rischio prodotto.

Le PA *deployer* di sistemi di IA ad alto rischio DEVONO verificare, preliminarmente alla loro adozione, che i sistemi siano conformi ai requisiti previsti dell'IA Act e chiedere ai fornitori di sistemi IA ad alto rischio di dimostrare e garantire, anche mediante adeguate certificazioni, i requisiti previsti all'art.16 dell'AI Act. Tra questi sono inclusi:

- la garanzia che i sistemi di IA ad alto rischio siano conformi ai requisiti dell'AI Act (sezione II)[26];
- l'esistenza di un solido sistema di gestione della qualità (art. 17);
- l'espletamento della procedura di valutazione della conformità richiesta dall'AI Act (art. 43);
- la redazione di tutta la documentazione pertinente necessarie per le verifiche (art. 18);

[26] I sistemi IA ad alto rischio devono soddisfare i seguenti requisiti definiti dalla sezione II dell'AI Act: sistema di gestione dei rischi (art. 9), dati e governance dei dati (art. 10), documentazione tecnica (art. 11), conservazione delle registrazioni (art. 12), trasparenza e fornitura di informazioni ai deployer (art. 13), sorveglianza umana (art. 14), accuratezza, robustezza e cibersicurezza (art. 15).

- l'istituzione di un sistema robusto per il monitoraggio successivo all'immissione sul mercato (art. 72).

Le PA DEVONO verificare che i fornitori di modelli di IA per finalità generale che presentino rischi sistemici abbiano assolto agli obblighi previsti all'art. 53 dell'AI Act, inclusa la documentazione relativa allo svolgimento di test e verifiche condotte da soggetti esterni e indipendenti rispetto al fornitore.

5.4. Conservazione della documentazione

Le PA DEVONO garantire la corretta conservazione della documentazione relativa all'impiego dei sistemi di IA ad alto rischio, in conformità con la normativa italiana ed europea sulla gestione documentale. Ciò include la conservazione dei dati e dei metadati necessari ad assicurare la tracciabilità dei processi decisionali automatizzati, nel rispetto della normativa in materia di protezione dei dati personali.

L'AI Act stabilisce l'obbligo di conservare documentazione tecnica sui sistemi di IA ad alto rischio, inclusa una descrizione del funzionamento, delle prestazioni e delle procedure di monitoraggio e controllo adottate per garantire la conformità ai requisiti di legge. Tale documentazione DEVE essere conservata per un periodo congruo rispetto all'utilizzo del sistema, in modo da garantire la possibilità di audit e verifiche da parte delle autorità competenti, e tenuto conto del periodo di conservazione della documentazione del procedimento che è supportato dal sistema di IA.

Le PA DEVONO gestire e conservare la documentazione dei sistemi di IA secondo le disposizioni del CAD e in conformità con le linee guida AgID sulla formazione, gestione e conservazione dei documenti informatici.

6. Governance etica dell'IA

Fermi restando gli obblighi di conformità all'AI Act (cfr. cap. 5) e ad ogni altra normativa unionale e nazionale, anche settoriale, applicabile, i quali DEVONO essere puntualmente adempiuti, le PA POSSONO valutare l'opportunità, in aggiunta ai vincoli predetti, di dotarsi di codici etici e di comportamento, anche in collaborazione con altre amministrazioni o adottando modelli definiti da enti sovraordinati o della stessa tipologia.

I codici etici e di comportamento sono strumenti adottati su base volontaria e sono finalizzati a:

- coadiuvare le PA nella selezione e nell'uso dei sistemi di IA, integrando e implementando le norme a vario titolo applicabili nello sviluppo o nell'uso dei sistemi IA;
- introdurre regole di utilizzo dei sistemi di IA da parte dei dipendenti; tali regole di utilizzo devono essere condivise con la dirigenza e/o vertici amministrativi.

I codici etici e di condotta NON POSSONO, in nessun caso, comportare un abbassamento delle tutele o il venir meno degli obblighi previsti dalla normativa vigente.

Le PA POSSONO sviluppare codici etici e di condotta, ai sensi dell'art. 95, par. 3 dell'AI Act, sia in qualità di fornitori sia di *deployer*, favorendo, nel processo di definizione degli stessi, il coinvolgimento dei portatori di interesse, anche privati (organizzazioni rappresentative della società civile, università, centri di ricerca associazioni di fornitori, enti regolatori).

Il contenuto dei codici etici e di condotta può riguardare la previsione di elementi e requisiti supplementari per lo sviluppo e l'utilizzo di sistemi di IA, che si aggiungono a quelli già obbligatori in forza della normativa applicabile. In particolare, ai sensi dell'art. 95, par. 1, dell'AI Act, i codici etici e di condotta dovrebbero incentivare l'applicazione volontaria a tutti i sistemi di IA adottati dalla PA, diversi dai sistemi di IA ad alto rischio, di alcuni o di tutti i requisiti di cui al capo III, sezione 2 dell'AI Act.

Nell'individuare eventuali elementi e requisiti supplementari, le PA DOVREBBERO considerare:

- le soluzioni tecniche disponibili e le best practice del settore;
- i principi etici sviluppati e/o in via di sviluppo in materia, come sancito dallo stesso art. 95 AI Act.

A tal fine, le PA DEVONO tenere in adeguata considerazione gli **Orientamenti etici per un'IA affidabile** del gruppo di esperti ad alto livello sull'intelligenza artificiale della Commissione Europea (AI HLEG)[27].

Le PA DOVREBBERO inoltre considerare, ai fini dello sviluppo e consolidamento dei codici etici, le principali fonti o iniziative europee o internazionali. Tra le iniziative in materia, a titolo esemplificativo, si possono richiamare i documenti redatti da:

[27] Orientamenti etici per un'IA affidabile https://data.europa.eu/doi/10.2759/640340.

- l'IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, gruppo che promuove iniziative nel settore, per esempio, l'*Ethically Aligned Design*[28];
- l'IEEE Standards Association, organo che contribuisce alla creazione di standard globali che promuovano l'innovazione responsabile nell'IA e nei sistemi autonomi; tra gli standard si possono richiamare i P7000 *Series*[29] ed P7010 *Series*[30];
- l'Organizzazione per la Cooperazione e lo Sviluppo Economico, che ha elaborato le raccomandazioni sull'IA, gli AI Principles[31].
- l'United Nations Educational, Scientific and Cultural Organization (UNESCO), che ha sviluppato le *Recommendation on the Ethics of Artificial Intelligence*[32].
- United Nations Secretary-General's High-level Advisory Body on Artificial Intelligence, che ha redatto il rapporto dal titolo *Governing AI for humanity*[33].
- Stato della Città del Vaticano, *Rome Call for Ethics*[34].
- Pontificia Commissione per lo Stato della Città del Vaticano, Decreto n° DCCII, *Linee Guida in materia di Intelligenza artificiale*[35].

In particolare, le raccomandazioni UNESCO evidenziano l'esigenza di mitigare i rischi prevedibili di conseguenze indesiderate e possibili abusi durante tutti gli stadi del ciclo di vita dei sistemi di IA: la ricerca, la progettazione e lo sviluppo fino alla installazione e utilizzo, inclusi manutenzione, funzionamento, monitoraggio, valutazione e dismissione.

Le Linee guida in materia di IA dello Stato della Città del Vaticano enunciano i principi generali tesi a valorizzare e promuovere un utilizzo etico e trasparente dell'IA in una dimensione antropocentrica e affidabile. Il documento sottolinea come l'IA non deve mai sostituirsi all'uomo e deve rispettare la sua autonomia. L'IA deve essere al servizio della persona e non dominarla, in modo che le decisioni ultime spettino sempre all'uomo.

Nel redigere un codice etico, le PA DOVREBBERO perseguire una serie di scopi (che si aggiungono a quelli già espliciti e tutelati dall'*AI Act*), In particolare, tali codici dovrebbero "*guidare la condotta degli attori coinvolti nello sviluppo e adozione dei sistemi di IA, indicando principi e valori di riferimento, nella direzione della promozione del benessere sociale e ambientale tramite l'IA*", ponendo l'accento su obiettivi etici quali:

- garantire elevati *standard* di competenza professionale e pratica etica;

[28] l'Ethically Aligned Design https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf.
[29] IEEE, Standard Model Process for Addressing Ethical Concerns during System Design https://standards.ieee.org/ieee/7000/6781/.
[30] IEEE, Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being https://standards.ieee.org/ieee/7010/7718/.
[31] OECD, Recommendation of the Council on Artificial Intelligence (2024) https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449.
[32] UNESCO, Recommendation on the Ethics of Artificial Intelligence (2024) https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.
[33] ONU, Governing AI for Humanity (2024) https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.
[34] Stato della Città del Vaticano, Rome call for ethics https://www.romecall.org.
[35] Stato della Città del Vaticano, N. DCCII - Decreto della Pontificia Commissione per lo Stato della Città del Vaticano recante "Linee Guida in materia di intelligenza artificiale" https://www.vaticanstate.va/images/N.%20DCCII.pdf

- promuovere la formazione continua del personale coinvolto nell'IA, al fine di sviluppare le competenze tecniche e le capacità utili ad affrontare le sfide etiche legate all'uso di tali tecnologie;
- promuovere la consapevolezza pubblica, quindi dei cittadini, sulle tecnologie di IA e sulle relative conseguenze;
- utilizzare l'IA per affrontare le sfide globali e garantire il bene pubblico;
- investire nella ricerca per mitigare i rischi legati alla sicurezza e alla società;
- contribuire allo sviluppo e all'adozione di standard tecnici internazionali per l'IA.

Le PA DEVONO garantire che i codici etici siano revisionati periodicamente, al fine di garantire il loro aggiornamento in base all'evoluzione tecnologica e normativa ed allineare tali iniziative agli *SDGs* in linea con il *Green Deal Europeo*. Le revisioni DOVREBBERO essere effettuate in consultazione con esperti, rappresentanti del settore pubblico e privato e associazioni di tutela dei diritti dei cittadini.

L'allegato D alle presenti Linee Guida, dal titolo "Schema di codice etico e di comportamento" relativo all'applicazione di sistemi di IA, fornisce uno strumento operativo che, nell'ambito dei principi generali relativi all'etica dell'IA, può contribuire ad orientare e sostenere le PA nella predisposizione di codici di comportamento con contenuti coerenti a quanto previsto dalla regolamentazione in materia.

7. Comunicazione

7.1. Misure di trasparenza

Le PA che adottano sistemi di IA DEVONO implementare misure di trasparenza finalizzate a garantire che gli utenti siano consapevoli del loro utilizzo e dell'impatto che tali sistemi possono avere sui processi decisionali. La trasparenza è un requisito fondamentale per assicurare la fiducia nei confronti delle tecnologie utilizzate, in conformità con i principi di trasparenza e responsabilità amministrativa sanciti dalle normative unionali.

L'AI Act prevede che i sistemi ad alto rischio e determinati sistemi di IA (AI Act art. 50) siano soggetti a requisiti stringenti di trasparenza. Le PA, a fronte del ruolo di fornitore o di *deployer,* DEVONO applicare o verificare che siano applicati tali requisiti. Questi includono:

- **informazioni sull'interazione con l'IA:** i fornitori DEVONO progettare e sviluppare i sistemi di IA in modo che gli utenti siano informati quando questi interagiscono con un sistema di IA. Questo significa che le persone devono essere consapevoli di quando stanno comunicando con una macchina anziché con un essere umano;
- **indicazione dei contenuti generati artificialmente:** i sistemi di IA che generano contenuti sintetici, come testi, immagini, audio o video, DEVONO identificare chiaramente tali contenuti come generati artificialmente; questo è particolarmente rilevante per i contenuti come deepfake, che devono essere etichettati come tali, per evitare rischi di manipolazione o disinformazione.
- **documentazione tecnica:** i fornitori di sistemi di IA, in particolare quelli considerati ad alto rischio (es. riconoscimento biometrico, sistemi di valutazione per l'accesso a servizi essenziali), DEVONO mantenere una documentazione tecnica che descriva il funzionamento del sistema; tale documentazione deve essere messa a disposizione delle autorità competenti in caso di necessità.
- **obblighi per sistemi ad alto rischio:** i fornitori DEVONO garantire che i sistemi di IA siano progettati per un'adeguata trasparenza, includendo istruzioni dettagliate per i responsabili del loro utilizzo, e che i sistemi ad alto rischio permettano la tracciabilità delle decisioni prese, affinché gli utenti comprendano come il sistema ha elaborato i dati e raggiunto una determinata decisione.

L'obiettivo principale di questi requisiti è assicurare che gli utenti siano sempre consapevoli dell'eventuale utilizzo di sistemi automatizzati nelle proprie interazioni con le PA e che possano esercitare i propri diritti (ad esempio di accesso, rettifica e opposizione), nel rispetto dei principi di legalità e trasparenza.

Le PA DEVONO essere trasparenti nell'utilizzo di sistemi di IA che trattano dati personali, dando adeguata, esaustiva e accessibile informativa agli interessati. come questi influenzano le decisioni e quali siano le conseguenze per gli utenti fruitori di servizi pubblici, rendendo note le finalità, i criteri e i metodi utilizzati dai sistemi di IA. Inoltre, devono assicurarsi che le informazioni fornite siano accessibili a tutti i soggetti interessati.

7.2. Obblighi di informativa

Nel rispetto del GDPR e dell'AI Act e come sopra accennato, le PA DEVONO garantire un elevato livello di trasparenza e chiarezza nell'impiego di sistemi di IA, soprattutto quando tali sistemi possono avere un impatto sui diritti e sugli interessi degli utenti.

In questo contesto, è fondamentale fornire un'informativa chiara, esaustiva e accessibile, affinché gli utenti possano comprendere il funzionamento dei sistemi di IA utilizzati dalla PA e i diritti che ne derivano in capo all'interessato. Oltre a quanto richiesto agli artt. 12-13-14 del GDPR in caso di trattamento di dati personali, l'informativa deve includere:

- **descrizione del sistema di IA e delle sue finalità:** le PA DEVONO fornire una descrizione dettagliata del sistema di IA utilizzato, specificandone la natura e gli obiettivi. L'informativa deve chiarire:
  - **tipo di IA impiegata:** se si tratta di un sistema di apprendimento automatico (machine learning), un algoritmo di analisi predittiva o un altro modello basato su IA;
  - **finalità del sistema:** le PA DEVONO indicare chiaramente le finalità per le quali il sistema è utilizzato, specificando se la soluzione di Intelligenza Artificiale sia impiegata per migliorare i servizi offerti, per l'automazione di processi amministrativi o per altre finalità legate al miglioramento dell'efficienza e dell'efficacia dell'azione amministrativa.
- **Criteri di decisione e impatto sui soggetti interessati:** le PA DEVONO spiegare in modo chiaro i criteri e i parametri su cui si basano le decisioni automatizzate prese dal sistema di IA. È essenziale che l'informativa comprenda:
  - **modalità di funzionamento del sistema:** indicando se la soluzione di Intelligenza Artificiale opera analizzando dati storici o se utilizza modelli statistici per formulare previsioni o raccomandazioni;
  - **possibili impatti delle decisioni automatizzate:** descrivendo l'eventuale impatto delle decisioni adottate dalla soluzione di IA sui cittadini, sia in termini di diritti che di eventuali conseguenze pratiche, come la possibilità di accedere a determinati servizi o benefici.
- **Spiegazione delle decisioni automatizzate:** i cittadini DEVONO essere informati della propria facoltà di ottenere una spiegazione comprensibile riguardo alle decisioni adottate tramite sistemi di IA; le PA DEVONO pertanto:
  - **garantire la trasparenza del processo decisionale:** fornire, su richiesta, informazioni chiare e semplici che permettano ai cittadini di comprendere le logiche sottostanti alle decisioni automatizzate che li riguardano;
  - **descrivere i criteri e i dati utilizzati:** essere in grado di spiegare quali dati sono stati utilizzati nel processo e come tali dati hanno influenzato l'esito della decisione.

- **Diritto di opposizione e intervento umano nel processo decisionale:** è fondamentale che le PA informino i propri utenti della possibilità di opporsi a decisioni automatizzate e di richiedere l'intervento umano nel processo decisionale, qualora questo possa avere effetti significativi sui loro diritti. A tale scopo è necessario:
    - **specificare le modalità per l'esercizio del diritto di opposizione:** le PA DEVONO indicare ai cittadini le procedure e i canali attraverso cui possono opporsi a una decisione automatizzata, evidenziando i tempi di risposta e le modalità per richiedere una revisione da parte di un operatore umano;
    - **garantire l'accesso a un intervento umano:** qualora un cittadino ritenga che la decisione automatizzata non sia accurata o giusta, DEVE essere garantita la possibilità di richiedere un intervento umano per riesaminare la decisione e, se necessario, modificarla.

Le PA DOVREBBERO elaborare informative dettagliate e personalizzate per ogni sistema di IA utilizzato, adattando la comunicazione alle specificità del contesto e alle esigenze degli utenti.

7.3. Adottare l'IA nella comunicazione istituzionale

La comunicazione è una leva strategica per la PA, per informare cittadini e imprese e per promuovere i servizi pubblici. Grazie a tecnologie avanzate come l'elaborazione del linguaggio naturale, il machine learning e l'automazione, l'IA può migliorare l'efficacia, l'accessibilità e l'interattività delle comunicazioni, contribuendo a un servizio pubblico più trasparente e inclusivo.

L'IA, quindi, può trasformare la comunicazione pubblica, pur richiedendo un'attenta gestione delle sfide etiche e operative.

Fra le opportunità e i vantaggi che l'IA porta con sé figurano:

- ottimizzazione dei processi creativi: analizzare tendenze e dati in tempo reale, generare contenuti testuali e visivi, campagne di comunicazione mirate e innovative;
- personalizzazione: personalizzare le esperienze degli utenti e rendere ogni interazione unica e significativa;
- automazione e risposte rapide: chatbot e assistenti virtuali rispondono in modo immediato e riducono i tempi di attesa;
- aumento dell'accessibilità: le informazioni possono essere più comprensibili anche ai cittadini con disabilità grazie a strumenti di traduzione o di sintesi vocale;
- riduzione dei costi: automatizzando compiti ripetitivi, si libera tempo prezioso da dedicare alla strategia creativa.

Le PA che adottano sistemi di IA ai fini della comunicazione istituzionale DEVONO seguire un percorso graduale e pianificato secondo quanto previsto al Capitolo 4. Inoltre, come già introdotto al Capitolo 8, è

opportuno che i dipendenti abbiano una formazione adeguata. Le PA che adottano l'IA nelle proprie attività di comunicazione:

- DEVONO redigere un piano strategico e operativo per l'uso dell'IA nelle azioni di comunicazione;
- DEVONO monitorare e valutare l'efficacia degli strumenti di IA utilizzati;
- DEVONO formare il personale sull'uso degli strumenti di IA;
- POSSONO adottare progetti pilota di IA per valutarne gli impatti e l'effettiva utilità.

Le PA possono utilizzare strumenti e tecnologie basate sull'IA per potenziare le proprie attività di comunicazione. In particolare, le PA:

- POSSONO utilizzare chatbot e assistenti virtuali che migliorano l'interazione con gli utenti e riducono i tempi di risposta;
- POSSONO usare strumenti di *sentiment analysis* per i social media e altre forme di interazione;
- POSSONO adottare piattaforme di monitoraggio e gestione dei social media per gestire in maniera automatizzata la pubblicazione di contenuti;
- POSSONO usare strumenti di sintesi automatica del linguaggio naturale (NLP) per sintetizzare documenti complessi o risposte a quesiti;
- POSSONO usare sistemi di traduzione automatica per consentire l'accesso ai contenuti ad un numero maggiore di utenti

L'uso dell'IA nella comunicazione può portare numerosi vantaggi, ma è fondamentale adottare un approccio responsabile e trasparente. In particolare, le PA che utilizzano IA per le proprie attività di comunicazione:

- DEVONO informare i cittadini sull'uso dell'IA e su come vengono gestiti i loro dati;
- DEVONO garantire la protezione dei dati personali eventualmente coinvolti nel trattamento;
- DEVONO essere trasparenti e segnalare l'uso di strumenti di IA;
- DEVONO garantire la massima accessibilità e comprensione da parte di tutti;
- NON DEVONO sviluppare una dipendenza eccessiva dall'IA o affidarsi completamente all'IA, garantendo sempre una supervisione umana;
- NON DEVONO adottare una comunicazione impersonale ma mantenere un tono umano ed empatico.

8. Formazione e sviluppo delle competenze

La PA, al fine di cogliere in modo consapevole i vantaggi dell'IA, DEVE lavorare su una serie di fattori abilitanti, di tipo gestionale, organizzativo e culturale oltre che tecnologico, tra cui riveste un ruolo chiave l'acquisizione e lo sviluppo di competenze adeguate a livello individuale e organizzativo.

La PA DEVE essere in grado di comprendere a fondo le potenzialità e le implicazioni dell'IA per migliorare la qualità, l'accessibilità e l'efficienza dei servizi pubblici, ad esempio attraverso l'automazione di attività di routine, l'analisi predittiva e il supporto decisionale.

Inoltre, la PA DEVE sviluppare competenze specifiche per poter governare e regolamentare l'utilizzo dell'IA, garantendo il rispetto di principi etici, di protezione dei dati personali e di trasparenza. Questo aspetto è fondamentale per promuovere un'adozione responsabile e sostenibile di queste tecnologie, che possano contribuire a migliorare l'equità e l'inclusione nell'erogazione dei servizi pubblici.

Infine, la capacità di offrire opportunità di formazione e sviluppo di competenze sull'IA può rendere il settore pubblico all'avanguardia nell'innovazione tecnologica e più attrattivo per i talenti, contribuendo a trattenere le migliori risorse e a rafforzare la competitività del sistema amministrativo nel suo complesso.

Il quadro strategico e regolatorio che è andato definendosi nel corso degli ultimi anni dedica particolare attenzione al tema dell'up-skilling connesso alla diffusione e adozione dell'IA. L'AI Act, in particolare, evidenzia come sia necessario procedere ad interventi differenziati di formazione in relazione al background, al ruolo assunto e alle previsioni di utilizzo delle soluzioni di IA. Ciò al fine di assicurare che tutti gli individui (o stakeholder) che si trovano ad interagire a vario titolo con soluzioni che si avvalgono di IA possano farlo in modo consapevole, sicuro, affidabile ed efficace.

Lo standard ISO/IEC 42001 sul sistema di gestione dell'IA stabilisce, inoltre, requisiti specifici relativi alle competenze del personale coinvolto nelle attività che influenzano le prestazioni dell'IA all'interno di una organizzazione, prevedendo la necessità di assicurare la costante presenza di competenze adeguate attraverso l'adozione di azioni volte a colmare le lacune rilevate ed il loro tracciamento e monitoraggio nel tempo.

In linea generale, è pertanto fondamentale la necessità di una mappatura attenta e comprensiva delle figure professionali già presenti nelle PA, con l'obiettivo di identificare chiaramente le responsabilità legate alla governance dell'IA. Questo processo permetterà di definire in maniera mirata i piani formativi e di up-skilling, garantendo che il personale disponga delle competenze necessarie per gestire in modo efficace e responsabile i sistemi IA.

Come richiamato anche nell'AI Toolkit predisposto dall'OCSE[36], alla stregua di ogni altra trasformazione digitale nel sistema pubblico, il raggiungimento della maturità nel ricorso all'IA richiede un ampio e variegato sistema di competenze a diversi livelli dell'organizzazione, che spazia dalle competenze di base a quelle di tipo specialistico e che richiede lo sviluppo di interventi di upskilling e formazione differenziati e mirati.

[36] OECD/UNESCO (2024), G7 Toolkit for Artificial Intelligence in the Public Sector, OECD Publishing, Paris, https://doi.org/10.1787/421c1244-en.

Vista la natura sempre più pervasiva delle applicazioni di IA in molti aspetti della vita sociale e del lavoro (anche pubblico), la capacità di interagire con tali applicazioni comporta in primo luogo la necessità di ampliare il ventaglio delle competenze digitali di base di tutti dipendenti pubblici, andando ad includere la cosiddetta *AI literacy,* che affianca alla comprensione di base dell'esistenza dell'IA e della sua capacità di influenzare il proprio lavoro (*AI awareness*) la comprensione di *come* e *perché* si esplica tale influenza, includendo la padronanza delle conoscenze sugli aspetti normativi, operativi e di dominio associati all'IA.

I dipendenti pubblici hanno necessità di riconoscere le caratteristiche essenziali delle applicazioni di IA sviluppate ad hoc o integrate nelle nuove release dei sistemi gestionali e di automazione di ufficio messe a disposizione dalla propria organizzazione, al fine di comprenderne le opportunità e le modalità di utilizzo e sfruttarne al meglio le potenzialità per rendere più veloce, efficiente, sicuro e consapevole il proprio lavoro. Al contempo, hanno necessità di acquisire consapevolezza rispetto alle implicazioni nella trasposizione nel contesto lavorativo delle modalità di interazione con applicazioni di IA ricorrenti nella vita privata (es. motori di ricerca basati su IA, applicazioni di IA generativa basati su licenza individuale, applicazioni *embedded* nei social media, etc.).

I dipendenti pubblici necessitano di essere formati anche sui principi etici legati all'uso dell'IA, inclusa la trasparenza nell'ambito di decisioni prese in maniera automatizzata e devono, altresì, sviluppare le competenze necessarie ad analizzare e gestire i rischi di sicurezza cui sono esposte le attività supportate dai sistemi di IA e i dati da essi utilizzati.

La piena consapevolezza delle modalità e delle condizioni ottimali di utilizzo delle applicazioni di IA, oltre ad assicurare la corretta adozione di tali sistemi, è estremamente importante anche per superare le fisiologiche resistenze che notoriamente ha incontrato il percorso di trasformazione digitale della PA. Nel caso dell'IA tali resistenze potrebbero addirittura risultare ulteriormente rafforzate da una non corretta interpretazione dell'attenzione normativa all'approccio basato sul rischio promosso dall'AI Act. In questa prospettiva, per cogliere effettivamente e in modo corretto le opportunità offerte, tutti i dipendenti pubblici DOVREBBERO acquisire un'alfabetizzazione sull'IA, in quanto ambito ormai imprescindibile delle cosiddette competenze digitali di base.

Accanto alla diffusione e al consolidamento di una cultura di base e condivisa tra tutte le amministrazioni è poi opportuno identificare e rafforzare competenze tecniche di dominio che riguardano specifici processi coinvolti dall'IA (es. la formazione stessa o il reclutamento) e/o e ambiti applicativi dell'IA quali, tra gli altri, la sanità, i trasporti o la formazione.

Le PA, ed in particolare gli Uffici per la trasformazione digitale (UTD) necessitano di rafforzare e/o sviluppare le competenze specialistiche connesse all'IA attraverso il ricorso a politiche mirate di reclutamento e *upskilling* volte soprattutto ad attrarre e trattenere i talenti. Tale esigenza nel settore pubblico è solo in parte mitigata dall'esternalizzazione delle attività di progettazione e sviluppo attraverso il ricorso a partner tecnologici di varia natura (enti in house o operatori di mercato), in quanto resta in ogni caso essenziale il possesso di competenze adeguate per formulare il fabbisogno dell'organizzazione e governare le partnership, ancor più con

il consolidarsi delle metodologie di sviluppo di tipo agile cui ricorrono sempre più frequentemente i fornitori tecnologici.

La maggior parte dei funzionari e dirigenti tecnici, per evidenti ragioni temporali, difficilmente hanno avuto modo di approfondire in modo sistematico, nella propria formazione universitaria, le tematiche di IA, che si sono prevalentemente sviluppate in modo rilevante negli ultimi anni. È quindi necessario, tra le altre, sviluppare competenze relative alle moderne metodologie e tecnologie di IA e machine learning, con riferimento anche agli aspetti tecnici delle problematiche relative ai rischi, nonché rafforzare la capacità di intercettare e gestire in maniera sicura e resiliente eventuali crisi dovute a deviazioni delle condizioni di funzionamento in caso di eventi indesiderati.

I profili professionali coinvolti nello sviluppo e gestione di progetti e soluzioni che si basano sull'AI sono molto variegati, in quanto la disciplina richiede competenze che coprono diversi settori come l'informatica, l'ingegneria, la matematica, il diritto delle tecnologie, senza trascurare gli aspetti umanistici. Lo standard ISO/IEC 42001 raccomanda di considerare la necessità di competenze diversificate in funzione delle diverse fasi ciclo di vita dell'AI, così come per garantire l'adeguata rappresentatività con riferimento ai dati usati per l'addestramento dei modelli di machine learning.

Tra le figure professionali sempre più critiche rientrano:

- **data engineer**, un esperto nella progettazione, nello sviluppo e nella gestione dei dati che si occuperà di garantire che i dati vengano raccolti, raffinati e resi disponibili in maniera efficiente e affidabile. La qualità del dato è cruciale per la restituzione di output affidabili;
- **machine learning engineer**, un esperto in algoritmi di machine learning e di programmazione in grado di progettare, implementare e ottimizzare algoritmi di machine learning fondamentali per il corretto funzionamento dei servizi digitali sviluppati con questa tecnologia;
- **prompt engineer**, l'esperto che istruisce l'IA generativa alla produzione di output efficaci, secondo le indicazioni ricevute;
- **deep learning engineer**, un esperto in reti neurali, deep learning e programmazione, in grado di applicare l'intelligenza artificiale a problemi più complessi, che richiedono grandi quantità di calcolo tramite l'apprendimento automatico, come ad esempio il riconoscimento vocale, computer vision e elaborazione del linguaggio naturale;
- **data scientist**, un esperto in analisi dei dati, machine learning e statistica che lavorando a stretto contatto con il machine learning engineer, analizza grandi quantità di dati, crea modelli predittivi e lavora per estrarre dati utilizzando algoritmi di IA e machine learning. Il data scientist trasformano i dati in valore, consentendo di prendere decisioni informate basate su dati concreti e affidabili;
- **AI engineer**, un esperto specializzato nella creazione e nell'implementazione di modelli di intelligenza artificiale attraverso attività di ricerca teorica in ambito IA. Questa figura professionale è fondamentale per la scelta e la valutazione del giusto modello di IA da applicare nell'implementazione di servizi software. Ogni modello ha le proprie caratteristiche, rendendolo adatto a specifiche casistiche, scelta cruciale per il successo del servizio implementato;

- **AI architect**, un esperto in grado di valutare e progettare l'architettura dei sistemi di intelligenza artificiale assicurando che siano scalabili, sicuri e performanti. All'occorrenza, collabora nell'integrazione dell'IA nelle infrastrutture esistenti. La scelta dell'architettura influenza la realizzazione della soluzione software incidendo su tempi, costi e prestazioni;
- **esperto di cybersecurity**, un esperto in sicurezza informatica in grado non solo di analizzare e prevenire le possibili minacce informatiche ma in grado di coniugare il tema della cyber security all'intelligenza artificiale, garantendo difese sempre più efficienti ed in grado di adattarsi autonomamente alle minacce esterne;
- **AI ethicist**, un esperto in grado di analizzare e valutare gli impatti sociali legati all'utilizzo dell'intelligenza artificiale, definendo linee guida e sensibilizzando chi realizza e utilizza gli strumenti di intelligenza artificiale. Il suo ruolo deve essere integrato in tutto il ciclo di vita di questi strumenti affrontando temi legati alla discriminazione, alla responsabilità, alla trasparenza e alla privacy;
- **giurista informatico**, un esperto in grado di combinare competenze legali e informatiche per fornire consulenze legali relative all'uso delle tecnologie informatiche che possano impattare su temi come la protezione dei dati personali, la sicurezza informatica e la proprietà intellettuale. Una figura professionale in grado anche di redigere e revisionare contratti relativi a software, licenze ecc., assicurando che siano conformi alle normative vigenti senza dimenticare le tematiche legate alla compliance, alla gestione delle controversie e all'analisi e la valutazione dei rischi derivanti dall'uso di una tecnologia come l'intelligenza artificiale;
- **esperto di protezione dei dati personali**, un esperto in grado di garantire che i processi, i sistemi e le applicazioni che utilizzano l'intelligenza artificiale rispettino il diritto fondamentale alla protezione dei dati personali e le normative sulla protezione dei dati personali, nazionali ed internazionali, come ad esempio il GDPR. Un professionista in grado di minimizzare i rischi legati alla raccolta, elaborazione e conservazione dei dati personali all'interno di sistemi di IA;
- **change manager**, figura in grado di valutare gli impatti dell'introduzione dell'IA e di elaborare strategie di cambiamento necessarie.
- **esperti umanisti**, più figure professionali umanistiche che collaborano con esperti in tecnologie informatiche per trattare questioni etiche, sociali, culturali e filosofiche, che inevitabilmente vengono coinvolte dall'uso e dallo sviluppo dell'intelligenza artificiale (ci si riferisce, ad esempio, a filosofi, sociologi, psicologi linguisti e altre professionalità dell'ambito umanistico.

Nella PA emerge pertanto la necessità di definire i profili professionali e le carriere di dirigenti e specialisti per l'AI, al fine di incentivare e guidare opportunamente i percorsi di reclutamento e up-skilling. In questa direzione va l'attività di qualificazione dei profili professionali relativi all'IA svolta dalla commissione tecnica di UNINFO, UNI/CT 526 "Attività professionali non regolamentate".

Insieme ai Responsabili per la trasformazione digitale e ai loro uffici, anche i dirigenti pubblici necessitano di un set di competenze e conoscenze adeguate ad assumere decisioni consapevoli e affidabili in merito a quali e quanti strumenti che ricorrono all'IA adottare, in virtù del loro ruolo determinante nei processi

decisionali connessi alla progettazione ed erogazione di servizi pubblici, nel rispetto e nella applicazione di norme e regole e più in generale nel processo di trasformazione del settore pubblico.

In particolare, i dirigenti pubblici necessitano di comprendere le proposte di progetti avanzate dagli esperti ed essere in grado di valutarne le opportunità, le implicazioni e gli impatti[37], affiancando alle competenze di base che accomunano tutti i dipendenti pubblici non specialisti IT la piena comprensione di principi, concetti e applicazioni in una prospettiva non solo tecnica ma anche - se non soprattutto - strategica, così come delle implicazioni di accessibilità, legali ed etiche per riuscire a identificare in modo compiuto le opportunità e le sfide connesse all'IA. Inoltre, in linea con quanto previsto dalle Linee guida per uno sviluppo sicuro dell'AI[38], alla stregua degli specialisti, ove responsabili dell'adozione dell'IA, necessitano di sviluppare le capacità di comprendere le minacce, i rischi e le relative mitigazioni ad essa associate. Per guidare opportunamente l'innovazione è infine necessario che i decisori abbiano piena comprensione dei fattori di complessità connessi all'implementazione delle soluzioni di IA e, come per qualunque altro progetto di cambiamento, le soft skills, le competenze manageriali e di leadership trasversali alle competenze abilitanti tutti i processi di transizione delle amministrazioni – in primis quella digitale[39].

8.1. Indicazioni operative per le PA

La definizione di attività formative per tutto il personale, la formazione continua specialistica per l'UTD, l'individuazione e selezione delle competenze necessarie alla gestione delle attività è un processo strategico che richiede piena consapevolezza da parte della PA e adeguata pianificazione degli interventi.

Un percorso che si snoda attraverso la valutazione delle esigenze formative, la definizione degli obiettivi, la definizione di programmi formativi, il coinvolgimento e l'incentivazione dei soggetti interessati, nonché di valutazione e aggiornamento continuo, all'interno di strumenti di programmazione (vedi PIAO) ben noti alle amministrazioni.

Al fine di dotarsi dell'adeguato sistema di competenze necessario per approcciare l'adozione di applicazioni che si basano sull'IA, le PA sono chiamate a introdurre una serie di interventi in coerenza con la propria strategia, anche avvalendosi delle opportunità offerte da iniziative di sistema realizzate o pianificate rispetto a questo ambito, principalmente su impulso del PNRR.

In particolare, a partire dall'analisi e descrizione dell'articolato sistema di competenze sull'IA che attraversa tutti i livelli di organizzazione deriva un set di indicazioni atte a guidare e orientare l'azione delle singole amministrazioni.

[37] ARISA (2023). AI Skills Needs Analysis, https://aiskills.eu/wp-content/uploads/2023/06/ARISA_AISkills-Needs-Analysis_V1.pdf
[38] Le linee guida sono pubblicate sul sito di ACN all'indirizzo https://www.acn.gov.it/portale/linee-guida-ia
[39] Si veda sul punto il Decreto del Ministero per la Pubblica Amministrazione del 28 settembre 2022 di adozione delle Linee guide sull'accesso alla dirigenza pubblica e il Decreto del Ministro per la Pubblica Amministrazione del 28 giugno 2023 di adozione del Framework delle competenze trasversali del personale di qualifica non dirigenziale delle pubbliche amministrazioni, nonché la Direttiva del Ministro per la pubblica amministrazione del 28 novembre 2023 che ha introdotto una specifica articolazione delle competenze di leadership che i dirigenti sono chiamati ad adottare e su cui devono essere valutati.

Le PA DEVONO promuovere l'*AI literacy* di tutti i dipendenti pubblici, riconosciuta nell'ambito della Direttiva MiPA sulla formazione 2025 a tutti gli effetti una competenza digitale di base[40], attraverso la definizione e il monitoraggio di specifici obiettivi di performance individuali connessi allo sviluppo di tali competenze. In particolare, le amministrazioni devono favorire il conseguimento degli obiettivi formativi definiti dalla Direttiva sulla formazione del Ministro della PA del 24 marzo 2023[41] per il 2024 e 2025 con riferimento al programma formativo disponibile su Syllabus denominato "Competenze digitali per la PA" e la fruizione, sempre su Syllabus, di programmi formativi riguardanti in modo specifico l'applicazione dell'Intelligenza Artificiale da parte dei dirigenti e dipendenti non specialisti IT.

Le PA DEVONO garantire la massima diffusione delle Linee Guida sull'IA realizzate da AgID presso i soggetti decisori coinvolti, o potenzialmente interessati, nella realizzazione e gestione di progetti e soluzioni che prevedono il ricorso all'IA in funzione delle specifiche competenze, affinché vengano adottate scelte consapevoli, coerenti e in grado di indirizzare tutti gli aspetti tecnologici, gestionali, organizzativi, etici e legali implicati.

Le PA DEVONO promuovere e favorire la partecipazione degli RTD, degli UTD e di altri attori chiave nell'attuazione della trasformazione digitale a learning communities sull'adozione dell'Intelligenza Artificiale, per stimolare il capacity building e la condivisione di buone pratiche tra le amministrazioni. La possibilità di condividere esperienze, conoscenze, metodologie e casi studio da parte delle amministrazioni apripista può rappresentare una soluzione efficace per irrobustire il sistema di conoscenze del settore pubblico e promuovere la creazione di un ambiente di apprendimento che incoraggi l'adozione dell'IA e la propensione a replicare o scalare progetti di successo. In particolare, le amministrazioni devono promuovere e favorire la partecipazione dell'UTD alla community promossa da AgID dedicata ai Responsabili per la Transizione al Digitale (progetto ReTe Digitale) che per sua natura rappresenta lo spazio privilegiato per la condivisione, collaborazione e confronto sull'IA. Le amministrazioni possono altresì promuovere la creazione/partecipazione ad altre learning community dedicate allo scambio, confronto e condivisione su specifiche tematiche connesse a tali tecnologie o specifici domini, anche circoscritto a specifiche tipologie di enti.

Le PA che adottano o intendono adottare progetti e soluzioni di IA DEVONO promuovere e favorire la definizione di obiettivi di performance dei dirigenti connessi alla partecipazione a percorsi formativi a supporto del *decision making* e della guida dei processi di adozione dell'IA, affiancando all'alfabetizzazione sull'IA l'approfondimento di tematiche di natura tecnica, normativa o etica e lo sviluppo di competenze gestionali, di leadership e delle soft skills a supporto della gestione dei processi di cambiamento, anche attraverso il ricorso a metodologie didattiche innovative. A tal fine le amministrazioni possono avvalersi oltre che della formazione disponibile sulla piattaforma Syllabus, dell'offerta formativa promossa dalla Scuola Nazionale dell'Amministrazione, dai suoi poli formativi territoriali e da Formez PA, di quella promossa dalle Università,

[40] Direttiva del Ministro della Pubblica Amministrazione "Valorizzazione delle persone e produzione di valore pubblico attraverso la formazione. Principi, obiettivi e strumenti del 14 gennaio 2025.
[41] Direttiva del Ministro della Pubblica Amministrazione, "Pianificazione della formazione e sviluppo delle competenze funzionali alla transizione digitale, ecologica e amministrativa promosse dal Piano Nazionale di Ripresa e Resilienza" del 23 marzo 2023

anche attraverso il programma "PA 110 e lode" finanziato dal Dipartimento della funzione pubblica della PCM, di altri interventi formativi a valere su fondi europei, nazionali o reginali e dell'autofinanziamento.

Le PA che realizzano progetti di trasformazione digitale basati sull'AI DEVONO promuovere la realizzazione di programmi formativi mirati e contestualizzati a supporto dei processi di adozione e change management. Se le competenze di AI *literacy* sono essenziali per complementare le competenze digitali chiave per i dipendenti pubblici e creare una cultura comune e condivisa abilitante la trasformazione digitale, l'effettivo dispiegamento di progetti mirati di applicazione di IA nelle singole amministrazioni o in determinate tipologie di amministrazioni (si pensi a progetti che coinvolgono le organizzazioni del comparto sanitario) richiede il ricorso a ulteriori investimenti in formazione e change management volti ad approfondire aspetti specifici di progettazione, funzionamento e adozione. Per far fronte a tali interventi, le amministrazioni possono sviluppare specifici progetti formativi anche di carattere innovativo (basati, tra gli altri, su seminari e conferenze, hackathon, programmi di mentorship), avvalendosi di progettualità promosse nell'ambito del PNRR, come nel caso dell'iniziativa PerformaPA, fondi europei, nazionali o regionali e dell'autofinanziamento.

Le PA coinvolte in attività di sviluppo, acquisizione e gestione dell'IA DEVONO favorire l'accesso a percorsi tecnico-specialistici rivolti all'UTD e in generale ai professionisti IT a partire da una rilevazione sistematica e periodica dei fabbisogni di competenze, anche avvalendosi della collaborazione con Università (es. progetto PA 110 e Lode), istituti di alta formazione e centri di competenza a livello nazionale e internazionale, e promuovendo l'acquisizione e il mantenimento di certificazioni professionali.

Le PA DEVONO, altresì, prevedere percorsi relativi alla sicurezza cibernetica dell'IA e concernenti la capacità di rispondere a minacce, incidenti e crisi cibernetiche nel contesto dell'IA.

Le PA DEVONO dare opportuna evidenza di ciascuno degli interventi formativi individuati nell'ambito della sezione della pianificazione della formazione prevista del PIAO nelle modalità e nei termini previsti per la programmazione, monitoraggio e rendicontazione previste dalla Direttiva MiPA sulla formazione 2025.

Le PA DOVREBBERO promuovere e incoraggiare una cultura dell'apprendimento continuo improntata alla formazione e all'arricchimento costante delle competenze di intelligenza artificiale. Questo approccio favorisce un ambiente in cui la crescita professionale viene valorizzata e incentivata, rendendo l'apprendimento una parte integrante della quotidianità lavorativa;

Le PA che si avvalgono o intendono avvalersi di soluzioni basate sull'IA POSSONO intraprendere azioni per favorire l'attrazione di talenti e di giovani specialisti, attraverso la promozione di iniziative che includono, tra gli altri, hackathon, attività di tirocinio, apprendistato e di dottorato. A riguardo le amministrazioni possono avvalersi anche delle opportunità di finanziamento prospettate da iniziative quali i dottorati innovativi promossi dal Ministro dell'università e ricerca, i tirocini e dottorati InPA promossi dal Dipartimento della funzione pubblica o promuovendo iniziative di raccordo con le Università e gli ITS.

Le PA che si avvalgono o intendono avvalersi di soluzioni basate sull'IA POSSONO promuovere attività di ricerca interdisciplinare attraverso iniziative di finanziamento o la partecipazione a progetti dedicati in collaborazione con enti di formazione, università, centri di ricerca e fornitori di soluzioni di IA.

Infine, poiché l'adozione dell'IA non può prescindere dallo sviluppo delle competenze digitali e sull'intelligenza artificiale dei destinatari dei servizi pubblici che se ne avvalgono, le amministrazioni, ed in particolare gli enti locali e le amministrazioni erogatrici di servizi, POSSONO contribuire a promuovere l'AI *awareness* e AI *literacy* dei cittadini, affinché riescano ad interagire in modo efficace, consapevole e sicuro con soluzioni di IA. I cittadini devono essere messi nelle condizioni di conoscere e sfruttare le potenzialità e gli impatti di tali tecnologie in modo da poter aumentare la propria propensione all'accesso ai servizi in modo consapevole, sicuro e rispettoso dei propri diritti. A tal fine le amministrazioni, ed in particolare gli enti territoriali possono promuovere presso i cittadini di riferimento (ad es. attraverso azioni informative e l'attivazione di collaborazioni con le amministrazioni attuatrici) le iniziative finanziate dal PNRR del Servizio civile digitale e della Rete dei servizi di facilitazione per favorire una crescente consapevolezza sui temi dell'intelligenza artificiale e sulle modalità di accesso ai servizi che si avvalgono di tali tecnologie[42], oltre che attivare progetti ad hoc.

[42] Il progetto Rete dei servizi di facilitazione digitale è coordinato dal Dipartimento per la trasformazione digitale ed attuato in collaborazione con Regioni, Province Autonome ed enti locali. Il progetto Servizio Civile Digitale è realizzato invece dal Dipartimento per la trasformazione digitale. in collaborazione con il Dipartimento per le politiche giovanili. Entrambi i progetti, realizzati nell'ambito della Misura 1.7 del PNRR, seppur con differenze di governance e operative, mirano all'attivazione di centri distribuiti su tutto il territorio nazionale dove i cittadini possono accedere a servizi di facilitazione e formazione per sviluppare e rafforzare le proprie competenze digitali di base. https://repubblicadigitale.gov.it/portale/progetti-del-dipartimento

9. Gestione e qualità dei dati

L'AI Act nel Considerando (67) evidenzia l'importanza della qualità dei dati quale prerequisito essenziale per sistemi di IA affidabili[43]. Lo stesso testo poi osserva che "*Per disporre di set di dati di addestramento, convalida e prova di elevata qualità è necessario attuare adeguate pratiche di governance e gestione dei dati*". Se ne ricava che la governance e la corretta gestione dei dati sono aspetti propedeutici a garantire la disponibilità e la qualità dei dati: è necessario contare su grossi volumi di dati per l'addestramento dell'IA ed è altrettanto necessario avere una qualità adeguata ad ottenere maggiore precisione e affidabilità dei risultati ricavati dai sistemi di IA.

Consapevole della rilevanza dei dati, anche al fine di creazione di valore, l'Unione Europea da alcuni anni ha definito un quadro normativo e regolatorio volto a favorire la circolazione dei dati all'interno dell'UE e in tutti i settori a vantaggio delle imprese, dei ricercatori e delle PA, garantendo, allo stesso tempo, protezione, diritti fondamentali, sicurezza e cybersicurezza, che sono tutti aspetti di cui bisogna tener conto quando si parla di IA. Tale quadro comprende, tra l'altro:

- l'adozione di una Strategia europea sui dati e l'identificazione di spazi di dati comuni;
- la definizione di una governance dei dati (con il Data Governance Act);
- la regolazione dell'accesso a tutti i dati generati dai prodotti connessi e dai servizi correlati (con il Data Act);
- la disponibilità di dati del settore pubblico per il riutilizzo (con la cosiddetta Direttiva Open Data).

Il tema della gestione dei dati, in ogni fase del processo connesso all'adozione, acquisizione e sviluppo di soluzioni di IA (training, documenti di ancoraggio, output, etc.) nel contesto pubblico, non può prescindere da un necessario coordinamento con la politica e le strategie già in atto a livello europeo (delineate innanzi) e nazionale in tema di dati e dati aperti e che la corretta gestione dei dati prodotti, sotto il profilo tecnico e negoziale, va accompagnata dalla capacità di acquisire e di riutilizzare dataset di terzi, a diverso titolo acquisiti, al fine di preservare la filiera di riutilizzabilità del dataset stesso.

In questo senso, assumono ancora più importanza le strategie di *licensing* connesse all'acquisizione, elaborazione, rilascio/pubblicazione e riutilizzo dei dataset pubblici, per cui l'omogeneizzazione delle licenze adottate e da adottarsi da parte del comparto pubblico introdotta, in termini di regolamentazione, con le "Linee Guida recanti regole tecniche per l'apertura dei dati e il riutilizzo dell'informazione del settore pubblico" (nel seguito linee guida Open Data)[44], diventa una imprescindibile necessità, evitando la parcellizzazione – specie per i dati ad alto valore di cui al Regolamento di esecuzione (UE) 2023/138.

Inoltre, oggi, con il diffondersi dell'analisi computazionale e di training, la selezione della licenza va orientata, come la produzione di un dato tecnicamente di qualità, anche con l'obiettivo di permetterne la più

[43] Dati di alta qualità e l'accesso a dati di alta qualità svolgono un ruolo essenziale nel fornire una struttura e garantire le prestazioni di molti sistemi di IA, in particolare quando si utilizzano tecniche che prevedono l'addestramento di modelli, al fine di garantire che il sistema di IA ad alto rischio funzioni come previsto e in maniera sicura e che non diventi una fonte di discriminazione vietata dal diritto dell'Unione.
[44] Adottate da AgID con la Determinazione del Direttore Generale n. 183/2023 ai sensi dell'art. 12 del D. Lgs. n. 36/2006

ottimale riutilizzabilità e la interpretabilità nella forma più trasparente anche da parte dei nuovi sistemi di intelligenza artificiale già disponibili o in corso di realizzazione.

Anche le clausole contrattuali già ipotizzate, per es. nella guida operativa sui dati di elevato valore, funzionali all'acquisizione di dataset prodotti da soggetti terzi, anche nel caso affidatari di attività in outsourcing, dovranno essere integrate con la previsione di una riutilizzabilità anche ai fini di un training o di un ancoraggio o comunque l'utilizzo per la produzione di output di sistemi di IA.

9.1. Tipologie di dati

Come detto in premessa, i dati costituiscono l'aspetto più rilevante e allo stesso tempo più critico nell'adozione di IA non solo nella PA, costituendo la base del funzionamento di qualsiasi progetto di IA.

Gli stessi dati prodotti dalla PA possono essere utilizzati per più scopi e condivisi con più utenti interni o esterni; allo stesso modo la PA può utilizzare dati provenienti da fonti esterne all'organizzazione.

I dati da utilizzare e/o utilizzati per l'IA possono essere caratterizzati a seconda di diversi parametri considerati per l'analisi. Considerando lo scenario in continua evoluzione, ad esempio in relazione agli algoritmi di machine learning che potrebbero originare altre caratteristiche specifiche per le informazioni, non è possibile produrre un catalogo completo e definitivo delle tipologie di dati per i sistemi di IA. Le classificazioni che seguono rappresentano, di conseguenza, una sintesi delle tipologie attualmente rilevate e potranno subire integrazioni o modifiche (es. la tendenza di definire data model standardizzati condurrà, auspicabilmente, alla riduzione o rimozione di dati non strutturati). In sintesi, la tecnologia così come la società progredisce, quindi anche la base di conoscenza su cui addestrare gli algoritmi deve essere aggiornata nel tempo.

Ai fini della costruzione del modello e di deployment del sistema di IA, l'AI Act identifica:

- **dati di addestramento**, definiti come quei dati utilizzati per addestrare un sistema di IA adattandone i parametri che può apprendere (cfr. art. 3, paragrafo 1, punto 29));
- **dati di convalida**, definiti come quei dati utilizzati per fornire una valutazione del sistema di IA addestrato e per metterne a punto, tra l'altro, i parametri che non può apprendere e il processo di apprendimento, al fine tra l'altro di evitare lo scarso (underfitting) o l'eccessivo (overfitting) adattamento ai dati di addestramento (cfr. art. 3, paragrafo 1, punto 30));
- **dati di prova**, definiti come quei dati utilizzati per fornire una valutazione indipendente del sistema di IA al fine di confermarne le prestazioni attese prima della sua immissione sul mercato o messa in servizio (cfr. art. 3, paragrafo 1, punto 32)).

Con riferimento alle fonti di dati, possono essere distinti:

- **dati interni** alla PA, cioè dati nativi generati dalle amministrazioni nell'adempimento delle proprie funzioni istituzionali, dati frutto di collaborazioni con altri soggetti oppure derivanti da operazioni di integrazione con altre fonti di dati;

- **dati esterni** alla PA, acquisiti tramite procedure di appalti oppure resi disponibili come open data da altre organizzazioni oppure ancora condivisi sulla base di specifiche norme (ad es. Data Governance Act) oppure derivanti da altre piattaforme (social media, dispositivi IoT, ecc.).

Se si considera la struttura del dato, si distinguono:

- **dati strutturati**, organizzati secondo schemi e tabelle rigide e relazionate definiti possibilmente da specifici data model (ad es. per i dati geografici le "Specifiche di contenuto sui database geotopografici"). Nel caso di tali dati, specialmente se basati su data model condivisi, possono essere previste operazioni semplificate di data cleaning;
- **dati non strutturati**, che, pur contenendo una grande quantità di informazione, non seguono un modello specifico di riferimento (es. immagini, video, file audio, documenti di testo, ecc..). In tali casi le operazioni di data cleaning risultano complesse (riduzione delle dimensioni, della complessità, dell'ambiguità e miglioramento dell'accuratezza, della completezza e dell'usabilità) considerando che il pre-processamento deve anche prevedere una trasformazione dei dati in un formato utilizzabile dagli algoritmi di IA;
- **dati semi-strutturati**, che contengono informazioni con caratteristiche ibride (formati XML, JSON, ecc..) e che sono caratterizzati da alcune proprietà organizzative che ne facilitano l'analisi, ossia contengono anche informazioni aggiuntive come metadati o tag che li rendono più organizzati rispetto ai dati non strutturati.

A seconda della fase di utilizzo/produzione dei dati, si possono avere:

- **dati di input**, definiti dall'AI Act come quei dati forniti a un sistema di IA o direttamente acquisiti dallo stesso, in base ai quali il sistema produce un output (cfr. art. 3, paragrafo 1, punto 33). Possono essere dati di addestramento, convalida o prova del modello (v. quanto indicato innanzi) oppure dati di pre-elaborazione;
- **dati di output**, che possono essere distinti in:
  - dati generati dal modello (testo, immagine, audio)
  - dati di performance
  - predizioni e relative probabilità ed affidabilità
- **dati del sistema di IA**, relativi ai seguenti aspetti:
  - parametri del modello;
  - metadati;
  - codice sorgente del modello;
  - dati sulla performance;
  - dati di log;
  - dati di stato;
  - dati temporanei.

Occorre tenere conto anche che l'impiego di tecniche di AI in ottica predittiva è comunque inevitabilmente associato all'utilizzo di **dati storici**, che rappresentano una sequenza di informazioni relative ad un evento rilevato ad intervalli regolari nel tempo. I dati storici costituiscono un elemento essenziale per l'implementazione di algoritmi statistici e tecniche di machine learning finalizzati all'analisi predittiva di molteplici tipologie di eventi. L'AI Act evidenzia che, specie nel caso di utilizzo di tali dati, le distorsioni possono, ad esempio, essere intrinseche ai set di dati di base.

Sono poi da tenere in considerazione i cosiddetti **dati sintetici**, creati artificialmente per l'addestramento dei sistemi di IA. L'efficacia dei dati sintetici deriva dalla difficoltà oggettiva, in molte situazioni, di raccogliere, strutturare, elaborare e validare dati reali o rispettare i requisiti di privacy. Ciò può frenare l'implementazione dei sistemi di IA. L'uso dei dati sintetici permette di addestrare ed eventualmente correggere gli algoritmi dei sistemi di IA prima che i dati del mondo reale siano disponibili. Per tali dati punti di attenzione possono riguardare l'effettiva protezione dei dati personali, la garanzia della rappresentatività rispetto ai dati reali e, quindi dell'appropriatezza delle proprietà statistiche richiamata anche dall'AI Act, e l'effettiva utilità di tali dati nell'implementazione dei sistemi di IA.

Chiaramente le classificazioni dei dati di cui sopra non sono da ritenersi a sé stanti, nel senso che, ad esempio, se consideriamo dati di input questi possono essere dati strutturati, non strutturati o semi-strutturati e contestualmente essere utilizzati per l'addestramento o la convalida o la prova.

Un ruolo cruciale nella disponibilità costante di dati può essere svolto dagli **spazi comuni di dati** previsti nella Strategia europea per i dati che ne ha delineato il percorso di creazione in una serie di settori strategici al fine di sfruttare al meglio il valore dei dati a vantaggio dell'economia e delle attività sociali degli stati membri. Tali spazi comuni, che nel tempo hanno allargato gli ambiti di interesse e che rappresentano elemento essenziale del mercato unico dei dati, possono facilitare la condivisione e la messa in comune di dati affidabili e sicuri in settori economici strategici e in ambiti di interesse pubblico.

Attualmente, gli spazi comuni europei di dati riguardano 14 domini: agricoltura, patrimonio culturale, energia, finanza, patto ecologico, salute, lingua, manifatturiero, media, mobilità, PA, ricerca e innovazione, abilità e turismo.

Il primo degli spazi comuni su cui si è avviata una regolamentazione specifica è quello dei dati sanitari, per i quali è stato recentemente raggiunto un accordo tra Parlamento europeo e Consiglio sulla proposta di Regolamento su tale spazio[45].

9.2. Caratteristiche dei dati

Come richiamato dalla Strategia Italiana per la IA 2024-2016, i dataset diventano anch'essi infrastrutture essenziali e trasversali a tutti gli ambiti di sviluppo e sfruttamento delle nuove potenzialità della IA della PA. È, quindi, necessario agire in forma coordinata e omogenea valorizzando prima di tutto la conoscenza utile a

[45] Proposta di REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO sullo spazio europeo dei dati sanitari

costituire, preservare tecnicamente e giuridicamente e pubblicare dataset di qualità ed effettivamente riutilizzabili.

Come azione propedeutica, le PA DOVREBBERO guidare una azione sistematica di analisi sullo stato e la qualità dei propri dataset, agendo per l'aggiornamento e la pubblicazione, ove non ancora realizzata, dei dati pubblicabili come open data, secondo le indicazioni delle linee guida Open Data, nonché la Guida operativa sulle serie di dati di elevato valore, rilasciando i dati secondo la CC BY 4.0 (o, se possibile, la CC0) o altra licenza altrettanto permissiva, come ad esempio la CDLA 2.0 permissive, che già prende in considerazione l'analisi computazionale in forma espressa.

Tale azione è anche in linea con il principio 2.1 del documento di raccomandazioni OCSE[46] che spinge i governi a investimenti anche in set di dati aperti che siano rappresentativi e rispettino la privacy e la protezione dei dati, per sostenere un ambiente per la ricerca e lo sviluppo dell'IA privo di pregiudizi dannosi e per migliorare l'interoperabilità e l'uso degli standard.

Inoltre, considerata la significativa quantità di dataset non pubblicabili come open data (in quanto, ad esempio, contenenti dati personali o dati protetti da qualche privativa), è fondamentale procedere a costruire un framework sicuro e comune per attuare nei limiti del possibile le disposizioni del Data Governance Act (Regolamento (UE) 2022/868). Sarà fondamentale in questo senso procedere nei termini previsti dal D. Lgs. n. 144/2024, strutturando percorsi di supporto ed accompagnamento agli enti pubblici nella definizione degli strumenti tecnici e contrattuali previsti.

I sistemi di IA apprendono e producono analisi affidabili se hanno a disposizione una adeguata quantità di dati di elevata qualità. È necessario, altresì, che sia garantito il rispetto delle normative e della regolamentazione sulla protezione dei dati e la sicurezza, aspetti che sono trattati in specifici paragrafi del presente documento.

Per quanto riguarda la qualità, le linee guida Open Data forniscono indicazioni specifiche facendo riferimento agli Standard UNI CEI ISO/IEC 25012:2014 e UNI CEI ISO/IEC 25024:2016. Tali Standard definiscono un insieme di caratteristiche per la qualità (e relative misure) suddivise in "inerenti" (accuratezza, aggiornamento (attualità), completezza, consistenza (coerenza), credibilità), "inerenti e dipendenti dal sistema" (accessibilità, comprensibilità, conformità, efficienza, precisione, riservatezza, tracciabilità) e "dipendenti dal sistema" (disponibilità, portabilità e ripristinabilità).

Con riferimento a tali caratteristiche, le linee guida Open Data, anche richiamando la Determinazione Commissariale n. 68/2013 di AgID, raccomandano che siano garantite almeno quattro caratteristiche di qualità dei dati, ovvero accuratezza, coerenza, completezza e attualità (cfr. par. 5.3 delle linee guida Open Data). Il rispetto di altre caratteristiche di qualità, rilevanti anche ai fini dei sistemi di IA, è garantito per via di obblighi

[46] Recommendation of the Council on Artificial Intelligence

derivanti da specifiche norme, come l'accessibilità (di cui alla legge n.4/2004 e alle relative Linee Guida AgID) o la riservatezza correlata alle indicazioni derivanti dal GDPR.

Le indicazioni riportate sono valide anche nel contesto dell'IA, atteso che l'approccio che si vuole seguire non è quello di definire un nuovo modello e nuovi processi relativi alla qualità dei dati piuttosto di integrare gli elementi, le misure e le procedure specifici per l'IA.

A partire, quindi, dalle raccomandazioni contenute nelle linee guida Open Data, considerando anche le caratteristiche di qualità individuate dall'AI Act, per i sistemi di IA DOVREBBERO essere garantite le caratteristiche di qualità dei dati di seguito riportate e derivanti dagli standard ISO/IEC 25012 e ISO/IEC 5259-2.

1. **Rappresentatività:** è il grado in cui i dataset riflettono la popolazione oggetto di studio (ISO/IEC 5259-2).
2. **Bilanciamento:** si riferisce alla distribuzione dei campioni per tutte le caratteristiche dei dataset (ISO/IEC 5259-2).
3. **Tracciabilità:** rappresenta il grado in cui i dati hanno attributi che forniscono una registrazione degli accessi ai dati e a tutte le modifiche effettuate ai dati in un contesto di utilizzo specifico (ISO/IEC 25012).
4. **Disponibilità:** è il grado in cui i dati hanno attributi che ne consentono il richiamo da parte di utenti autorizzati e/o applicazioni in un contesto di uso specifico. I dati devono essere archiviati in modo organizzato e strutturato, con sistemi adeguati a recuperare rapidamente le informazioni necessarie per l'addestramento e l'implementazione degli algoritmi (ISO/IEC 25012).
5. **Credibilità:** è il grado in cui i dati hanno attributi considerati veri e credibili da parte degli utilizzatori in uno specifico contesto d'uso (ISO/IEC 25012).

La quantità dei dati è un altro fattore importante per addestrare efficacemente i sistemi di IA. In particolare, essa è rilevante nei Big Data, definiti nello standard ISO/IEC 20546:2019 come insiemi di dati estesi - principalmente con caratteristiche di volume, varietà, velocità e/o variabilità - che richiedono una tecnologia scalabile per l'archiviazione, la manipolazione, la gestione e l'analisi efficienti. Per i big data gli Standard evidenziano 4 caratteristiche: 2 di struttura e 2 di qualità. Le caratteristiche di struttura sono il volume e la variabilità delle sorgenti. Grandi volumi di dati e la loro variabilità possono richiedere l'uso di strumenti automatizzati, come indicato nella norma ISO/IEC 5259-1.

In sintesi, relativamente al volume, i sistemi di IA DOVREBBERO essere alimentati da grandi quantità di dati per addestrare modelli complessi. Analogamente, i dati DEVONO altresì includere una varietà di esempi rappresentativi delle possibili situazioni che i sistemi di AI si trovano a valutare.

Un altro aspetto rilevante è "fornire informazioni chiare e di facile comprensione sulle fonti di dati/ingressi, sui fattori, sui processi e/o sulla logica che hanno portato alla previsione, al contenuto, alla raccomandazione o alla decisione, per consentire a coloro che sono interessati da un sistema di IA di comprendere l'output" come ribadito nel principio 1.3 "Transparency and explainability" del documento di

raccomandazioni OCSE innanzi citato. A tale proposito, la Strategia Italiana per l'IA 2024-2026 ha previsto un programma mirato alla definizione di un registro di dataset e modelli, che siano costruiti secondo principi di trasparenza e fairness, che siano eticamente affidabili by design e che siano riusabili per accelerare le soluzioni delle aziende italiane. Secondo la Strategia, tutti i progetti finanziati nell'ambito della stessa strategia nazionale o comunque che riceveranno finanziamenti pubblici saranno tenuti a riportare i dataset utilizzati e i modelli prodotti nel registro, in accordo a linee guida che definiranno i livelli di accesso e le modalità di riuso.

In relazione a questo, si segnala che nell'ambito dell'azione SEMIC della Commissione Europea è in corso di definizione un modello semantico nel campo dell'apprendimento automatico, il cui obiettivo è estendere l'uso di DCAT-AP (il profilo di metadati utilizzato per i dati aperti). Il modello MLDCAT-AP[47], attualmente allo stato "Candidate Recommendation", facilita le descrizioni standardizzate di un processo di apprendimento automatico, insieme ai relativi set di dati, alla qualità misurata sui set di dati e alla citazione dei documenti.

Oltre che nel senso più tipico di dati messi "al servizio" dell'IA, rileva anche la relazione inversa, di IA quale strumento migliorativo della qualità dei dati stessi, del loro livello di rappresentatività, completezza e interoperabilità. Sotto questo aspetto è possibile armonizzare classificazioni e descrizioni di item di medesima tipologia, standardizzare, bonificare e migliorare l'accuratezza degli archivi, estrarre termini chiave da utilizzare successivamente per la ricerca, modellare e rappresentare gli argomenti trattati da un corpus testuale e in generale sintetizzare e riuscire a governare un patrimonio informativo che per sua natura si manifesta in gran parte non strutturato e che attraverso l'IA può essere organizzato e "dominato".

Un ulteriore aspetto in questo ambito riguarda la necessità e la possibilità, anche tramite l'ausilio di AI, di rendere i dati tra loro "parlanti", quindi, in altri termini interoperabili e integrati, estendendo la base di conoscenza e abilitando grandi opportunità per i dati stessi di rigenerarsi e di esprimere le loro potenzialità, in aggiunta a quelle che già possono avere con un utilizzo limitato ai "silos" all'interno dei quali sono stati prodotti o acquisiti.

9.3. Processi e governance dei dati

Per una corretta gestione dei dati nell'adozione di soluzioni di AI nella PA, è necessario delineare un processo strutturato e responsabile che copra l'intero ciclo di vita del dato, garantendo che raccolta, archiviazione, elaborazione, analisi, monitoraggio e aggiornamento siano sicuri e conformi alle normative vigenti. È necessario, inoltre, fornire indicazioni per garantire qualità e disponibilità dei dati, protezione dei dati personali e sicurezza dei dati, resilienza contro bias e distorsioni, adozione di pratiche etiche e trasparenti e conformità continua alle normative in evoluzione.

A questo si aggiunge che l'azione delle PA non può prescindere da (o derogare a) una serie di valori, principi e anche adempimenti quali quelli di apertura, trasparenza, accountability e del perseguimento di una

[47] https://semiceu.github.io/MLDCAT-AP/releases/2.0.0/

maggiore efficienza accompagnata da una non minore efficacia (per esempio, aumentando le disuguaglianze), nonché da una verifica di economicità complessiva, tenuto conto anche degli impatti ambientali.

Tutto questo non dovrebbe prefigurare, tuttavia, la definizione di nuovi processi di gestione dei dati per l'IA separati dalle procedure seguite (finora) nell'organizzazione. Deve piuttosto essere definito con chiarezza (se non ancora codificato e realizzato) un processo strutturato di gestione dei dati più ampio che sovrintenda il ciclo di vita del dato a prescindere dall'utilizzo e in cui siano integrate le procedure e le pratiche relative al contesto dell'IA. La gestione dei dati ai fini dei sistemi di IA, quindi, può essere considerata come un sottoinsieme della gestione complessiva dei dati.

Le Linee guida Open Data includono una sezione relativa agli aspetti organizzativi in cui è definito un percorso di preparazione dei dati (v. figura 2 delle medesime Linee Guida) come frutto di una catena di processi e di una serie di attività di analisi ed elaborazione finalizzate al miglioramento della qualità e dell'accesso al dato stesso, che poi sono i requisiti di base richiesti per i sistemi di IA.

Nell'ottica, indicata innanzi, di considerare un processo generale di gestione dei dati considerando, in aggiunta, gli aspetti e le caratteristiche relativi all'IA, il percorso delineato dalle Linee Guida Open Data richiamate può essere opportunamente revisionato e adattato in modo che ricomprenda fasi e attività specifiche per l'IA. A tale scopo, sebbene riferiti ai sistemi di IA ad alto rischio, possono essere considerati gli aspetti indicati dall'AI Act all'art. 10 relativi alle pratiche di governance e gestione dei dati, cioè:

a) le scelte progettuali pertinenti;

b) i processi di raccolta dei dati e l'origine dei dati, nonché la finalità originaria della raccolta nel caso di dati personali;

c) le operazioni di trattamento pertinenti ai fini della preparazione dei dati, quali annotazione, etichettatura, pulizia, aggiornamento, arricchimento e aggregazione;

d) la formulazione di ipotesi, in particolare per quanto riguarda le informazioni che si presume che i dati misurino e rappresentino;

e) una valutazione della disponibilità, della quantità e dell'adeguatezza dei set di dati necessari;

f) un esame atto a valutare le possibili distorsioni suscettibili di incidere sulla salute e sulla sicurezza delle persone, di avere un impatto.

Alcuni degli aspetti indicati sono già o possono essere contemplati nelle varie fasi previste nel processo menzionato innanzi. Le attività di cui ai punti b) e d), per esempio, possono essere considerati nelle fasi indicate come 'identificazione' e 'analisi'; quelle di cui al punto c) nelle fasi denominate 'analisi' e 'arricchimento'.

Allo stesso modo può essere considerato il ciclo di vita dei dati descritto nella norma ISO/IEC 8183, che considera 10 fasi (per ciascuna di esse sono riportate alcune specifiche):

- **Fase 1**. Concezione dell'idea (obiettivo di business e metriche);
- **Fase 2.** Requisiti di business (obiettivi, strategia, requisiti di business e utente, conformità);

- **Fase 3**. Pianificazione dei dati (quantità di dati necessari, fonte, dati sintetici, formato, sicurezza, privacy);
- **Fase 4**. Acquisizione dei dati (da fonti interne, terze parti, dati aperti);
- **Fase 5**. Preparazione dei dati (pulizia, trasformazione, normalizzazione, organizzazione dei dati, etichettatura, ricampionamento, codifica, verifica dell'integrità, provenienza, anonimizzazione o pseudonimizzazione dei dati);
- **Fase 6**. Costruzione del modello (addestramento di un algoritmo di ML, combinazione di conoscenze umane);
- **Fase 7**. Implementazione del sistema (il sistema di IA entra in funzione nell'ambiente di destinazione);
- **Fase 8**. Funzionamento del sistema (analisi dei dati, visualizzazione dei dati, trasmissione dei dati, archiviazione dei dati);
- **Fase 9**. Dismissione dei dati (cancellazione sicura, archiviazione, riutilizzo, conservazione per l'audit);
- **Fase 10**. Disattivazione del sistema (cessazione dell'elaborazione dei dati, conservazione dei log).

Anche in questo caso alcune delle fasi indicate sono sovrapponibili o comunque contemplabili nel processo indicato nelle linee guida Open Data: le attività previste nelle fasi 3 e 4 della ISO/IEC 8183 possono essere considerate nelle fasi delle linee guida identificate con "identificazione" e "analisi"; quelle previste nella fase 5 della ISO/IEC 8183 possono essere considerate nelle fasi delle linee guida identificate con "analisi", "arricchimento", "modellazione e documentazione" e "validazione".

Se ne ricava che un modello di processo di gestione complessiva dei dati, adattato, come detto, da quello delle linee Guida Open Data con le integrazioni derivanti dall'AI Act e dagli Standard ISO citati, potrebbe essere come riportato in Figura 6. Le fasi identificate con celle con sfondo grigio sono quelle relative ad attività e operazioni finalizzate esclusivamente all'IA.

Figura 6. Modello di processo di gestione dei dati.

Come evidente, diverse sottofasi sono generali e, quindi, implementabili per diverse finalità, non solo per l'IA, a seconda dei procedimenti di interesse (pubblicazione dati aperti, gestione dei dati protetti, adozione di sistemi di IA, ecc.).

Di seguito un approfondimento delle fasi e delle attività indicate nello schema precedente.

Fase di Identificazione:

L'attività di pianificazione è relativa alla definizione di una strategia di raccolta e gestione dei dati che garantisca la qualità, la varietà e la pertinenza dei dati per l'obiettivo specifico di volta in volta perseguito. L'obiettivo è, quindi, quello di identificare le esigenze di dati necessari per i vari scopi e a programmare le attività necessarie all'utilizzo degli stessi (per esempio relativamente alla sicurezza). Nel caso della gestione dei dati per l'IA, essa include: quantità di dati necessari, fonte, necessità di dati sintetici (vedi ISO/IEC 8183, fase 3).

L'attività di raccolta include sia la ricognizione di dati (già prevista nelle linee guida Open Data) derivanti da fonti interne ossia prodotti internamente all'amministrazione o comunque ivi gestiti (ad esempio, derivanti da operazioni di mashup), sia l'acquisizione di dati provenienti da altre fonti, come API, sensori IoT, social media, immagini, dati pubblici/open data di altre PA (vedi ISO/IEC 8183, fase 4).

L'attività di analisi dei vincoli è finalizzata all'individuazione delle barriere che impediscono l'utilizzo dei dati per un certo scopo. Nel caso dell'IA, per esempio, verifica della finalità originaria della raccolta nel caso di dati personali.

Fase di Analisi:

L'attività di analisi della qualità è finalizzata a verificare il livello di qualità originario dei dati raccolti al fine di procedere ad opportune operazioni di pulizia e bonifica.

L'attività di pulizia/bonifica è finalizzata al miglioramento della qualità dei dati attraverso la rimozione o la correzione di duplicati, errori e valori mancanti utilizzando processi basati sui dati o adottando azioni di bonifica basate su processi (vedi ISO/IEC 8183, fase 5).

Come già indicato nelle linee guida Open Data, l'analisi del processo organizzativo che produce e gestisce il dato è necessaria per fare in modo che la produzione di quel dato sia consolidata e diventi stabile, secondo la frequenza di aggiornamento e le modalità di rilascio adottate.

Fase di Preparazione:

Nella fase di preparazione sono state incluse le fasi identificate come "Arricchimento" e "Modellazione e documentazione" (riferibili alle operazioni di normalizzazione e organizzazione dei dati) previste nelle linee guida Open Data (a cui si rimanda per il dettaglio), con l'aggiunta di alcune operazioni specifiche di pre-processing finalizzate a rendere i dati adeguati all'addestramento dei modelli di IA. Tali operazioni possono includere trasformazione, ricampionamento, etichettatura, codifica, anonimizzazione o pseudonimizzazione dei dati (vedi ISO/IEC 8183, fase 5).

Fase di Validazione:

Nell'attività di valutazione della qualità, già prevista nelle linee guida Open Data, sono considerate le operazioni di verifica dell'integrità (vedi ISO/IEC 8183, fase 5) e quelle relative alla valutazione di possibili distorsioni, uno degli aspetti indicato dall'AI Act.

Fase di Utilizzazione:

Nella fase di utilizzazione sono incluse le attività di pubblicazione finalizzate a garantire la disponibilità dei dati e di conservazione dei dati (che deve avvenire in modo sicuro e scalabile, comunemente tramite infrastrutture cloud che garantiscano scalabilità elastica, archiviazione distribuita, sicurezza e crittografia dei dati), sono state incluse le fasi 6, 7 e 8 previste dallo standard ISO/IEC 8183.

Fase di monitoraggio:

Il monitoraggio costante della qualità e della freschezza dei dati, che per loro natura non sono statici, in particolare quando si opera in contesti di IA, deve essere accompagnato da revisioni e aggiornamenti periodici per tenere conto di eventuali cambiamenti nel dominio, nei comportamenti rilevati o nelle relative fonti di raccolta. Sono poi da considerare le operazioni di dismissione dei dati (vedi ISO/IEC 8183, fase 9), come la cessazione dell'elaborazione dei dati e conservazione dei log.

In sintesi, le PA DEVONO garantire un'efficace e robusta gestione del ciclo di vita dei dati in cui devono essere presidiati in particolare i seguenti processi:

- **Raccolta e Archiviazione**: I dati DEVONO essere raccolti e archiviati in modo sicuro e conforme alle normative vigenti.
- **Elaborazione e Analisi**: I dati DEVONO essere elaborati e analizzati utilizzando tecniche avanzate per garantirne l'integrità e la qualità.
- **Monitoraggio e Aggiornamento**: I dataset DEVONO essere aggiornati e monitorati mediante dei processi di governance per mantenere la rilevanza e l'accuratezza delle analisi.

La gestione dei dati, anche nel contesto dell'adozione di soluzioni di IA, non può prescindere da una corretta impostazione della loro governance, con l'individuazione chiara di ruoli e relative responsabilità. Anche questo aspetto è trattato, relativamente a quell'ambito, nelle Linee Guida Open Data, in cui si richiama la Circolare n. 3 del 1° ottobre 2018 del Ministro per la PA che raccomanda di prevedere, per il Responsabile per la Transizione al Digitale (RT), anche il potere di costituire gruppi tematici per singole attività e/o adempimenti. Nel processo di apertura e pubblicazione dei dati, pertanto, si raccomanda di costituire, all'interno dell'organizzazione dell'Ente, un apposito gruppo di lavoro dedicato al processo di apertura dei dati, prevedendo, ove possibile, le strutture e le figure adatte e necessarie a tale scopo, tenendo in considerazione i referenti dei singoli domini e prevedendo altresì il necessario coinvolgimento del Responsabile della protezione dei dati laddove siano coinvolti dati personali.

L'individuazione di apposite strutture di coordinamento o gruppi di lavoro esistenti o da istituire, anche all'interno dell'ufficio del RTD, è prevista anche nella Direttiva del Sottosegretario di Stato alla Presidenza del Consiglio dei ministri con delega di funzioni in materia di innovazione tecnologica e transizione digitale, concernente "*Misure per l'attuazione dell'articolo 50-ter del decreto legislativo 7 marzo 2005, n. 82*"[48]. Tra i compiti da affidare a tali strutture, secondo la Direttiva, è incluso anche il conseguimento di obiettivi di data governance e di razionalizzazione delle banche dati esistenti interne alla PA, a garanzia dell'univocità e della qualità del dato, e la promozione della condivisione del patrimonio informativo detenuto dalle PA.

Con l'obiettivo di evitare la proliferazione di strutture organizzative dedicate ai dati, è auspicabile che la governance dei dati nelle PA, anche in relazione ai dati per l'IA sia ricondotta ad un'unica struttura di coordinamento che sovrintenda alla loro gestione complessiva come innanzi delineata (e che quindi copra tutti gli aspetti indicati: open data, attuazione art. 50-ter, dati per IA, ...).

In questo contesto, è rilevante anche il tema della collaborazione istituzionale trattato anche nel Piano Triennale. Analogamente a quanto indicato nel Piano, anche nel processo di gestione dei dati alcune amministrazioni possono svolgere il ruolo di coordinamento (hub nazionali e/o regionali), come intermediarie, attuatrici e promotrici di soluzioni verso gli Enti locali, in ascolto dei loro bisogni e complementari rispetto alle disomogenee sensibilità e capacità espresse dalle diverse realtà locali.

Dette amministrazioni assumono un ruolo di intermediazione e raccordo verso i livelli territoriali inferiori, per le possibilità tecnologiche, la capacità di spesa e le economicità possibili grazie alla propria scala territoriale più ampia. Anche nell'ambito della gestione dei dati, si ritiene fondamentale il ruolo svolto dalle amministrazioni sovraordinate in qualità di accentratori di infrastrutture, servizi e contratti per la raccolta e la gestione, a cura e con il presidio diretto di ciascun ente del territorio, di un patrimonio informativo pubblico integrato a supporto dell'IA. Tale approccio garantisce la maggiore ampiezza delle soluzioni, una gestione omogenea e uniforme, economie di scala nelle fasi di raccolta e pre-trattamento dei dati.

Questo può tradursi nella promozione di iniziative locali per la valorizzazione dei dati, come la creazione di data center o di soluzioni di High Performance Computing interregionali, la promozione di piattaforme hub open data, lo sviluppo di partenariati con enti locali, università, centri di ricerca e aziende per progetti di analisi e utilizzo dei dati, l'investimento in programmi di formazione e sviluppo di competenze nel settore dei dati per dipendenti pubblici e cittadini.

Una preliminare mappatura del livello di maturità e dei fabbisogni delle diverse amministrazioni del territorio riguardo alle dimensioni suddette è certamente utile a suggerire sviluppi omogenei su tali dimensioni, individuando scenari e strategie che il livello di governance sovraordinato può percorrere e suggerire al territorio, o assumere direttamente ove opportuno. Tale mappatura è peraltro utile a stimolare una consapevolezza e una volontà di adottare una strategia di gestione e valorizzazione dei dati, che in talune situazioni vengono necessariamente raccolti per finalità tipicamente amministrative, ma non sono poi utilizzati o lo sono a

[48] https://www.governo.it/sites/governo.it/files/Decreto20231205_Direttiva_PDND.pdf

compartimenti stagni, secondo approcci innovativi che sono guidati, secondo una logica "bottom-up" da parte di poche figure che ne hanno compreso individualmente i vantaggi.

L'intervento delle amministrazioni capofila, in rapporto alla realtà degli Enti presenti sul territorio, può avere in definitiva a riguardo una serie di aspetti, tra i quali:

- l'implementazione e la gestione centralizzata di infrastrutture e strumenti, secondo una logica di tipo "cloud", da mettere a disposizione di ciascun Ente, accentrando una serie di investimenti e di competenze e realizzando evidenti economie di scala (è di rilievo in questo ambito anche la possibilità di federare risorse computazionali tra livello centrale, regionale e inter-regionale, come possibile prospettiva di messa in condivisione di risorse di computing della PA italiana, in logica di cooperazione istituzionale);
- la disponibilità stessa dei dati, che sovente possono essere di interesse trasversale e della cui raccolta e condivisione, quindi, il livello superiore può farsi aggregatore ed orchestratore a beneficio di tutto il territorio, garantendo standard di qualità e aggiornamento omogenei e anche in tal caso generando evidenti economie;
- la messa a disposizione di strumenti contrattuali, in qualità di soggetto aggregatore, dedicati alla realizzazione di progettualità aventi ad oggetto la valorizzazione e il riutilizzo di dati. Per ciascun Ente locale, singolarmente preso, può essere difficoltoso, oneroso e non efficace attivare procedure pubbliche di selezione di fornitori dedicati a questa tipologia di servizi innovativi, dando peraltro luogo potenzialmente a iniziative non coordinate;
- l'attivazione di iniziative di stimolo verso gli Enti per l'introduzione di nuove modalità di gestione, trattamento e valorizzazione del patrimonio informativo pubblico, che possano incentivare il loro impegno e la loro partecipazione attiva e che possono essere basate su una combinazione di contributi economici, supporto tecnico, formazione e sensibilizzazione;
- la previsione di sinergie tra gli enti, i quali possono stringere accordi per condividere mezzi, risorse umane e competenze necessarie a portare avanti progettualità di comune interesse.

Si presume che iniziative di stimolo quali quelle indicate favoriscano l'introduzione di pratiche più strutturate ed avanzate di gestione e valorizzazione dei dati da parte degli Enti locali, aiutando le amministrazioni a diventare più efficienti, trasparenti e in grado di offrire ulteriori servizi nell'ambito delle rispettive competenze, vedendo ciascuna amministrazione come inserita in un "ecosistema amministrativo digitale", in un network collaborativo tra enti pubblici, aziende private e cittadini, finalizzato alla condivisione e alla valorizzazione dei dati, che può portare a soluzioni innovative in vari settori, quali ad esempio quello dei trasporti, dell'energia o della gestione ambientale.

10. Protezione dei dati personali

I sistemi di IA, nel contesto delle presenti Linee guida, possono essere utilizzati per lo svolgimento e/o il supporto di attività che comportano il trattamento di dati anche personali: l'adozione di tali sistemi da parte delle PA, pertanto, DEVE intervenire nel rispetto del diritto fondamentale alla protezione dei dati personali.

In tale ottica, al momento dell'adozione, la PA DEVE porre un'attenzione primaria alla valutazione e alla verifica della conformità normativa del sistema di IA e all'impatto che questo avrà sui diritti degli interessati coinvolti, anche adeguando il proprio modello organizzativo e le correlate misure tecniche e organizzative di sicurezza.

La PA, in particolare, DEVE verificare il rispetto dei principi enunciati all'art. 5 del GDPR, compresa la responsabilizzazione della PA stessa che ne farà uso in qualità di titolare del trattamento, nel rispetto della normativa unionale e nazionale in materia di protezione dei dati personali e dei provvedimenti e pareri emessi dall'European Data Protection Board e dal Garante per la protezione dei dati personali.

Qualora il sistema di IA sia adottato per lo svolgimento di attività che comportano il trattamento di dati personali, risulta di primaria rilevanza che la PA effettui un'analisi di tale sistema sotto lo specifico profilo della protezione dei dati personali, al fine di garantire e dimostrare il rispetto di quanto segue[49]:

- il trattamento dei dati personali mediante il sistema di IA avviene in modo lecito, corretto e trasparente nei confronti dell'interessato;
- i dati personali sono raccolti mediante il sistema di IA per finalità determinate, esplicite e legittime e, successivamente, sono trattati compatibilmente con tali finalità;
- i dati personali trattati a mezzo del sistema di IA sono adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per cui sono trattati;
- i dati personali trattati a mezzo del sistema di IA sono esatti e, se necessario, aggiornati, adottando la PA tutte le misure ragionevoli per cancellare o rettificare tempestivamente i dati inesatti rispetto alle finalità per cui sono trattati;
- i dati personali trattati a mezzo del sistema di IA sono conservati in una forma che consenta l'identificazione degli interessati per un arco di tempo non superiore al conseguimento delle finalità per cui sono trattati;
- la PA utilizza il sistema di IA in modo da garantire un livello di sicurezza dei dati personali adeguato al rischio, individuando e attuando misure tecniche e organizzative adeguate a proteggere i dati da

49 Sugli aspetti oggetto di trattazione, si vedano più estesamente i provvedimenti dello European Data Protection Board e del precedente Article 29 Working Party nonché e del Garante per la protezione dei dati personali.
Fra questi, a titolo di primario riferimento nel contesto in esame, si richiamano ad esempio le "*Linee guida sul processo decisionale automatizzato relativo alle persone fisiche e sulla profilazione ai fini del regolamento 2016/679*", adottate dall'Article 29 Working Party il 3 ottobre 2017, nella versione emendata e adottata in data 6 febbraio 2018, consultabili al link: https://ec.europa.eu/newsroom/article29/items/612053/en.

violazioni di sicurezza che possano comportare, illecitamente o accidentalmente, la perdita, la modifica, la divulgazione non autorizzata, l'accesso ai dati personali trasmessi, conservati o comunque trattati;

- la PA garantisce sempre all'interessato l'esercizio dei propri diritti in materia di protezione dei dati personali;
- la PA istruisce e designa il personale a cui attribuisce specifici compiti e funzioni connessi al trattamento di dati personali a mezzo del sistema di IA adottato;
- qualora intenda determinare finalità e mezzi del trattamento congiuntamente ad altro soggetto o utilizzare altro soggetto per il trattamento dei dati personali mediante il sistema di IA, la PA agisce sempre ai sensi degli artt. 26 e 28 del GDPR;
- nell'utilizzo del sistema di IA, la PA garantisce la protezione dei dati personali fin dalla progettazione e per impostazione predefinita;
- prima di procedere al trattamento di dati personali mediante il sistema di IA adottato e altresì periodicamente, la PA effettua una valutazione d'impatto sulla protezione dei dati personali ai sensi dell'art. 35 del GDPR, delle Linee guida dell'Article 29 Working Party[50] e dei provvedimenti del Garante per la protezione dei dati personali[51], individuando i rischi e le misure tecniche e organizzative idonee a mitigarli e, qualora risulti un rischio elevato in assenza di misure di attenuazione, consulta il Garante per la protezione dei dati personali;
- le categorie particolari di dati personali, come individuate agli artt. 9-10 del GDPR, sono trattate a mezzo di sistemi di IA solo nel rispetto di quanto stabilito dalla normativa unionale e nazionale in materia di protezione dei dati personali e ai sensi dello stesso AI Act.

Nel contesto oggetto delle presenti Linee guida, in ogni caso, è necessario che la PA che adotta sistemi di IA per lo svolgimento di proprie attività sia ben conscia di rivestire il ruolo di titolare del trattamento dei dati personali ai sensi dell'art. 4, n. 7) del GDPR e, pertanto, di avere la responsabilità (anche in ottica di responsabilizzazione, ai sensi dell'art. 5, par. 2, del GDPR) del trattamento effettuato tramite lo strumento di IA adottato, anche nei casi in cui l'utilizzo intervenga per il tramite del fornitore di riferimento.

Come evidenziato dal Garante per la protezione dei dati personali, inoltre, la PA DEVE porre la massima attenzione a tre principi cardine che devono necessariamente governare l'utilizzo di algoritmi e sistemi di IA nell'esecuzione di compiti di rilevante interesse pubblico:

[50] Si vedano le "*Linee guida in materia di valutazione d'impatto sulla protezione dei dati e determinazione della possibilità che il trattamento 'possa presentare un rischio elevato' ai fini del regolamento (UE) 2016/679*" dell'Article 29 Working Party del 4 aprile 2017, come modificate e adottate da ultimo il 4 ottobre 2017 e fatte proprie dall'European Data Protection Board il 25 maggio 2018, consultabili al link: https://ec.europa.eu/newsroom/article29/items/611236/en.

[51] Si veda il provvedimento del Garante per la protezione dei dati personali n. 467 dell'11 ottobre 2018, recante l'"*Elenco delle tipologie di trattamenti soggetti al requisito di una valutazione d'impatto sulla protezione dei dati ai sensi dell'art. 35, comma 4, del Regolamento (UE) n. 2016/679*", al link: https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9058979.

"*1. il principio di conoscibilità, in base al quale l'interessato ha il diritto di conoscere l'esistenza di processi decisionali basati su trattamenti automatizzati e, in tal caso, di ricevere informazioni significative sulla logica utilizzata, sì da poterla comprendere;*

2. il principio di non esclusività della decisione algoritmica, secondo cui deve comunque esistere nel processo decisionale un intervento umano capace di controllare, validare ovvero smentire la decisione automatica (c.d. human in the loop);

3. il principio di non discriminazione algoritmica, secondo cui è opportuno che il titolare del trattamento utilizzi sistemi di IA affidabili che riducano le opacità, gli errori dovuti a cause tecnologiche e/o umane, verificandone periodicamente l'efficacia anche alla luce della rapida evoluzione delle tecnologie impiegate, delle procedure matematiche o statistiche appropriate per la profilazione, mettendo in atto misure tecniche e organizzative adeguate. Ciò, anche al fine di garantire, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori, visti i potenziali effetti discriminatori che un trattamento inesatto di dati sullo stato di salute può determinare nei confronti di persone fisiche (cfr. considerando n. 71 del Regolamento)"[52].

Su specifici aspetti di protezione dei dati personali trattati nel contesto dei modelli di IA, si è recentemente espresso altresì lo European Data Protection Board con il parere n. 28 del 17 dicembre 2024, utile strumento di indirizzo per la PA in merito alla natura dei modelli di IA in relazione alla definizione di dato personale, alle circostanze in cui i modelli di intelligenza artificiale potrebbero essere considerati anonimi e alla relativa dimostrazione, all'adeguatezza dell'interesse legittimo come base giuridica per il trattamento dei dati personali nel contesto dello sviluppo e dell'implementazione dei modelli di IA e al possibile impatto di un trattamento illecito di dati personali nello sviluppo di un modello di IA sulla liceità del successivo trattamento o funzionamento del modello di IA[53].

[52] Si veda il "*Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*" al link https://www.garanteprivacy.it/documents/10160/0/Decalogo+per+la+realizzazione+di+servizi+sanitari+nazionali+attraverso+sistemi+di+Intelligenza+Artificiale.pdf/a5c4a24d-4823-e014-93bf-1543f1331670?version=2.0. Si ritiene di chiarire che, sebbene il citato provvedimento sia specificamente rivolto al contesto sanitario, i principi cardine sopra citati sono individuati come necessari per una consapevole e corretta gestione di algoritmi e sistemi di IA nell'esecuzione di compiti di rilevante interesse pubblico.

[53] Si veda il parere n. 24/2024 dell'EDPB, consultabile al link: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.

11. Sicurezza cibernetica

La sicurezza dei sistemi di IA è un requisito essenziale per l'adozione, l'acquisizione e lo sviluppo dell'IA nella PA. Se da un lato, infatti, l'IA sembra poter fornire utili strumenti per rispondere alla crescente necessità di migliorare l'efficienza e l'efficacia nella gestione e nell'erogazione dei servizi pubblici[54], dall'altro, questa tecnologia introduce nuovi rischi, minacce e vulnerabilità che devono essere presi in considerazione nella sua implementazione.

La sicurezza di un sistema di IA può essere definita[55] come l'insieme di strumenti, strategie e processi implementati per identificare e prevenire le minacce e gli attacchi che potrebbero compromettere la riservatezza, l'integrità o la disponibilità di un modello di IA o di un sistema abilitato all'IA.

In considerazione della loro natura inerentemente **socio-tecnica** – in cui elementi sociali (l'influenza delle dinamiche sociali e l'impatto sulle persone che li usano o che ne sono condizionati) e tecnici (quali dataset, algoritmi, modelli) risultano strettamente intrecciati tra loro – i sistemi di IA sono caratterizzati da specifiche peculiarità nell'ambito della sicurezza in particolare con riferimento ai *rischi* ai quali sono soggetti e agli *attacchi* dei quali sono vittima.

11.1. Tassonomie di attacco

In aggiunta ai tradizionali attacchi cyber[56] all'infrastruttura ICT ospitante, i sistemi di IA sono soggetti ad attacchi a componenti specifiche dell'IA quali, ad esempio, i modelli e i dati di addestramento. La conoscenza delle varie tassonomie di attacco permette di attuare azioni di protezione e contenimento mirate e contestualizzate.

Come tassonomia di riferimento per gli attacchi ai sistemi di IA può essere usata quella sviluppata dal NIST[57] che prevede le seguenti macro-categorie di attacchi:

- evasion attacks;
- poisoning attacks;
- privacy attacks;
- abuse attacks;

Le prime tre categorie riguardano sia i modelli di IA predittiva che quelli di IA generativa, mentre l'ultima solo i modelli di IA generativa[58].

Nei successivi paragrafi sono discussi brevemente queste categorie e possibili strategie di mitigazione.

[54] Agenzia per l'Italia Digitale (AGID), Piano Triennale per l'informatica nella Pubblica Amministrazione - edizione 2024 -2026, aggiornamento 2025, https://www.agid.gov.it/it/agenzia/piano-triennale.

[55] La definizione è tratta dall'articolo pubblicato sul sito di MITRE ATLAS all'indirizzo https://atlas.mitre.org/resources/ai-security-101. MITRE ATLAS è una *knowledge base* delle tattiche e tecniche avversarie relative ai sistemi di IA.

[56] Attacchi nei quali l'attore malevolo fa uso di TTP (tattiche, tecniche e procedure che caratterizzano il comportamento di un attaccante per raggiungere il proprio obiettivo) tipiche del dominio cyber.

[57] National Institute of Standards and Technology (NIST), Adversarial Machine Learning – Taxonomy and Terminology of Attacks and Mitigations, 2024, .

[58] I sistemi di IA predittiva identificano schemi e relazioni o fanno previsioni sulla base dei dati di addestramento, i sistemi di IA generativa creano nuovi contenuti sulla base dei dati di addestramento.

11.1.1. Evasion attacks

Questa categoria di attacchi ha come obiettivo quello di generare un errore nella classificazione del modello introducendo perturbazioni (spesso impercettibili per l'uomo) negli *input* del modello, denominate *adversarial examples* (esempi avversari).

Questi attacchi sono progettati per sfruttare le vulnerabilità nel processo decisionale del modello. Possono indurre il modello a prevedere un valore desiderato dall'attaccante o determinare una riduzione dell'accuratezza del modello[59].

Come possibili strategie di mitigazione, è possibile applicare l'**advesarial training**, riaddestrando il modello con *adversarial examples* etichettati correttamente, il **randomized smoothing** e il **formal verification**, con i quali si cerca di rendere invariante il modello all'eventuale rumore introdotto da un avversario aumentandone la robustezza.

11.1.2. Poisoning attacks

Questa categoria di attacchi ha come obiettivo degradare le prestazioni di un modello o fargli generare uno specifico risultato alterando (*avvelenando*) i dati di addestramento del modello. Un esempio di attacco consiste nel cosiddetto *label flipping* nel quale l'attaccante cambia l'etichetta dei dati di addestramento con l'obiettivo di addestrare il modello sulla base dell'etichetta da lui scelta[60].

Questi attacchi possono essere distinti in:

- **availability poisoning**: determinano una violazione della disponibilità del modello tramite una degradazione delle sue prestazioni su tutti i *sample* di dati. Sono rilevabili monitorando le prestazioni del modello, possibili strategie di mitigazione prevedono la sanitizzazione dei dati di addestramento e l'apprendimento robusto (possono, ad esempio, essere addestrati molteplici modelli);
- **targeted poisoning**: determinano una violazione dell'integrità del modello alterandone la previsione su un numero ridotto di *sample* mirati. Possibili strategie di mitigazione prevedono l'implementazione di controlli di sicurezza sull'origine e sull'integrità dei dati;
- **backdoor poisoning**: analogamente agli attacchi *targeted poisoning* determinano una violazione dell'integrità del modello, in questo caso tuttavia l'obiettivo è indurre in errore il modello in risposta a uno specifico *sample* di dati (denominato *trigger*). Possibili strategie di mitigazione prevedono la sanitizzazione dei dati di addestramento, la ricostruzione del *trigger*, l'ispezione e la sanitizzazione del modello;
- **model poisoning**: modificano direttamente il modello addestrato iniettandogli funzionalità malevole. Possono determinare una violazione sia dell'integrità che della disponibilità e avvengono

[59] Ad esempio, in *K. Eykholt et al., "Robust Physical-World Attacks on Deep Learning Visual Classification," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 1625-1634, doi: 10.1109/CVPR.2018.00175* è stato mostrato come sia possibile attraverso questa categoria di attacchi indurre in errore un modello addestrato nel riconoscere la segnaletica stradale, come i segnali di stop, apponendovi dei piccoli adesivi.

[60] Ad esempio, nell'ambito dell'addestramento di un filtro anti-spam della posta elettronica, un avversario potrebbe cambiare le etichette dei dati di training da *spam* a *no-spam*, per indurre il modello addestrato a non filtrare correttamente i messaggi di posta elettronica che contengono spam.

generalmente nell'ambito del cosiddetto apprendimento federato in cui sistemi *client* inviano aggiornamenti locali del modello a un *server* che li aggrega in un modello globale. Sono inoltre possibili in scenari di *supply chain* nei quali sono acquisiti modelli o relativi componenti che sono stati *avvelenati*. Possibili strategie di mitigazione prevedono l'individuazione ed esclusione degli aggiornamenti malevoli o (nel caso di avvelenamenti del modello tramite *backdoor*) l'ispezione e la sanitizzazione del modello.

11.1.3. Privacy attacks

Questa categoria di attacchi ha come obiettivo compromettere le informazioni degli utenti *ricostruendole* a partire dai dati di addestramento. Possono essere distinti in:

- **data reconstruction**, ricostruiscono le informazioni a partire dalle informazioni aggregate;
- **membership inference**, determinano se un particolare *record* è stato incluso nel dataset utilizzato per l'addestramento di un modello, compromettendo le informazioni dell'utente;
- **model extraction**, ottengono informazioni estraendo informazioni sul particolare modello utilizzato, come la sua architettura o i suoi parametri.
- **property inference**, accedono a informazioni globali sulla distribuzione dei dati di addestramento interagendo con il modello.

Per la mitigazione degli attacchi di ricostruzione è stato proposto di far uso di tecniche per migliorare la privacy come la *privacy differenziale*, che – attraverso opportune manipolazioni dei dati – permette di fissare un limite su quanto un attaccante, con accesso ai risultati dell'algoritmo, può inferire su ogni singolo *record* del dataset.

Possibili strategie di mitigazione prevedono inoltre limitazioni al numero di interrogazioni dell'utente al modello, e di rilevamento delle interrogazioni sospette al modello.

11.1.4. Abuse attacks

Questa categoria di attacchi ha come obiettivo alterare il comportamento di un sistema di IA generativa per adattarlo ai propri scopi come, ad esempio, perpetrare frodi, distribuire malware e manipolare informazioni.

Possibili strategie di mitigazione prevedono l'uso di metodi come l'apprendimento rinforzato con *feedback* umano, il filtraggio degli input o il rilevamento di valori anomali di output (*outliers*).

11.2. Gestione del rischio cibernetico

Come osservato nel paragrafo 11.1, i sistemi di IA sono sistemi caratterizzati da specifici rischi la cui gestione rappresenta un elemento imprescindibile per lo sviluppo e l'utilizzo responsabile dell'IA

L'AI Act prevede, all'articolo 15, che i sistemi di IA ad alto rischio siano progettati e sviluppati in modo tale da conseguire un adeguato livello di accuratezza, robustezza e sicurezza cibernetica e operare in modo coerente con tali aspetti durante tutto il loro ciclo di vita.

Un'efficace strutturazione di un processo di gestione dei rischi di un sistema di IA deve necessariamente considerare le caratteristiche distintive di tali sistemi. A tale scopo è possibile fare riferimento a framework e standard specializzati, quale, ad esempio, il Risk Management Framework per l'IA (AI RMF)[61] sviluppato dal *National Institute of Standards and Technology (NIST)*.

Il framework è disegnato per assistere organizzazioni e individui, definiti come *Ai Actors*[62], e fornisce un approccio strutturato per identificare, valutare e mitigare i rischi associati ai sistemi di IA. L'obiettivo è quello di minimizzare i potenziali impatti negativi e massimizzare quelli positivi in modo da avere sistemi di IA affidabili.

Potenziali impatti negativi derivanti dall'uso dei sistemi di IA sono classificati a seconda della tipologia di attore coinvolto:

- **impatti sugli individui**, come ad esempio gli impatti sulle libertà civili, sulla sicurezza fisica/psicologica o sulla sfera economica di un individuo;
- **impatti sulle organizzazioni**, come ad esempio gli impatti sulle attività commerciali, sulla reputazione o derivanti dalla compromissione di un'organizzazione;
- **Impatti sull'ecosistema**, come ad esempio gli impatti su elementi e risorse interconnessi e interdipendenti, sul sistema finanziario, sulla catena di approvvigionamento o sulle risorse naturali e l'ambiente.

Il cosiddetto *Core* del framework individua le seguenti quattro funzioni (a loro volta suddivise in categorie e sottocategorie) per supportare le organizzazioni nella gestione dei rischi posti dai sistemi di IA:

- **GOVERN**: promuovere una cultura di gestione del rischio all'interno delle organizzazioni che progettano, sviluppano, acquisiscono e adottano sistemi di IA.
- **MAP**: stabilire il contesto nel quale inquadrare i rischi di un sistema di IA, identificare i rischi e i relativi fattori di rischio;
- **MEASURE**: analizzare, valutare, confrontare e monitorare il rischio e i relativi impatti. Utilizza le informazioni acquisite dalla precedente funzione e fornisce indicazioni a quella successiva;
- **MANAGE:** definire le priorità e ad agire sui rischi identificati. Il trattamento del rischio comprende piani di risposta, recupero e comunicazione di incidenti o eventi.

La funzione GOVERN è trasversale e si applica a tutte le fasi del processo di gestione del rischio. Le funzioni MAP, MEASURE e MANAGE comprendono componenti della funzione GOVERN (in particolare quelle riguardanti la conformità o la valutazione) e possono essere utilizzate in contesti specifici e in determinate fasi del ciclo di vita dei sistemi di IA.

[61] National Institute of Standards and Technology (NIST), Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023, https://doi.org/10.6028/NIST.AI.100-1.

[62] L'organizzazione per la cooperazione e lo sviluppo economico (OCSE) definisce gli *AI Actors* come coloro i quali svolgono un ruolo attivo nel ciclo di vita dei sistemi di IA, compresi le organizzazioni e gli individui che distribuiscono o gestiscono l'IA.

Per supportare le organizzazioni nell'uso del Framework, il NIST ha sviluppato un playbook[63] allineato a ciascuna sottocategoria delle quattro funzioni.

L'individuazione dei rischi avviene tipicamente a partire dalle minacce cui può essere esposto un sistema e dai suoi asset, sui quali tali minacce tentano di sfruttarne le vulnerabilità.

In ragione di ciò, nei seguenti paragrafi sono elencati, sulla base del modello del ciclo di vita e categorie di asset e minacce relative ai sistemi di IA.

11.3. Asset

Un sistema di IA è costituito da un insieme di *asset*. Per asset si intende tutto ciò che ha un valore per un individuo o un'organizzazione e che quindi deve essere protetto. In aggiunta agli asset caratteristici dell'IA (come, ad esempio, i modelli e gli iperparametri) sono qui considerati anche gli asset dell'infrastruttura ICT (come, ad esempio, le reti di comunicazione e i sistemi operativi)

Gli asset di un sistema di IA possono essere categorizzati[64] in (tra parentesi è riportata la corrispondente fase del ciclo di vita del sistema di IA):

- **dati**, come ad esempio: dati grezzi (acquisizione dei dati), dati di valutazione (tuning del modello), dati etichettati (pre-elaborazione dei dati), dati di test (addestramento del modello);
- **modelli**, come ad esempio: algoritmi (addestramento del modello), iperparametri (tuning del modello), algoritmi di addestramento (selezione del modello);
- **attori**, come ad esempio: fornitori di servizi cloud (raccolta dei dati, addestramento del modello, tuning del modello), proprietari dei dati (definizione dell'obiettivo, raccolta dei dati, esplorazione dei dati), fornitori dei dati (raccolta dei dati);
- **processi**, come ad esempio: acquisizione dei dati (raccolta dei dati), etichettamento dei dati (pre-elaborazione dei dati), comprensione dei dati (esplorazione e convalida dei dati);
- **strumenti**, come ad esempio: reti di comunicazione (raccolta dei dati), database (raccolta dei dati), sistemi operativi (distribuzione del modello, mantenimento del modello), interfacce (utente e di gestione) e API (esterne);
- **artefatti**, come ad esempio: politiche di gestione dei dati (raccolta dei dati), architettura del modello (selezione del modello, distribuzione del modello), casi d'uso (comprensione del business).

È fondamentale effettuare una mappatura sistemica degli asset critici che compongono il sistema di IA. Ciò permette di tracciare e sorvegliare le risorse più rilevanti, come i dati, i modelli, le infrastrutture e i processi operativi, garantendo che siano adeguatamente protetti lungo tutte le fasi del ciclo di vita del sistema.

63 NIST AI RMF Playbook https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

64 La tassonomia completa è riportata nell'allegato A del documento di ENISA "Artificial Intelligence Cybersecurity Challenges" https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challengesLa tassonomia completa è riportata nell'annesso A di [3].

11.4. Minacce

Ogni fase del ciclo di vita è caratterizzata da una o più minacce[65]. In accordo a quanto indicato nel documento di ENISA "Artificial Intelligence Cybersecurity Challenges"[66] possono essere individuate le seguenti categorie di minacce:

- **attività malevole/abuso**: azioni intenzionali che prendono di mira sistemi, infrastrutture e reti ICT mediante atti malevoli con l'obiettivo di sottrarre, alterare o distruggere un obiettivo specifico (ad esempio[67] data poisoning e backdoors nel modello);
- **eavesdropping/ intercettazioni/ hijacking**: azioni volte a intercettare, interrompere o prendere il controllo di una comunicazione di terzi senza consenso (ad esempio divulgazione del modello e furto di dati);
- **attacchi fisici**: azioni che mirano a distruggere, esporre, alterare, disattivare, rubare o ottenere un accesso non autorizzato a risorse fisiche come infrastrutture, hardware o interconnessioni consenso (ad esempio sabotaggio del modello e DDOS).
- **danno non intenzionale**: azioni non intenzionali che causano distruzione, danni a proprietà o persone e che si traducono in un guasto o in una riduzione dell'utilità (ad esempio riduzione dell'accuratezza dei dati e compromissione della selezione delle caratteristiche);
- **guasti o malfunzionamenti**: funzionamento parziale o totalmente insufficiente di un asset hardware o software (ad esempio scarsità di dati e degradazione delle performance del modello);
- **interruzioni**: interruzioni impreviste del servizio o diminuzione della qualità al di sotto di un livello richiesto (ad esempio interruzione dell'infrastruttura/sistemi, interruzione delle reti di telecomunicazione);
- **disastro**: incidente improvviso o catastrofe naturale che causa ingenti danni (ad esempio disastri naturali e fenomeni di cambiamento climatico);
- **legale**: azioni legali di terzi (ad esempio a causa di divulgazione di informazioni personali e profilazione degli utenti).

Con l'evoluzione continua delle tecnologie AI, nuove vulnerabilità e rischi possono emergere, spesso in modo imprevedibile. Pertanto, le organizzazioni devono dotarsi di processi e strumenti per individuare tempestivamente le minacce emergenti, analizzarle e adattare le loro strategie di difesa. Il monitoraggio delle minacce emergenti non riguarda solo le tecniche di attacco tradizionali, ma include anche il rilevamento di nuovi

[65] Una medesima minaccia può far riferimento a più fasi del ciclo di vita.

[66] Negli annessi B e D del documento di ENISA "Artificial Intelligence Cybersecurity Challenges" https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges sono riportati, rispettivamente, la tassonomia completa delle minacce e la mappatura con le fasi del ciclo di vita. Per ogni categoria di minacce sono altresì individuati le specifiche minacce, le dimensioni di sicurezza potenzialmente impattate e gli asset impattati. Negli annessi B e D del documento di ENISA sono riportati, rispettivamente, la tassonomia completa delle minacce e la mappatura con le fasi del ciclo di vita. Per ogni categoria di minacce sono altresì individuati le specifiche minacce, le dimensioni di sicurezza potenzialmente impattate e gli asset impattati.

[67] Il paragrafo 11.1 tratta le varie categorie di attacchi associati a questo tipo di minacce.

tipi di exploit basati su progressi tecnologici, cambiamenti normativi o nuove modalità di interazione tra i sistemi.

11.5. Obiettivi di sicurezza

In questa sezione sono enunciati gli obiettivi di sicurezza che devono guidare l'adozione dell'IA da parte della PA.

Gli obiettivi sono mutuati dalle *Linee guida per uno sviluppo sicuro dell'IA*[68] promosse dal *National Cyber Security Centre del Regno Unito (NCSC)* e alle quali hanno aderito 35 Agenzie di 18 Paesi, tra le quali l'Agenzia per la Cybersicurezza Nazionale[69].

Le raccomandazioni fornite per il raggiungimento dei già citati obiettivi sono da intendersi come pratiche di base, le cui implicazioni devono essere comprese e valutate attentamente in fase di realizzazione.

Resta in capo a ciascuna PA la valutazione, in esito alla differente esposizione alle minacce e alla propria analisi del rischio, in merito all'individuazione e conseguente raggiungimento di ulteriori obiettivi di sicurezza per il rafforzamento della sicurezza cibernetica dei propri sistemi di IA.

1. **Adottare l'IA in modo responsabile.** Modelli, applicazioni e sistemi di IA devono essere adottati dopo essere stati sottoposti alla valutazione di sicurezza. Dovrebbero essere inoltre indicati chiaramente agli utenti le eventuali limitazioni note, i potenziali guasti, gli aspetti di sicurezza di cui sono responsabili e come (e dove) i loro dati potrebbero essere utilizzati, consultati o conservati (ad esempio, se sono utilizzati per riaddestrare il modello o se siano revisionati da addetti della PA o terze parti).
2. **Identificare, tracciare, mantenere e proteggere gli asset.** Devono essere previsti processi e strumenti per identificare, tracciare, mantenere e proteggere gli asset[70] dei sistemi di IA prevedendo, ad esempio, specifici controlli per la gestione e protezione dei dati e dei contenuti generati dai sistemi di IA. Devono essere inoltre implementati controlli per proteggere la riservatezza, l'integrità e la disponibilità dei log dei sistemi di IA.
3. **Proteggere la catena di approvvigionamento.** La sicurezza della catena di approvvigionamento (supply chain) dei vari componenti dei sistemi di IA (come, ad esempio, dati e modelli, librerie software, API esterne, ecc.) deve essere valutata e monitorata durante l'intero ciclo di vita dei sistemi di IA. I componenti devono essere acquisiti da fornitori verificati e devono essere adeguatamente protetti e documentati. È fondamentale analizzare la gestione dei dati da parte delle terze parti, assicurandosi che siano implementate misure di sicurezza adeguate, come la crittografia e i controlli di accesso. Inoltre, è necessario richiedere ai fornitori che le loro misure di sicurezza siano allineate con le politiche di sicurezza adottate e con l'analisi dei rischi effettuata.

[68] National Cyber Security Centre (NCSC), Guidelines for secure AI system development.
[69] Le linee guida sono pubblicate sul sito di ACN all'indirizzo https://www.acn.gov.it/portale/linee-guida-ia.
[70] Per l'elenco delle categorie di asset si faccia riferimento al paragrafo 11.3.

4. **Proteggere il modello e i dati.** I modelli e i dati dei sistemi di IA devono essere protetti da accessi impropri o manomissioni. Un attaccante può essere in grado di ricostruire la funzionalità di un modello o i dati su cui è stato addestrato, accedendo direttamente ad essi (acquisendo i pesi del modello) o indirettamente (interrogando il modello tramite un'applicazione o un servizio). Può inoltre manomettere i modelli, i dati o le richieste durante o dopo la fase di addestramento del modello, rendendone il risultato non affidabile. È possibile proteggere il modello e i dati dall'accesso adottando le best practice di cybersecurity e implementando controlli sull'interfaccia di interrogazione del modello per rilevare e prevenire i tentativi di accesso, modifica ed esfiltrazione delle informazioni.
5. **Monitorare il comportamento del sistema e degli input.** Devono essere monitorati i risultati e le prestazioni del modello e del sistema di IA in modo da poter osservare e rispondere a cambiamenti improvvisi e graduali del comportamento che influiscono sulla sicurezza come, ad esempio, potenziali intrusioni e compromissioni. In linea con i requisiti di protezione dei dati, anche personali, gli input ai sistemi di IA (come, ad esempio, le richieste di inferenza e le interrogazioni) devono essere monitorati e registrati per consentire verifiche di conformità, audit, analisi e ripristino in caso di compromissione o uso improprio.
6. **Sviluppare un piano di risposta agli incidenti.** Deve essere definito e attuato un piano di risposta agli incidenti che riflette i diversi scenari di minaccia. Il piano deve essere revisionato periodicamente e al verificarsi di eventi interni (come, ad esempio, l'aggiornamento di piani strategici o modifiche organizzative), eventi esterni (come l'evoluzione del conteso normativo e legislativo) o mutamenti dell'esposizione alle minacce e ai relativi rischi regolarmente valutati con l'evoluzione del sistema e della ricerca in generale.
7. **Formare e sensibilizzare il personale sulle minacce e sui rischi.** Gli attori responsabili dell'adozione dell'IA devono comprendere le minacce e le relative mitigazioni. I data scientist e gli sviluppatori devono essere formati allo sviluppo sicuro del software e alle pratiche di sistemi di IA sicuri e responsabili. Per approfondimenti sulla formazione e sullo sviluppo delle competenze fare riferimento al Capitolo 8 delle presenti linee guida.
8. **Proteggere l'infrastruttura ICT.** Le infrastrutture ICT che ospitano i sistemi di IA devono essere adeguatamente protette. Le vulnerabilità delle infrastrutture ICT possono essere infatti sfruttate da un attaccante per perpetrare attacchi al sistema di IA (come, ad esempio, la compromissione del modello o la degradazione delle sue prestazioni). Pertanto, devono essere adottate misure di cybersecurity adeguate ai connessi rischi sulle infrastrutture ICT utilizzate in ogni parte del ciclo di vita dei sistemi di IA. Per le misure di cybersecurity applicabili alle infrastrutture ICT è possibile far riferimento alle linee guida per il rafforzamento della resilienza di cui all'articolo 8 della legge 28 giugno 2024, n. 90 emanate da ACN.
9. **Proteggere le identità e gli accessi.** La gestione delle identità e degli accessi (IAM) è fondamentale per proteggere i sistemi di IA, in particolar modo quando si trattano dati sensibili. I principali rischi

associati includono accessi non autorizzati, furti di dati e manipolazioni dei modelli, che possono compromettere l'integrità e la riservatezza delle informazioni. Al fine di mitigare questi rischi e mantenere la sicurezza complessiva dei sistemi di IA, è fondamentale implementare controlli centralizzati che consentano una supervisione efficace degli accessi, definire chiaramente ruoli e permessi per garantire un accesso appropriato e utilizzare l'autenticazione multi-fattore (MFA) per aggiungere un ulteriore strato di protezione.

11.6. Gestione integrata della sicurezza dei sistemi di IA

Per perseguire in modo efficace gli obiettivi di sicurezza delineati nel paragrafo precedente è opportuno integrare la gestione della sicurezza dei sistemi di IA in un ciclo che comprenda almeno le seguenti fasi, rappresentate in Figura 7:

Figura 7. Ciclo per la gestione della sicurezza dei sistemi di IA

1 **Sicurezza in tutte le fasi del ciclo di vita dell'IA:** la sicurezza deve essere integrata come elemento centrale lungo tutto il ciclo di vita dei sistemi di IA, dalla definizione dell'obiettivo fino alla manutenzione del modello. Ciò implica assicurare gli obiettivi di sicurezza delineati nel precedente paragrafo lungo tutto il ciclo di vita. Le best practice includono la protezione dei dati, l'analisi delle minacce e l'implementazione di contromisure specifiche per ciascuna fase, come indicato nella tabella seguente. Di seguito sono riportate le pratiche di sicurezza per ciascuna fase del ciclo di vita di un sistema di IA.

1.1 *fase:* **definizione dell'obiettivo**

pratiche di sicurezza: stabilire criteri di sicurezza e privacy già in questa fase per garantire che l'obiettivo rispetto ai requisiti normativi e operativi.

1.2 *fase:* **raccolta, esplorazione e pre-elaborazione dei dati**

pratiche di sicurezza: applicare tecniche di anonimizzazione, pseudonimizzazione e crittografia sui dati sensibili. Verificare la sicurezza delle sorgenti esterne e l'affidabilità dei fornitori di dati. Garantire che i dati siano trattati in modo conforme ai principi di minimizzazione dei dati e

integrità. Proteggere l'accesso durante l'esplorazione e il trattamento. Assicurare che i dati siano protetti contro modifiche non autorizzate durante questa fase. Verificare la correttezza dell'anonimizzazione e rimozione dei dati sensibili.

1.3 *fase:* **selezione delle caratteristiche**

pratiche di sicurezza: garantire che le informazioni sensibili non siano esposte e che i dati trattati siano protetti contro accessi non autorizzati.

1.4 *fase:* **selezione del modello**

pratiche di sicurezza: valutare il modello scelto per identificare eventuali vulnerabilità note o potenziali. Assicurarsi che il modello selezionato abbia una robustezza adeguata contro attacchi adversarial (le vulnerabilità possono variare in base al modello, valutare di integrare i dettagli nelle sezioni apposite).

1.5 *fase:* **addestramento del modello**

pratiche di sicurezza: assicurarsi che i dati utilizzati per l'addestramento siano privi di contaminazioni o bias.

1.6 *fase:* **tuning del modello**

pratiche di sicurezza: controllare che le modifiche agli iperparametri non espongano il modello a nuove vulnerabilità. Proteggere il processo di tuning da accessi non autorizzati.

1.7 *fase*: **trasferimento dell'apprendimento**

pratiche di sicurezza: verificare la sicurezza dei modelli pre-addestrati e la loro provenienza per evitare l'inserimento di vulnerabilità attraverso il trasferimento di apprendimento. Applicare tecniche di protezione contro attacchi specifici ai modelli trasferiti.

1.8 *fase:* **distribuzione del modello**

pratiche di sicurezza: implementare sistemi di controllo degli accessi per la distribuzione e protezione del modello. Utilizzare certificati digitali per garantire l'integrità del modello distribuito.

1.9 *fase:* **manutenzione del modello**

pratiche di sicurezza: monitorare costantemente il comportamento del modello per rilevare eventuali anomalie o compromissioni. Utilizzare sistemi di logging per tracciare modifiche e accessi al modello.

1.10 *fase:* **valutazione del raggiungimento dell'obiettivo**

pratiche di sicurezza: effettuare un'analisi delle prestazioni del modello in ambienti controllati per assicurarsi che continui a operare in modo sicuro e affidabile.

2 **Audit regolari e controlli di conformità**: le organizzazioni devono effettuare audit interni ed esterni per verificare la conformità alle normative e alle policy aziendali, con particolare attenzione alla protezione dei dati, alla sicurezza del modello e al rispetto degli standard internazionali di sicurezza. Questi audit devono includere attività di test di penetrazione e simulazioni di attacco per valutare la resilienza del sistema.

3 **Revisione e aggiornamento continuo delle misure di sicurezza**: data la natura in rapida evoluzione delle minacce, le misure di sicurezza devono essere costantemente aggiornate. È necessario condurre valutazioni periodiche del rischio e adottare nuove tecnologie di protezione (come l'introduzione di nuovi algoritmi di crittografia o l'uso di tecniche di protezione avanzata contro attacchi adversarial).

Allegati

A. Valutazione del livello di maturità nell'adozione di IA

Il documento fornisce uno strumento per la valutazione del livello di maturità tecnica e organizzativa di una PA ai fini dell'adozione dell'IA (cfr. par. 4.2)

Il modello di maturità[71] proposto consente di valutare il livello di sviluppo di un'organizzazione attraverso un percorso strutturato in più dimensioni, ognuna delle quali rappresenta un aspetto cruciale per l'integrazione dell'IA nei processi dell'ente. Per ogni dimensione, il modello definisce vari livelli di maturità che permettono di identificare il grado di preparazione dell'ente e di guidare la sua evoluzione verso una piena capacità operativa nell'uso dell'IA.

Il modello si pone l'obiettivo di supportare le PA nel percorso di adozione dell'IA, identificando le aree chiave su cui intervenire e definendo gli elementi essenziali che caratterizzano i diversi stadi di maturità che gli enti possono raggiungere. Il modello nasce dall'analisi di numerosi casi di implementazione di IA in ambito pubblico, fornendo una metodologia solida e basata su evidenze empiriche per comprendere le aree su cui lavorare e le priorità da definire, al fine di agevolare la transizione da una PA tradizionale a una PA AI-ready, ovvero pronta all'uso dell'IA.

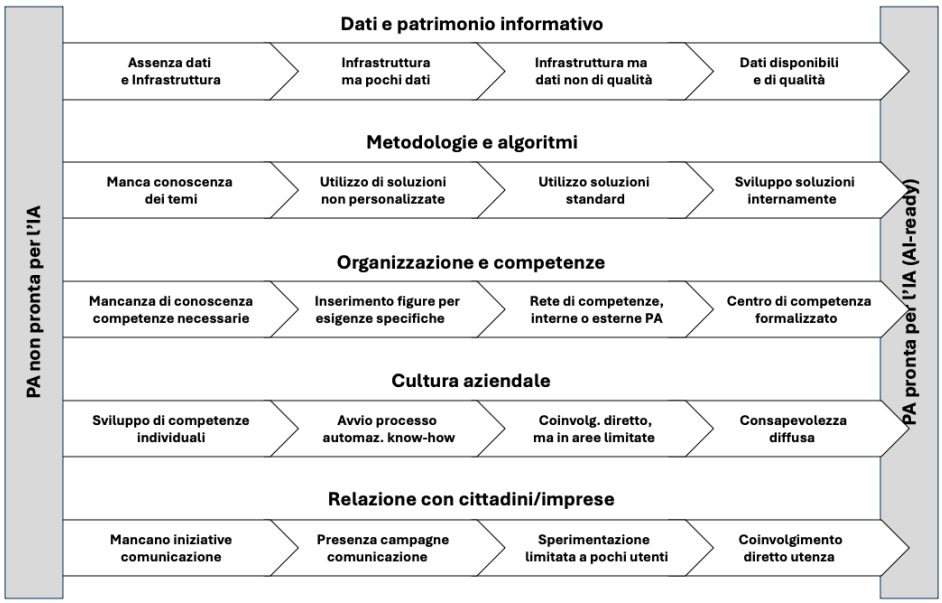

Figura A1. Modello di maturità per l'adozione di soluzioni di Intelligenza Artificiale in ambito pubblico

Come indicato in Figura A1, il modello si articola su cinque dimensioni, ognuna contraddistinta da quattro livelli di progressiva maturità:

1. *Dati e patrimonio informativo*: questa dimensione valuta la qualità e la disponibilità del patrimonio informativo dell'ente, indispensabile per lo sviluppo di soluzioni basate sull'IA; si suddivide in quattro stadi di maturità:

[71] Il modello di maturità è stato sviluppato dagli Osservatori Artificial Intelligence e Agenda Digitale del Politecnico di Milano

- *Assenza di dati e infrastruttura*: non esiste un'infrastruttura adeguata all'acquisizione e gestione dei dati necessari allo sviluppo di progetti di IA;
- *Infrastruttura presente ma dati insufficienti*: esiste un sistema di raccolta dati, ma il volume o la profondità temporale dei dati sono inadeguati;
- *Infrastruttura e dati presenti, ma di bassa qualità*: vi è una disponibilità di dati sufficiente, ma questi presentano carenze qualitative, come incompletezza o assenza di meta-informazioni rilevanti;
- *Dati disponibili e di alta qualità*: l'ente dispone di un'infrastruttura solida e di dati di alta qualità, adatti per l'uso dell'IA, sia per funzioni di routine sia di apprendimento.

2. *Metodologia e algoritmi*: questa dimensione riguarda la capacità dell'ente di sviluppare e applicare algoritmi e metodologie per progetti di IA. Si articola in:
   - *Mancanza di conoscenza sui temi*: l'ente non possiede le competenze necessarie ad avviare un progetto di Intelligenza Artificiale efficace;
   - *Utilizzo di soluzioni non personalizzate*: sono adottate soluzioni di mercato pronte all'uso, senza un adattamento alle specifiche necessità dell'ente;
   - *Utilizzo di soluzioni standard adattate*: l'ente impiega soluzioni di mercato, ma le adatta alle proprie esigenze attraverso competenze interne;
   - *Sviluppo di soluzioni ad hoc*: l'ente è in grado di sviluppare soluzioni Intelligenza Artificiale interne, con algoritmi e metodologie personalizzate, integrando eventualmente risorse esterne.
3. *Organizzazione e competenze*: questa dimensione esamina l'organizzazione interna e la disponibilità di competenze per la gestione di progetti di Intelligenza Artificiale:
   - *Assenza di competenze rilevanti*: l'ente non ha conoscenze o risorse specializzate per la gestione di soluzioni di Intelligenza Artificiale;
   - *Inserimento di figure specializzate per progetti pilota*: l'ente ha introdotto competenze specifiche per affrontare esigenze di un singolo progetto sperimentale;
   - *Rete di competenze diffuse*: esiste una rete di competenze, interne o esterne, che sono attivamente coinvolte nei processi di adozione di soluzioni di IA;
   - *Centro di competenza formalizzato*: l'ente ha formalizzato un centro di competenza dedicato all'Intelligenza Artificiale, con strutture organizzative e processi di coordinamento definiti.
4. *Cultura aziendale*: questa dimensione valuta il livello di consapevolezza e apertura dell'organizzazione verso l'integrazione dell'Intelligenza Artificiale nei processi lavorativi:
   - *Sviluppo di competenze individuali limitate*: l'organizzazione è focalizzata principalmente sulla gestione del core business, con scarsa attenzione all'innovazione tecnologica;
   - *Avvio dell'automazione del know-how*: l'automazione dei processi conoscitivi è avviata, ma il personale è coinvolto solo marginalmente;
   - *Coinvolgimento attivo del personale in aree limitate*: il personale è coinvolto attivamente nella sperimentazione di soluzioni di Intelligenza Artificiale, ma solo in ambiti ristretti;

- *Consapevolezza diffusa*: l'organizzazione è ampiamente consapevole dei benefici dell'Intelligenza Artificiale e i dipendenti sono disposti a rivedere le proprie mansioni per integrarsi con le nuove tecnologie.

5. *Relazione con cittadini e imprese*: questa dimensione misura il livello di interazione e coinvolgimento degli utenti finali (cittadini e imprese) nei progetti di Intelligenza Artificiale:
   - *Assenza di iniziative di comunicazione*: non esistono iniziative mirate a coinvolgere o informare gli utenti sui progetti di Intelligenza Artificiale;
   - *Presenza di campagne informative*: sono presenti campagne di comunicazione per informare i cittadini sull'utilizzo dell'Intelligenza Artificiale nei servizi pubblici;
   - *Sperimentazione limitata a cluster di utenti*: sono coinvolti gruppi ristretti di utenti in progetti di sperimentazione congiunta delle soluzioni di Intelligenza Artificiale;
   - *Coinvolgimento attivo e feedback diretto*: i cittadini partecipano attivamente al processo di innovazione dei servizi pubblici basato su soluzioni di Intelligenza Artificiale, fornendo feedback utili e contribuendo alla co-creazione dei servizi.

La struttura di tale strumento permette alle PA di autovalutarsi, o di essere valutate, consentendo così di mappare con precisione lo stato attuale di maturità e delineare gli obiettivi futuri per una piena integrazione dell'IA nei processi operativi e strategici.

Grazie a questa auto-diagnosi, le PA possono identificare con maggiore chiarezza le aree su cui concentrare gli sforzi e stabilire le priorità, per ciascuna dimensione, che consentano una crescita graduale ma mirata.

È fondamentale che, pur nella loro autonomia, le dimensioni siano valutate con una visione d'insieme per garantire un'adozione armonica dell'IA. Diversamente, si rischia che le azioni intraprese non siano coordinate, impedendo così una reale crescita equilibrata durante il percorso di adozione.

Ad esempio, in Figura A2, è rappresentata una PA di medie dimensioni carente sul fronte di organizzazione e competenze oltre a quello della cultura aziendale (profilo AS IS). Il modello aiuta a traguardare uno scenario di maturità a tendere (profilo TO BE) non troppo ambizioso ma realizzabile e suggerisce livelli omogenei di maturità sulle cinque dimensioni prima di immaginare future evoluzioni verso una PA AI-ready.

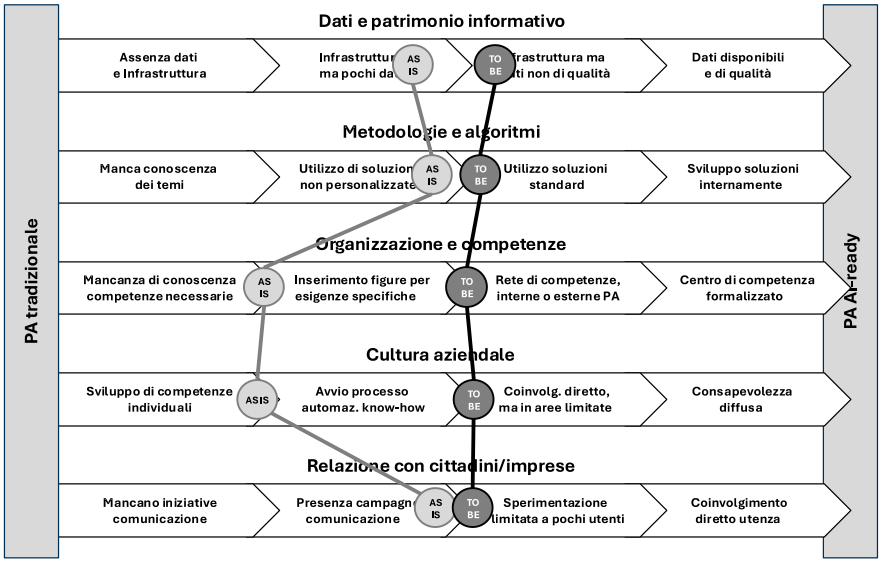

Figura A2. Esempio di applicazione del modello di maturità per l'adozione di soluzioni di IA in ambito pubblico

Un aspetto di notevole valore aggiunto del modello è la sua flessibilità: le cinque dimensioni sono potenzialmente autonome tra loro, consentendo di identificare per ciascuna di esse diversi livelli di maturità. Questo approccio permette di cogliere potenziali disomogeneità all'interno dello stesso ente, contribuendo ad allineare così visioni differenti a seconda delle competenze e dei ruoli dei vari soggetti coinvolti.

Ad esempio, in Figura A3, sono riportate le auto-mappature del livello di maturità di un ente da parte di tre diversi dipendenti pubblici che lavorano al suo interno. Se si rileva sostanziale omogeneità nella valutazione della maturità dell'ente sulle dimensioni relative a metodologie e algoritmi, organizzazione e competenze e cultura aziendale, il modello mostra che sulle dimensioni relative a dati e patrimonio informativo e relazione con cittadini/imprese si registrano opinioni molto diverse, probabilmente legate allo specifico ruolo giocato dai dipendenti pubblici. Prima di traguardare scenari a tendere sarebbe buona cosa allineare le varie visioni. È frequente che, come rappresentato in figura, ci siano diversi punti di vista soggettivi da riconciliare. Il modello aiuta a valutare l'eterogeneità dei diversi punti di vista e ad attivare riflessioni per una loro convergenza.

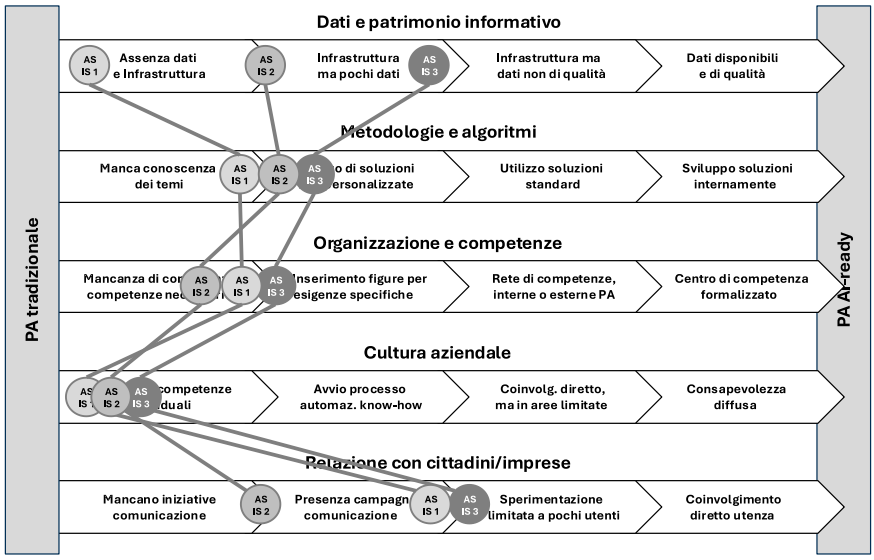

Figura A3. Esempio di applicazione del modello di maturità per l'adozione di soluzioni di IA in ambito pubblico

Un ulteriore aspetto rilevante è che il modello non prevede un percorso obbligatorio verso il massimo livello di maturità per ogni ente. Le piccole PA, con risorse limitate e minori complessità, potrebbero non necessitare di raggiungere livelli elevati in tutte le dimensioni, mentre grandi enti, con una maggiore complessità operativa e maggiori risorse, potrebbero ambire a una maturità completa per gestire in modo efficace il proprio patrimonio di dati e processi legati all'IA, cogliendone appieno i benefici.

Il modello non solo fornisce una mappatura dettagliata del percorso evolutivo degli enti, ma rappresenta anche uno strumento strategico fondamentale per identificare gap e priorità, allineando in modo concreto visioni e strategie per una transizione graduale e omogenea verso una PA AI-ready.

Un aspetto di notevole valore aggiunto del modello è la sua flessibilità: le cinque dimensioni sono potenzialmente autonome tra loro, consentendo di identificare per ciascuna di esse diversi livelli di maturità. Questo approccio permette di cogliere potenziali disomogeneità all'interno dello stesso ente, contribuendo ad allineare così visioni differenti a seconda delle competenze e dei ruoli dei vari soggetti coinvolti. Tuttavia, è fondamentale che, pur nella loro autonomia, le dimensioni siano valutate con una visione d'insieme per garantire un'adozione armonica dell'IA. Diversamente, si rischia che le azioni intraprese non siano coordinate, impedendo così una reale crescita equilibrata durante il percorso di adozione.

Un ulteriore aspetto rilevante è che il modello non prevede un percorso obbligatorio verso il massimo livello di maturità per ogni ente. Le piccole Amministrazioni, con risorse limitate e minori complessità, potrebbero non necessitare di raggiungere livelli elevati in tutte le dimensioni, mentre grandi enti, con una maggiore complessità operativa e maggiori risorse, potrebbero ambire a una maturità completa per gestire in modo efficace il proprio patrimonio di dati e processi legati all'IA, cogliendone appieno i benefici.

Il modello non solo fornisce una mappatura dettagliata del percorso evolutivo degli enti, ma rappresenta anche uno strumento strategico fondamentale per identificare gap e priorità, allineando in modo concreto visioni e strategie per una transizione graduale e omogenea verso una PA AI-ready.

B. Valutazione del rischio

Il documento propone un modello di valutazione del rischio che dovrebbe essere applicato per tutti i sistemi IA con rischio diverso da "rischio minimo o nessun rischio" (per i quali è consigliata un'analisi semplificata) così come definito dall'AI Act.

Tale modello sarà arricchito in futuro da ulteriori approfondimenti con maggiore dettaglio tecnico, seguendo l'evoluzione delle attività in corso di definizione di metodologie specifiche[72] e norme tecniche europee ed internazionali (cfr. allegato E).

B.1. Modello di valutazione, testing e monitoraggio del rischio nell'adozione di IA nella PA

Il modello di analisi del rischio dovrebbe prevedere una differenziazione in termini di tipologia e valorizzazione dei KPI di valutazione in funzione della maturità della PA nel gestire l'adozione di una soluzione e delle caratteristiche del caso d'uso.

A tal proposito, è consigliabile che ogni PA definisca il sistema IA che intende adottare in termini di casi d'uso. In **Errore. L'origine riferimento non è stata trovata.**, si suggeriscono dei criteri per la definizione i n maniera univoca di casi d'uso (cfr. par. 4.5):

- *Dominio*: ambito di business o funzione organizzativa beneficiaria (ad esempio: Direzione X);
- *Applicazione*: finalità dell'utilizzo (ad esempio: Incremento della produttività);
- *Tecnologia*: tipologia di modello IA (ad esempio: IA generativa, modello LLM);
- *Destinatario*: utenza e volumi potenziali (ad esempio: cittadini, oltre 10.000).

[72] HUDERIA - risk and impact assessment of AI systems https://www.coe.int/en/web/artificial-intelligence/huderia-risk-and-impact-assessment-of-ai-systems

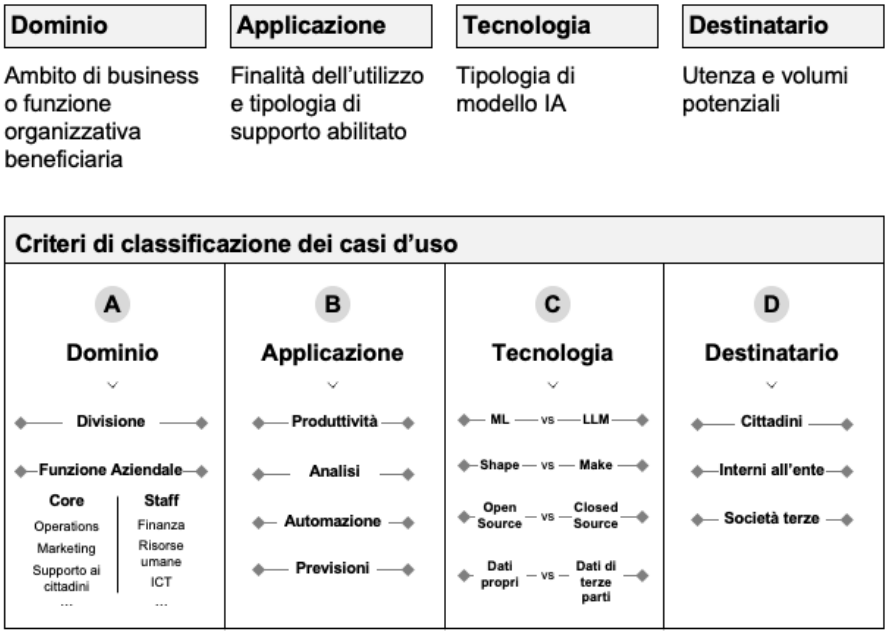

Figura B1. Criteri di classificazione e mappatura dei casi d'uso.

Per ogni caso d'uso le PA dovrebbero definire indicatori di prestazioni specifici (KPI).

Per ogni caso d'uso, i KPI devono essere valorizzati con soglie minime di accettazione opportunamente legate al livello di maturità del sistema di IA che si sta sviluppando, al di sotto delle quali l'ulteriore sviluppo del caso d'uso non sarà perseguibile (ad esempio, violazione di normative e regolamenti, livello di rischio inaccettabile, costi fuori budget). Data la definizione a livello di caso d'uso, si consiglia di esprimere le soglie minime di accettazione sottoforma di range di valori.

Come illustrato in **Errore. L'origine riferimento non è stata trovata.**, a partire dalla definizione delle m etriche e soglie minime per lo specifico caso d'uso, il modello proposto mira a svolgere un'investigazione *end-to-end* della rischiosità a livello di singolo caso d'uso nell'adozione di una tecnologia basata sull'Intelligenza Artificiale e ne prevede un monitoraggio continuo durante tutto il suo ciclo di vita (definito secondo le metriche stabilite dall'OECD):

- *Valutazione ex ante*: attraverso la diagnosi preventiva delle opportunità e dei potenziali rischi specifici, la definizione di requisiti minimi di compliance e di performance (limitatamente a quelle metriche per le quali è possibile l'identificazione di valori e soglie ex ante) identificati nella fase di pianificazione e design del sistema IA;
- *Valutazione in fase di testing*: tramite una metodologia di testing che consenta di valutare la conformità effettiva e sostanziale ai requisiti normativi, etici e di performance espressi nella fase valutativa di pianificazione e design;
- *Monitoraggio continuo:* nella fase di operatività e monitoraggio, grazie alla definizione di una dashboard riepilogativa che permetta la visualizzazione dei requisiti definiti per tramite di KPI etici, di

compliance e di performance e un raffronto con gli effettivi risultati misurati durante tutto il ciclo di vita del caso d'uso.

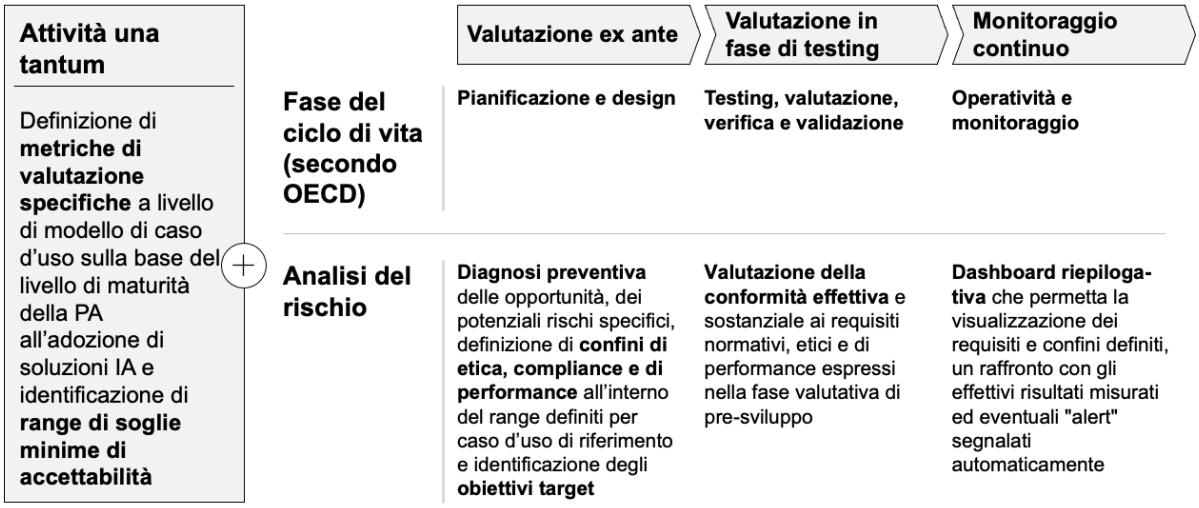

Figura B.2. Analisi del rischio per le fasi del ciclo di vita di una tecnologia IA.

Entrando maggiormente nel dettaglio, si analizzano nel seguito l'attività una tantum a livello di caso d'uso e le tre fasi principali in cui l'analisi e il controllo del rischio sono fondamentali al fine di rilasciare un prodotto tecnologico sicuro per la PA e i suoi utenti finali.

B.2. Attività una tantum: definizione di metriche e range di soglie minime per la valutazione del rischio di un modello di caso d'uso

L'attività di identificazione delle metriche sulla base del livello di maturità all'adozione di soluzioni IA della PA e la successiva definizione di range di soglie minime può essere svolta al momento di definizione di un nuovo caso d'uso e aggiornata solo in caso di necessità. Si definisce pertanto attività una tantum.

Si suggeriscono di seguito le metriche standard adottabili da una PA semplice, potenzialmente personalizzabili da PA più complesse:

- *Performance*: attiene a tre principali caratteristiche dei modelli IA, eventualmente integrabili con ulteriori grandezze (es., rilevanza delle risposte, linguaggio utilizzato):
  - *Accuratezza*: capacità del modello di generalizzare, fare previsioni o fornire risposte accurate e affidabili rispetto al reale contesto di riferimento;
  - *Robustezza*: capacità del modello di mantenere performance stabili e accurate, anche in presenza di dati incompleti, anomali e in continua evoluzione nel tempo;
  - *Efficienza*: utilizzo richiesto dal modello di risorse computazionali, come CPU, GPU e memoria, e tempo necessario per elaborare le previsioni/risposte;
- *Compliance*: si riferisce al rispetto da parte del modello di norme e regolamenti interni ed esterni alla Pubblica Amministrazione di riferimento, in particolare ad esempio:

- *Conformità normativa:* afferenti a norme e leggi relativi all'utilizzo dell'IA e alla gestione e protezione dei dati;
- *Policy di IA Responsabile:* definiti internamente ed esternamente alla Pubblica Amministrazione e basate su principi etici e linee guida di enti preposti come l'UE (incl. valori come data protection, data governance, trasparenza, responsabilità e informazione, adozione di standard tecnici, ecc);
- *Gestione del rischio:* regolamenti interni alla Pubblica Amministrazione afferenti al livello di tolleranza / predisposizione al rischio;
- *Cybersicurezza:* garantita dalla presenza di processi e software che rendono robusta dal punto di vista di attacchi hacker la struttura della soluzione IA;

Il vincolo etico (ad esempio: accessibilità, inclusività, non discriminazione) deve essere rispettato a prescindere dal soddisfacimento dei requisiti di performance e compliance e una mancata soddisfazione di tale caratterista rende la soluzione IA non adatta allo sviluppo.

B.3. Valutazione ex ante del singolo caso d'uso

Per ognuno dei KPI identificati a livello di caso d'uso nel paragrafo precedente, è necessario definire delle soglie minime oltre le quali la PA non è disposta a considerare lo sviluppo di una tecnologia IA. Tali soglie dovranno rientrare nei range definiti per il modello di caso d'uso di riferimento e tenere conto di eventuali regolamenti o scelte interne all'Amministrazione per il singolo caso d'uso.

Il raffronto tra le soglie minime e il caso d'uso andrà svolto tramite valorizzazione attesa delle metriche selezionate per il singolo caso d'uso, ove possibile (ad esempio, non sarà possibile valorizzare le metriche di robustezza ed efficienza, mentre l'accuratezza potrà essere valorizzata in termini di disponibilità attesa dei dati necessari al modello IA).

In questo modo viene abilitato un "filtro" scalabile, sintetizzato in Figura B.3, in grado di supportare una valutazione preliminare dei casi d'uso, promuovendo alla fase successiva di sviluppo solo quelle soluzioni che soddisfano le soglie minime di performance e compliance, minimizzando contemporaneamente i rischi reputazionali, operativi e normativi.

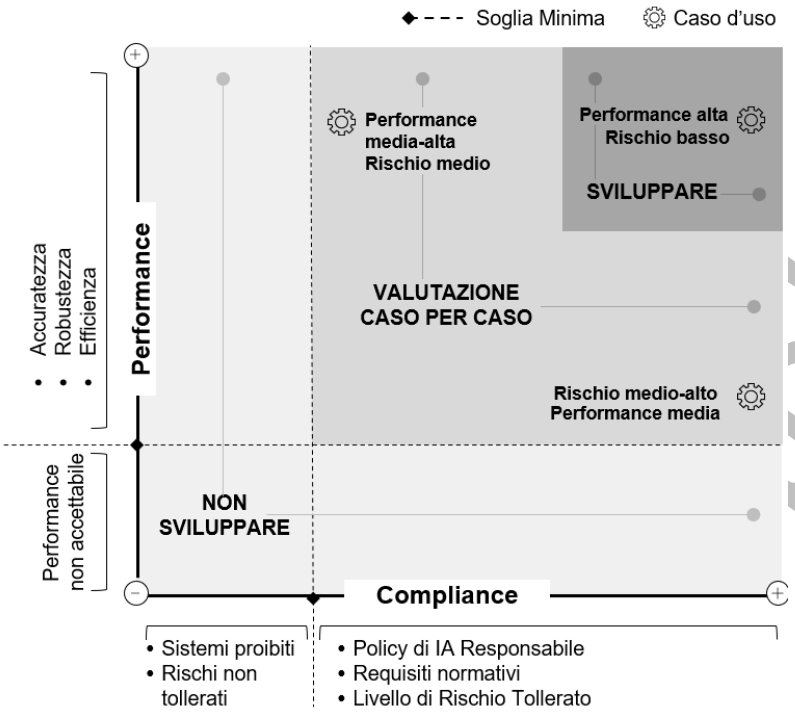

Figura B.3. "Filtro" di valutazione casi d'uso IA

Infine, in tale fase ogni PA dovrebbe definire dei target a livello di singolo caso d'uso per ogni metrica ai quali mirare nella fase successiva di sviluppo e testing della tecnologia di Intelligenza Artificiale.

In Figura B.4 si riporta una sintesi illustrativa di una scheda di valutazione ex ante di un caso d'uso IA di esempio, che aggrega e rende accessibile e fruibile ai decisori:

- La classificazione del singolo caso d'uso, secondo i criteri predefiniti;
- Le analisi di fattibilità tecnica, la valutazione dell'impatto atteso ed ulteriori allegati tecnico-economici;
- I requisiti etici, di compliance e di performance, con indicazione delle soglie minime e dei target attesi.

Caso d'uso #1

Overview

A **Dominio**: Cliente X - Contact Center (CC)
B **Applicazione**: Produttività degli operatori del CC
C **Tecnologia**: Generative AI (LLM Open Source)
D **Destinatario**: Operatori del Contact Center

Allegati di dettaglio:

Dettaglio requisiti | Prototipo della UX | Stima di Impatto | Analisi di fattibilità

Balanced Scorecard

COMPLIANCE

KPIs	Range soglia minima	Soglia minima use-case	Valutazione ex-ante	Target
IA Act	R.. min xx%	R. al yy%		zz
Data Protection	R.. min xx%	R. al yy%		zz
Data Governance	R.. min xx%	R. al yy%		zz
Trasparenza	R.. min xx%	R. al yy%		zz
Responsabilità	R.. min xx%	R. al yy%		zz
Informazione	R.. min xx%	R. al yy%		zz
Standard tecnici	R.. min xx%	R. al yy%		zz
Cybersecurity	R.. min xx%	R. al yy%		zz

PERFORMANCE

KPIs	Range soglia minima	Soglia minima use-case	Valutazione ex-ante	Target
Accuratezza	Tra x% e y%	zz%		bb%
Robustezza	Tra x% e y%	zz%	*non stimabile*	bb%
Efficienza	Tra x% e y%	zz%	*non stimabile*	bb%

Figura B.4. Scheda riassuntiva di esempio per un caso d'uso IA

B.4. Valutazione in fase di testing del caso d'uso IA

La potenziale limitata interpretabilità dei modelli IA, richiede che l'adozione di misure di valutazione preliminari sia combinata con metodologie di testing sulla tecnologia effettivamente sviluppata, con l'obiettivo di confermare la rispondenza ai requisiti di eticità, compliance e performance attesi nella fase valutativa ex ante. A tal fine, risulta utile attivare una metodologia che consenta di:

- Identificare e ricreare i casi limite non presenti nel dataset di riferimento (ad esempio, dati errati, campi non attesi);
- Sottoporre tali casi all'elaborazione del modello di IA;
- Confrontare i risultati del modello IA con l'output atteso.

I risultati ottenuti andranno tracciati tramite appositi KPI di etica, compliance e performance e confrontati con i risultati attesi e di target prospettati nella fase di valutazione ex ante.

B.5. Monitoraggio continuo del caso d'uso IA

La natura dinamica delle tecnologie basate sull'Intelligenza Artificiale prevede un monitoraggio continuo del rispetto dei requisiti etici, di compliance e performance minimi, che devono essere testati per tutta la durata del ciclo di vita della soluzione IA.

Ove possibile, tali test dovrebbero essere svolti tramite controlli automatizzati ad elevata frequenza, con l'obiettivo di garantire un controllo puntuale su tutti i parametri della tecnologia e al contempo fornire segnalazioni immediate nel caso di anomalie. Si consiglia pertanto l'introduzione di soglie di allerta per ogni metrica a livello di singolo caso d'uso (potenzialmente assimilabili ai valori di soglia minimi definiti nella fase di valutazione ex ante) e la segnalazione di "alert" proattivi nella dashboard di monitoraggio del modello IA nel momento in cui una o più metriche dovessero assumere valori al di sotto di tali soglie.

C. Valutazione d'impatto

Il documento fornisce uno strumento di supporto per le PA che dovranno effettuare la valutazione dell'impatto (cd. AIIA, Artificial Intelligence Impact Assessment)[73] dell'introduzione di un sistema di IA.

Lo strumento di seguito descritto e pubblicato al link (*il link sarà individuato a valle della consultazione pubblica*) è strutturato come una check list ed è suddiviso in sei schede più due appendici, che riassumono i principali aspetti che è opportuno valutare prima di adottare un sistema di IA.

Ogni scheda presenta domande a risposta chiusa (si, no, non applicabile) con descrizione e/o motivazione obbligatoria, e domande a risposta aperta con descrizione e/o motivazione obbligatoria.

Sono presenti domande obbligatorie (O) e domande facoltative (F).

Le schede sono strutturate come segue:

- **C1. Domande generali**. La scheda si compone di una serie di domande introduttive, ed è suddivisa nelle sezioni "scopo del sistema" e "impatto del sistema". Nella prima sezione, "scopo del sistema", deve essere riportato l'esito della valutazione del rischio (cfr. livello di rischio) del sistema di IA; qualora il rischio sia inaccettabile, il sistema non potrà essere introdotto e la valutazione di impatto non andrà completata.
- **C2. Diritti fondamentali ed equità.** La scheda è costituita dalle sezioni "Diritti fondamentali", "Bias" e "Partecipazione degli stakeholders".
- **C3. Robustezza tecnologica**. La scheda è suddivisa nelle sezioni "Accuratezza", "Affidabilità", "Implementazione tecnica", "Replicabilità" e "Spiegabilità".
- **C4. Governance dei dati**. La scheda è suddivisa nelle sezioni "Qualità e integrità dei dati" e "Privacy e protezione dei dati personali".
- **C5. Gestione del rischio.** La scheda è suddivisa nelle sezioni "Gestione del rischio informatico", "Procedura alternativa" e "Attacchi hacking e corruzione del sistema".
- **C6. Accountability.** La scheda è suddivisa nelle sezioni "Comunicazione", "Verificabilità", "Archiviazione", "Sostenibilità".

Infine, è presente un'appendice:

- **Appendice.1 - Livello di rischio,** che consente di mappare il livello di rischio (inaccettabile, elevato, medio e basso).

[73] Lo strumento è stato sviluppato prendendo come riferimento l'AI Impact Assessment (AIIA) introdotto dal governo dei Paesi Bassi. Per maggiori informazioni si veda il report "AI Impact Assessment: A tool to set up responsible AI projects" https://www.government.nl/documents/publications/2023/03/02/ai-impact-assessment.

D. Modello di codice etico

SCHEMA DI CODICE ETICO E DI COMPORTAMENTO

relativo all'applicazione dei sistemi di Intelligenza Artificiale

CAPO I

DISPOSIZIONI GENERALI

Premessa

L'IA può avere un impatto significativo sui diritti umani, libertà fondamentali e sulla dignità umana e l'uguaglianza. Per tutelare in modo efficace ambiti sensibili come la sfera personale e sociale, le questioni etiche riguardanti i sistemi di Intelligenza artificiale DEVONO essere rilevanti per tutti gli stadi del ciclo di vita degli stessi, dalla ricerca, la progettazione e lo sviluppo fino alla installazione e utilizzo, inclusi manutenzione, funzionamento, monitoraggio, valutazione e dismissione.

Lo sviluppo di sistemi di IA antropocentrici va considerato come compito della società nel suo complesso e non consiste nella semplice regolamentazione della tecnologia.

Art. 1

(*Definizioni*)

Ai fini del presente documento, i termini di seguito indicati assumono il significato riportato nel presente articolo in relazione a ciascuno di essi, in coerenza con le definizioni previste dal Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024:

a) «**AI Act**»: il Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024, che istituisce un quadro giuridico uniforme per quanto riguarda lo sviluppo, l'immissione sul mercato, la messa in servizio e l'uso di Sistemi di AI nell'Unione per la diffusione di un'Intelligenza Artificiale antropocentrica e affidabile, che garantisca la protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea;

b) «**Codice Etico di AI**»: il presente codice etico, istituito ai sensi dell'art. 95 dell'AI Act, che definisce le regole etiche da rispettare nell'utilizzo di Sistemi di AI da parte dell'Ente;

c) «**Codice di Comportamento**»: il codice di comportamento che definisce i doveri minimi di diligenza, lealtà, imparzialità e buona condotta e i principi cogenti che i dipendenti dell'Ente sono tenuti ad osservare nello svolgimento delle rispettive attività lavorative;

d) «**Ente**»: il soggetto che adotta il presente Codice Etico di AI e ne agevola e promuove la diffusione;

e) «**Intelligenza Artificiale**» o anche «IA»: l'abilità di un sistema di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività, analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere obiettivi specifici;

f) «**Orientamenti**»: i principi, gli indirizzi e le previsioni in materia di etica e di AI adottati da autorità pubbliche, anche a livello internazionale, tra i quali, a titolo esemplificativo e non esaustivo, ricorrono i seguenti documenti:

- "*Carta etica europea sull'utilizzo dell'IA nei sistemi giudiziari e negli ambiti connessi*", adottata dalla Commissione europea per l'efficienza della giustizia (CEPEJ) il 3 dicembre 2018;
- "*Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni - «Creare fiducia nell'intelligenza artificiale antropocentrica»*", elaborata dal gruppo di esperti di alto livello istituito dalla Commissione Europea e pubblicata in data 8 aprile 2019;
- "*Recommendation on the Ethics of Artificial Intelligence*", pubblicata dall'UNESCO il 23 novembre 2021;
- "*Principles for the ethical use of artificial intelligence in the United Nations system*", pubblicato il 27 ottobre 2022 dal Consiglio di coordinamento dei capi esecutivi del sistema delle Nazioni Unite.

g) «**Sistema di Intelligenza Artificiale**» o «**Sistema di IA**»: un sistema automatizzato progettato per funzionare con livelli di autonomia variabili, 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. 2

(*Oggetto*)

Conformemente a quanto previsto dall'AI Act e dagli Orientamenti, nell'ambito dei Sistemi di Intelligenza Artificiale il presente Codice Etico di IA:

a) reca i princìpi guida dell'azione amministrativa dell'Ente nonché del comportamento dei soggetti che operano presso l'Ente, specificando i doveri di lealtà, imparzialità, diligenza ed operosità previsti per i dipendenti pubblici dal Codice di Comportamento;

b) incentiva l'applicazione volontaria dei requisiti specifici di cui al Capo III, Sezione 2, dell'AI Act ai Sistemi di Intelligenza Artificiale non ad alto rischio, attraverso l'adozione delle soluzioni tecniche disponibili, il rispetto delle *best practice* di settore e la definizione di obiettivi chiari e indicatori chiave di prestazione diretti a misurare l'effettivo conseguimento dei predetti obiettivi.

Art. 3

(*Ambito di applicazione*)

1. I principi e i contenuti del Codice Etico di IA si rivolgono al personale dipendente dell'Ente e ai soggetti che a vario titolo entrano in contatto con l'Ente stesso nell'ambito dell'esercizio dell'attività amministrativa e contrattuale, compresi i collaboratori, i consulenti, gli appaltatori e gli altri enti per le finalità che ne interessano gli ambiti di cooperazione.

2. L'Ente agevola e promuove la diffusione del Codice Etico di IA e predispone ogni possibile strumento che ne favorisca la conoscenza e la piena applicazione.

Art. 4

(*Finalità di interesse pubblico*)

1. L'Ente applica i Sistemi di Intelligenza Artificiale per il perseguimento di finalità di interesse pubblico in modo tale da incrementare il livello di qualità di vita della collettività di riferimento, nel rispetto delle relative attribuzioni istituzionali e di quanto previsto dall'AI Act e dagli Orientamenti.
2. L'impiego dei Sistemi di Intelligenza Artificiale nell'ambito dei processi organizzativi è funzionale a implementare l'efficienza degli apparati dell'Ente e semplificare i processi decisionali amministrativi, nel rispetto dei principi di economicità, efficacia, efficienza, imparzialità, pubblicità, trasparenza e correttezza.
3. L'Ente è tenuto a impiegare i Sistemi di Intelligenza Artificiale in modo da rendere l'individuo parte attiva del progresso umano e scientifico, tutelandone i valori, le libertà e i diritti, nel rispetto dell'autonomia e della libertà personale.

CAPO II

PRINCIPI GENERALI

Art. 5

(*Valori e principi etici condivisi*)

1. Le previsioni del presente Codice Etico di IA si fondano sui valori etici condivisi a livello globale, quali il rispetto e la protezione dei diritti umani, delle libertà fondamentali e della dignità umana, l'attenzione all'ambiente e agli ecosistemi, la tutela delle diversità e dell'inclusione, nonché la garanzia di vivere in un ambiente pacifico, giusto, basato su un futuro interconnesso a beneficio di tutti.
2. Ferma restando l'osservanza degli obblighi stabiliti dall'AI Act e dagli Orientamenti, l'impiego di Sistemi di Intelligenza Artificiale da parte dell'Ente avviene nel rispetto dei principi di trasparenza, responsabilità, sicurezza, sostenibilità ambientale, non discriminazione e tutela della riservatezza dei dati personali.

Art. 6

(*Trasparenza*)

1. L'Ente rende conoscibili e spiegabili, anche con un linguaggio non tecnico, l'elenco delle soluzioni di Intelligenza Artificiale nella disponibilità dell'Ente o in corso di acquisizione da parte del medesimo, dando atto delle modalità e degli ambiti di impiego nonché delle funzionalità connesse.
2. Su richiesta di ogni soggetto interessato, l'Ente fornisce ogni informazione utile a descrivere l'impatto potenziale dell'utilizzo di tali soluzioni nei diversi ambiti di impiego.
3. L'Ente conserva una documentazione dettagliata e accessibile riguardante lo sviluppo, l'implementazione e l'uso dei Sistemi di IA. Questa documentazione include informazioni sui dati utilizzati, sugli algoritmi impiegati e sui processi decisionali automatizzati.

Art. 7

(*Responsabilità*)

1. L'Ente garantisce che ogni decisione critica assunta tramite Sistemi di Intelligenza Artificiale sia previamente sottoposta all'apprezzamento finale degli esseri umani per garantirne l'accuratezza e l'equità.
2. L'Ente assicura la sorveglianza sulle decisioni adottate tramite Sistemi di Intelligenza Artificiale, verificando l'impatto che queste possono avere rispetto alle decisioni assunte secondo le modalità tradizionali.

Art. 8

(*Sicurezza*)

1. Nel rispetto di quanto previsto dall'AI Act e dagli Orientamenti, l'Ente utilizza, acquisisce e implementa Sistemi di Intelligenza Artificiale che adottino misure di sicurezza robuste per prevenire accessi non autorizzati, attacchi informatici e altre minacce.
2. Al fine di mantenere la protezione dei Sistemi di Intelligenza Artificiale in uso, con cadenza periodica l'Ente esegue controlli sui requisiti di sicurezza ed effettua aggiornamenti.
3. L'Ente adotta e diffonde buone pratiche interne per la prevenzione dei danni conseguenti all'uso di Sistemi di Intelligenza Artificiale, anche tramite la predisposizione di procedure di *risk assessment* e *management* con la finalità di identificare situazioni critiche e potenziali scenari di rischio.

Art. 9

(*Sostenibilità ambientale*)

1. L'Ente valuta e riduce al minimo l'impatto dei Sistemi di Intelligenza Artificiale sulla sostenibilità ambientale, anche attraverso l'adozione di strumenti di acquisito di soluzioni conformi ai parametri di cui al successivo comma 3.
2. L'utilizzo delle soluzioni di Intelligenza Artificiale è funzionale al raggiungimento degli obiettivi di sviluppo sostenibile, così come individuati dall'ONU all'interno dell'«*Agenda 20230 per lo Sviluppo Sostenibile*».
3. La verifica della sostenibilità ambientale delle soluzioni di Intelligenza Artificiale può avvenire attraverso il richiamo ai parametri fissati in relazione al principio «*Do No Significant Harm*».

Art. 10

(*Non discriminazione*)

1. L'Ente facilita la progettazione inclusiva e diversificata dei Sistemi di Intelligenza Artificiale, anche attraverso la creazione di gruppi di sviluppo inclusivi e diversificati e la promozione della partecipazione dei portatori di interessi a tale processo, nel rispetto del principio di non discriminazione e dei relativi corollari.

2. Le decisioni e i processi automatizzati derivanti dall'applicazione dell'Intelligenza Artificiale assicurano la prevenzione di *bias* cognitivi, riconoscendo e rimuovendo gli stessi *bias* dai *set* di dati, dalle scelte tecniche e tecnologiche, nonché dall'interpretazione dei risultati, al fine di prevenire qualunque impatto discriminatorio.
3. L'Ente prevede che i Sistemi di Intelligenza Artificiale progettati e adoperati siano in grado di essere utilizzati anche da parte di soggetti disabili che necessitano di tecnologie assistive o configurazioni particolari.

Art. 11

(Privacy)

1. Nell'ambito dello sviluppo, dell'acquisizione e dell'utilizzo di Sistemi di Intelligenza Artificiale, l'Ente garantisce il rispetto della disciplina in materia di protezione dei dati personali, nonché degli orientamenti internazionali e nazionali adottati in materia.
2. Nell'ambito dei trattamenti di dati personali effettuati mediante l'utilizzo di Sistemi di Intelligenza Artificiale, l'Ente assicura il rispetto dei princìpi di liceità, correttezza, trasparenza, esattezza, limitazione delle finalità, minimizzazione, limitazione della conservazione, integrità e riservatezza, nonché dei seguenti principi individuati nell'ambito degli orientamenti di settore:
   a) comprensibilità, conoscibilità e rilevanza delle informazioni fornite ai soggetti interessati circa le modalità di funzionamento della soluzione di AI adottata e della relativa logica utilizzata;
   b) non esclusività della decisione algoritmica;
   c) non discriminazione algoritmica, mediante l'adozione di sistemi antropocentrici e affidabili, idonei a tutelare i diritti e le libertà fondamentali degli interessati, prestando particolare attenzione ai diritti dei soggetti vulnerabili e, in particolare, dei minori.

CAPO III
RAPPORTI CON IL PERSONALE

Art. 12

(*Reclutamento e gestione del personale*)

1. Nella ricerca e selezione del personale, nella costituzione del rapporto di lavoro e nella successiva gestione, l'Ente può avvalersi di Sistemi di Intelligenza Artificiale al fine di conseguire indicatori utili da sottoporre all'apprezzamento finale umano.
2. I Sistemi di Intelligenza Artificiale richiamati al comma che precede garantiscono, in quanto compatibili e conformemente ai regolamenti interni, il rispetto dei principi di pari opportunità e parità di genere.

Art. 13

(*Utilizzo dei Sistemi di AI nell'attività lavorativa*)

1. Il personale dell'Ente è tenuto a utilizzare i Sistemi di Intelligenza Artificiale nel rispetto degli obblighi di comportamento posti dal Codice di Comportamento.
2. Il personale è incoraggiato a esplorare e comprendere i Sistemi di Intelligenza Artificiale messi a disposizione dall'Ente, nella consapevolezza dei rischi associati all'uso dell'Intelligenza Artificiale generativa, inclusi *bias* e disinformazione.
3. Il personale sottopone a continui processi di verificazione gli *output* dell'Intelligenza Artificiale generativa, evitando di recepirli acriticamente nell'ambito della propria attività lavorativa.

Art. 14

(*Formazione*)

1. In conformità all'assunto per cui i Sistemi di Intelligenza Artificiale sono sviluppati al servizio delle persone, l'Ente adotta le misure necessarie affinché il proprio personale e gli eventuali ulteriori soggetti coinvolti nell'utilizzo dei Sistemi di Intelligenza Artificiale possano servirsene in modo consapevole e responsabile, anche attraverso l'organizzazione periodica di corsi di formazione focalizzati sulla materia.
2. L'Ente promuove l'alfabetizzazione in materia di Intelligenza Artificiale con particolare riguardo alle persone che si occupano dello sviluppo, del funzionamento e dell'uso dell'Intelligenza Artificiale.

CAPO IV

RAPPORTI CON L'ESTERNO

Art. 15

(*Cooperazione con altri enti*)

1. L'Ente collabora con altre istituzioni e organizzazioni, di natura pubblica e privata, per condividere informazioni sulle best practice per l'utilizzo e la protezione di Sistemi di Intelligenza Artificiale, nel rispetto delle modalità e delle forme previste dall'ordinamento.
2. L'Ente promuove la partecipazione a progetti di ricerca congiunti con università e centri di ricerca, al fine di esplorare nuove applicazioni dell'Intelligenza Artificiale e migliorare le pratiche esistenti.

Art. 16

(*Coinvolgimento degli* stakeholder)

1. L'Ente adotta il modello «*multistakeholder*», che si fonda sul coinvolgimento delle diverse parti interessate, promuovendo il coinvolgimento dei cittadini, delle associazioni e delle imprese negli ambiti che riguardano l'uso dei Sistemi di Intelligenza Artificiale, al fine di colmare le eventuali asimmetrie informative in materia.
2. Nel rispetto delle relative attribuzioni istituzionali e degli strumenti previsti dall'ordinamento, l'Ente partecipa al dibattito pubblico sulle implicazioni etiche, giuridiche ed economiche dei Sistemi di Intelligenza Artificiale.

3. L'Ente adotta un approccio inclusivo, tenendo conto anche degli interessi e delle esigenze specifiche delle piccole e medie imprese e delle *start-up* nella progettazione e nell'uso dei Sistemi di Intelligenza Artificiale, al fine di accelerare il processo di innovazione interno.

Art. 17

(*Procedure di acquisizione di beni e servizi*)

Nell'ambito delle procedure di acquisizione di Sistemi di Intelligenza Artificiale non ad alto rischio da aggiudicarsi con il criterio dell'offerta economicamente più vantaggiosa, l'Ente valuta la possibilità di adoperare criteri premiali in relazione alle proposte che garantiscano il rispetto dei requisiti fissati dal Capo III, Sezione 2, dell'AI Act.

CAPO V

GOVERNANCE

Art. 18

(*Comitato Etico*)

1. L'Ente istituisce un meccanismo di supervisione per monitorare l'uso dei Sistemi di Intelligenza Artificiale e garantire che siano utilizzati in conformità con il presente Codice Etico di IA.
2. Ai fini del comma che precede, l'Ente può istituire, con proprio provvedimento, un Comitato Etico per l'Intelligenza Artificiale con l'obiettivo di osservare lo sviluppo e l'utilizzo di Sistemi di Intelligenza Artificiale in conformità alle linee guida e agli orientamenti condivisi a livello nazionale ed europeo.
3. Al Comitato Etico possono essere attribuiti i seguenti compiti, nel rispetto dei vincoli fissati dalla normativa di riferimento e dai regolamenti interni:
   a) promuovere l'adozione di misure volte a incentivare il rispetto dei requisiti specifici anche per i Sistemi di Intelligenza Artificiale non ad alto rischio;
   b) garantire e monitorare la costante applicazione dei valori e dei principi etici dell'Intelligenza Artificiale;
   c) fornire supporto al personale nell'interpretazione delle regole di utilizzo dei Sistemi di Intelligenza Artificiale, nella successiva fase applicativa e durante tutto il ciclo di vita degli stessi Sistemi di IA;
   d) risolvere eventuali incertezze derivanti dall'interpretazione del presente Codice Etico di IA;
   e) coinvolgere i portatori di interessi pubblici nell'ambito della progettazione e dell'implementazione di Sistemi di Intelligenza Artificiale.
4. Il Comitato Etico può essere composto da un numero di soggetti dispari, in misura non inferiore a 3, individuati tra i soggetti dotati delle necessarie competenze all'interno del personale in servizio, e rimane in carica per un biennio.

Art. 19

(*Adozione di indirizzi*)

Nell'ambito dell'utilizzo di Sistemi di Intelligenza Artificiale per lo svolgimento delle proprie attività, il Comitato Etico può adottare specifici indirizzi e prassi interne dirette a conformare l'utilizzo dei Sistemi di IA nella prospettiva di garantire la protezione dei diritti umani, delle libertà fondamentali e della dignità umana, la sostenibilità ambientale e la garanzia dell'opposizione a qualsiasi forma di isolamento o discriminazione.

CAPO VI

DISPOSIZIONI FINALI

Art. 20

(*Pubblicità*)

Il Codice Etico di IA è pubblicato sul sito internet dell'Ente e della pubblicazione viene data notizia a tutti i dipendenti mediante specifica comunicazione tramite *e-mail* e, eventualmente, anche ai fornitori mediante l'inserimento di apposita informativa all'interno della documentazione delle procedure pubbliche nelle quali sono coinvolti.

Art. 21

(*Approvazione e aggiornamento*)

1. Ogni modifica e integrazione al presente Codice Etico di IA dovrà essere apportata con le stesse modalità adottate per la sua approvazione iniziale.
2. Il presente Codice Etico di IA è sottoposto ad aggiornamento ogni anno su proposta del Comitato Etico in base all'evoluzione tecnologica e normativa, previa consultazione degli esperti e dei portatori di interesse del settore.

E. Norme tecniche in ambito IA

Il documento presenta una selezione delle norme tecniche in ambito IA attualmente pubblicate.

L'elenco esaustivo delle norme tecniche a livello nazionale, europeo ed internazionale e lo stato di avanzamento delle attività di normazione tecnica è disponibile alla pagina della commissione tecnica UNI/CT 533 "Intelligenza artificiale" https://www.uninfo.it/partecipare/aree/ai-intelligenza-artificiale.html.

- **ISO/IEC 22989:2022** Information technology - Artificial intelligence - Artificial intelligence concepts and terminology
- **ISO/IEC 42001:2023** Information technology - Artificial intelligence - Management system
- **ISO/IEC 23894:2023** Information technology - Artificial intelligence - Guidance on risk management
- **ISO/IEC 5259-1:2024** Artificial intelligence -Data quality for analytics and machine learning (ML) - Part 1: Overview, terminology, and examples
- **ISO/IEC 5259-2:2024** Artificial intelligence -Data quality for analytics and machine learning (ML) - Part 2: Data quality measures
- **ISO/IEC 5259-3:2024** Artificial intelligence -Data quality for analytics and machine learning (ML) - Part 3: Data quality management requirements and guidelines
- **ISO/IEC 5259-4:2024** Artificial intelligence -Data quality for analytics and machine learning (ML) - Part 4: Data quality process framework
- **ISO/IEC 5338:2023** Information technology -Artificial intelligence -AI system life cycle processes
- **ISO/IEC 5339:2024** Information technology -Artificial intelligence -Guidance for AI applications
- **ISO/IEC 5392:2024** Information technology -Artificial intelligence -Reference architecture of knowledge engineering
- **ISO/IEC TS 4213:2022** Information technology -Artificial intelligence -Assessment of machine learning classification performance
- **ISO/IEC TR 5469:2024** Artificial intelligence -Functional safety and AI systems
- **ISO/IEC 8183:2023** Information technology -Artificial intelligence -Data life cycle framework
- **ISO/IEC TS 8200:2024** Information technology -Artificial intelligence -Controllability of automated artificial intelligence systems
- **ISO/IEC TS 12791** Information technology -Artificial intelligence -Treatment of unwanted bias in classification and regression machine learning tasks
- **ISO/IEC TR 17903:2024** Information technology -Artificial intelligence -Overview of machine learning computing devices
- **ISO/IEC 20546:2019** Information technology -Big data -Overview and vocabulary
- **ISO/IEC TR 20547-1:2020** Information technology -Big data reference architecture -Part 1: Framework and application process

- **ISO/IEC TR 20547-2:2018** Information technology -Big data reference architecture -Part 2: Use cases and derived requirements
- **ISO/IEC 20547-3:2020** Information technology -Big data reference architecture -Part 3: Reference architecture
- **ISO/IEC TR 20547-5:2018** Information technology -Big data reference architecture -Part 5: Standards roadmap
- **ISO/IEC 23053:2022** Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
- **ISO/IEC TR 24027:2021** Information technology -Artificial intelligence (AI) -Bias in AI systems and AI-aided decision making
- **ISO/IEC TR 24028:2020** Information technology -Artificial intelligence -Overview of trustworthiness in artificial intelligence
- **ISO/IEC TR 24029-1:2021** Artificial Intelligence (AI) -Assessment of the robustness of neural networks -Part 1: Overview
- **ISO/IEC 24029-2:2023** Artificial intelligence (AI) -Assessment of the robustness of neural networks -Part 2: Methodology for the use of formal methods
- **ISO/IEC TR 24030:2024** Information technology -Artificial intelligence (AI) -Use cases
- **ISO/IEC TR 24368:2022** Information technology -Artificial intelligence -Overview of ethical and societal concerns
- **ISO/IEC TR 24372:2021** Information technology -Artificial intelligence (AI) -Overview of computational approaches for AI systems
- **ISO/IEC 24668:2022** Information technology -Artificial intelligence -Process management framework for big data analytics
- **ISO/IEC TS 25058:2024** Systems and software engineering -Systems and software Quality Requirements and Evaluation (SQuaRE) -Guidance for quality evaluation of artificial intelligence (AI) systems
- **ISO/IEC 25059:2023** Software engineering -Systems and software Quality Requirements and Evaluation (SQuaRE) -Quality model for AI systems
- **ISO/IEC 38507:2022** Information technology -Governance of IT -Governance implications of the use of artificial intelligence by organizations
- **ISO/IEC 25012:2008** (UNI ISO/IEC 25012:2014) Software engineering -Software product Quality Requirements and Evaluation (SQuaRE) -Data quality model

F. Casi d'uso

Di seguito è riportato lo schema descrittivo del caso d'uso introdotto al paragrafo 4.5.

1. **Dominio applicativo.**
   Rappresenta l'ambito o il settore specifico in cui un sistema di IA è utilizzato. La ISO/IEC TR 24030:2024 "Information technology - Artificial intelligence (AI) - Use cases" individua diciotto domini applicativi: agricoltura, marketing digitale, e-commerce/e-business, istruzione, energia, fintech, sanità, robotica domestica e di servizio, ICT (Tecnologie dell'informazione e della comunicazione), assicurazioni, gestione della conoscenza, legale, manifattura, media e intrattenimento, mobilità, pubblica amministrazione, vendita al dettaglio, trasporti. Per ragioni statistiche è opportuno che il dominio applicativo siano rappresentato anche secondo le categorie COFOG Classificazione spesa pubblica https://www.istat.it/classificazione/classificazione-internazionale-della-spesa-pubblica-per-funzione-cofog/

2. **Modello di distribuzione.**
   Modalità di distribuzione dell'applicazione IA, tra cui: sistemi on-premise, embedded, servizi cloud, ibridi (edge computing, ecc.).

3. **Obiettivi.**
   Finalità del sistema di IA ovvero il risultato atteso (es.: miglioramento dell'efficienza operativa mediante l'utilizzo di IA nella gestione documentale, cfr. par. 4.3), i beneficiari (es.: cittadinanza, imprese, altre PA, operatori della PA), il contesto in cui l'IA dovrà operare, indicando eventuali restrizioni o limitazioni funzionali (es.: integrazione con il sistema documentale esistente, gestione dei documenti classificati, utilizzo esclusivo delle funzionalità di raccomandazione).

4. **Descrizione.**
   Esposizione dettagliata di come il sistema di IA è utilizzato per ottimizzare processi o generare inferenze.
   Devono essere specificati:
   - 4.1. **decisioni, previsioni o raccomandazioni** *generate dal sistema di IA;*
   - 4.2. **capacità e caratteristiche** *del sistema di IA per il dominio di applicazione (ad esempio, analisi di grandi volumi di dati, elaborazione in tempo reale, personalizzazione avanzata);*
   - 4.3. **contesto decisionale**, *ovvero l'ambiente in cui il sistema opera, e quale decisione umana viene supportata o aumentata;*
   - 4.4. **dinamiche dell'ambiente decisionale**, *come la complessità, i tempi di risposta richiesti o le variabili esterne che influenzano le decisioni.*

5. **Caratteristiche dei dati**
   - 5.1. **Origine**.
     Fonte dei dati utilizzati dal sistema di IA, come ad esempio banche dati interne, banche dati esterne (es.: open data, dati acquistati), dati ottenuti da collaborazioni con altre istituzioni.
   - 5.2. **Varietà.**
     Contenuto e formato dei dati, eventuale conformità a standard di formato, modelli logici e classificazioni semantiche.
     - **formato**: *dati strutturati (es.: database relazionali, fogli di calcolo), dati non strutturati (es.: testi, immagini, video), dati semi-strutturati (es.: XML, JSON).*
     - **contenuto**: *dati quantitativi (discreti o continui), dati qualitativi (che esprimono categorie), testo, serie temporali, immagini, video, dati di elevato valore (edifici, indirizzi, dati meteorologici, dati statistici, ecc.), dati generati da sensori (es.: IoT), banche dati di interesse nazionale, dati dinamici.*
     - **standard**: *standard riconosciuti a livello nazionale e/o internazionale cui si conformano le banche dati utilizzate (es.: modelli, specifiche, ontologie, vocabolari controllati).*
   - 5.3. **Velocità.**
     Rapidità di creazione, memorizzazione, analisi o visualizzazione dei dati, inclusa l'elaborazione in tempo reale.
   - 5.4. **Variabilità.**
     Cambiamenti nella velocità, struttura, formato, semantica o qualità dei dati.

5.5. Qualità.
Completezza e accuratezza dei dati rispetto al contenuto semantico e alla sintassi. Conformità con le quattro caratteristiche di qualità (accuratezza semantica e sintattica, coerenza, completezza, attualità) del modello di qualità dei dati definito dalla ISO/IEC 25012:2008 Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model https://www.iso.org/standard/35736.html.

5.6. Dati personali.
Presenza di dati personali o sensibili o altri attributi regolamentati dal punto di vista legale (cfr. cap. 10).

5.7. Organizzazione.
Struttura di memorizzazione e gestione dei dati (es. centralizzata, distribuita, mista).

6. Indicatori di prestazione (Key Performance Indicator - KPI).
Indicatori utilizzati per valutare le prestazioni o l'utilità del sistema di IA, misurando il raggiungimento degli obiettivi prefissati. (es.: Recall (sensibilità), Precision (valore predittivo positivo), Customer satisfaction (Soddisfazione del cliente), Algorithm accuracy (Accuratezza dell'algoritmo), Task completion rate (Tasso di completamento dei compiti), Cost Reduction (Riduzione dei Costi), Efficiency (Efficienza).

7. Caratteristiche del caso d'uso

7.1. Obiettivo.
Identifica il compito principale supportato dal caso d'uso, con particolare riferimento agli obiettivi e ambiti prioritari di applicazione individuati al paragrafo 4.3.

7.2. Funzionalità.
Specifica il mix di funzionalità dei sistemi di IA secondo la classificazione delle presenti Linee guida (cfr. par. 4.5.1).

7.3. Ruolo della PA.
Specifica se la PA riveste il ruolo di fornitore o deployer della soluzione di IA.

7.4. Livello di automazione.
Definisce il livello di automazione del sistema IA nel caso d'uso, variando da piena autonomia a completa supervisione umana. Si utilizzi la classificazione dei livelli di automazione dei sistemi di IA (ISO/IEC 22989:2022 Information technology - Artificial intelligence - Artificial intelligence concepts and terminology https://www.iso.org/standard/74296.html.)

7.5. Modelli di IA.
Specifica i metodi, modelli o framework di IA utilizzati per l'implementazione del caso d'uso, specificandone le caratteristiche (es.: OSS, proprietario, ecc.).

7.6. Livello di Rischio.
Specifica il livello di rischio del sistema di IA utilizzato secondo la classificazione dell'AI Act (cfr. par. 3.3).

7.7. Piattaforma.
Indica la piattaforma hardware o software utilizzata per lo sviluppo e il deployment del sistema IA.

7.8. Topologia.
Descrive l'architettura di rete utilizzata per il deployment del sistema, ad esempio distribuita, centralizzata o ibrida.

7.9. Fornitori.
Descrive quali fornitori sono coinvolti nella catena del valore del caso d'uso e con quali finalità.

7.10. Best practice e precursori.
Descrive, referenziandole, le eventuali implementazioni dello stesso caso d'uso da parte di altre PA o di altri soggetti.

8. Minacce e vulnerabilità.
Indica le minacce e le vulnerabilità rilevanti per il caso d'uso, come bias, uso improprio o non previsto del sistema di IA, minacce alla sicurezza, sfide alla responsabilità e rischi per la privacy.

9. Sfide e problematiche.
Descrive le principali sfide e problematiche affrontate nel caso d'uso. Tra queste possono rientrare difficoltà tecniche, ostacoli normativi, questioni etiche o sociali e sfide legate all'implementazione e all'adozione del sistema di IA.

10. Considerazioni sulla fiducia (trustworthiness).

Descrive come il caso d'uso affronta le caratteristiche della fiducia inclusi nei principi definiti al paragrafo 3.4.

10.1. Mitigazione dei bias.

Descrizione di come sono identificati e mitigati i bias, tra cui: bias cognitivi umani, bias di conferma, bias nei dati, bias statistici.

10.2. Questioni etiche e sociali.

Descrive come le questioni etiche e sociali relative al sistema di IA siano state identificate, analizzate e mitigate. Sono da considerare come riferimento i principi etici fondamentali e i requisiti per un'IA affidabile individuati a livello europeo dall'EU High-Level Expert Group oltre alle considerazioni espresse al capitolo 6.

10.3. Spiegabilità.

Descrive il grado di spiegabilità dei risultati del sistema e come l'organizzazione ha affrontato obiettivi e sfide legati alla spiegabilità.

10.4. Controllabilità e supervisione umana (Controllability and human oversight).

Descrive il grado di controllabilità del sistema di IA e le modalità attraverso cui l'organizzazione affronta obiettivi e sfide connesse a questa caratteristica. Viene illustrato come il sistema di IA consenta a un operatore umano o a un agente esterno di intervenire nel funzionamento del sistema, specificando quali componenti del sistema possono essere controllati, da chi e in quali circostanze. Un aspetto chiave della controllabilità è l'identificazione di quale agente (es. fornitori di servizi, produttori, utenti finali, o autorità regolatorie) può controllare quali componenti del sistema di IA.

10.5. Predicibilità.

Descrive il grado di predicibilità dei risultati del sistema di IA e come l'organizzazione gestisce gli obiettivi e le sfide legate alla predicibilità.

10.6. Trasparenza.

Descrive del livello di trasparenza del sistema di IA e come l'organizzazione intende affrontare gli obiettivi e le sfide per garantire la trasparenza.

10.7. Verifica.

Descrive come l'organizzazione affronta gli obiettivi e le sfide legati alla verifica dei requisiti e delle funzionalità del sistema di IA.

10.8. Robustezza, affidabilità e resilienza.

Descrizione il livello di robustezza, affidabilità e resilienza del sistema di IA e come l'organizzazione gestisce gli obiettivi e le sfide legati a queste caratteristiche.

11. Conformità ai requisiti dell'IA Act (solo per i sistemi ad alto rischio).

Descrizione di come sono soddisfatti i requisiti alla sezione 2 dell'AI Act.

12. Utilizzo degli standard e opportunità di standardizzazione.

Descrive le opportunità o i requisiti di standardizzazione associati al caso d'uso. Include riferimenti a standard attualmente utilizzati o che potrebbero essere sviluppati per migliorare l'adozione dell'IA.

G. Funzionalità dell'IA

Il documento presenta la classificazione delle principali funzionalità dei sistemi di IA introdotta al paragrafo 4.5.1. delle linee guida sull'adozione dell'IA nella PA.

Nell'ordine sono presentate con i propri identificativi: funzionalità, eventuali sotto-funzionalità ed esempi di applicazione.

F.1. Analisi e classificazione del testo. L'IA può analizzare e interpretare grandi quantità di dati testuali e linguistici (comunicazioni e documenti), classificandoli automaticamente per argomento e analizzando tematiche specifiche. Questa capacità è particolarmente utile per gestire volumi elevati di comunicazioni e documenti, consentendo di identificare rapidamente le priorità e ottimizzare la gestione delle informazioni.

F.1.a. Classificazione di comunicazioni e documenti: organizzazione e categorizzazione di documenti e messaggi agevolando la gestione documentale.

F.1.b. Analisi delle comunicazioni: identificazione dei bisogni e delle richieste degli utenti (cittadinanza, imprese e altre PA) attraverso l'elaborazione di comunicazioni al fine di migliorare la tempestività e la precisione delle risposte.

F.1.c. Analisi dei feedback sui servizi: rilevazione automatica del livello di soddisfazione degli utenti (cittadinanza, imprese, altre PA) e delle aree di miglioramento dei servizi**.**

F.2. Analisi di contenuti non testuali. L'IA può essere applicata per analizzare contenuti non testuali, come immagini, video e audio, estraendo informazioni significative.

F.2.A. Analisi di immagini e video. L'IA analizza immagini e video per riconoscere persone, animali, oggetti e scene, effettuando operazioni di estrazione di informazioni visive. Questo approccio è utile per applicazioni che richiedono un monitoraggio visivo e la gestione delle immagini.

F.2.A.a. Monitoraggio ambientale: analisi dei cambiamenti territoriali e pianificazione delle attività di prevenzione e intervento in caso di emergenza.

F.2.A.b. Monitoraggio del traffico: analisi dei flussi di traffico in tempo reale e supporto alla gestione della circolazione.

F.2.A.c. Monitoraggio delle infrastrutture: analisi delle immagini per rilevare anomalie in infrastrutture o ambienti urbani al fine della manutenzione preventiva.

F.2.B. Analisi (non linguistica) di audio. L'IA estrae informazioni da segnali audio, utilizzandole per classificare, memorizzare o recuperare contenuti, senza analisi linguistica. Questo tipo di analisi consente di interpretare e catalogare segnali sonori per varie finalità.

F.2.B.a. Monitoraggio ambientale: identificazione di eventi anomali o critici (dissesto idro-geologico, inquinamento acustico) al fine di interventi di emergenza o di prevenzione.

F.2.B.b. Monitoraggio del traffico: rilevamento di segnali sonori per monitorare la circolazione stradale e individuare situazioni di emergenza.

F.2.B.c. Tutela del patrimonio culturale: monitoraggio acustico in musei, siti archeologici o edifici storici per rilevare indicatori di degrado o intrusioni.

F.3. Generazione di nuovo contenuto. L'IA può essere utilizzata per creare contenuti nuovi e originali in vari formati (testo, immagini, audio video, modelli 3D), rispondendo a specifiche richieste dell'utente.

F.3.A. Generazione di linguaggio, conversazioni e traduzioni. L'IA è in grado di generare contenuti in forma di testo o conversazioni in risposta a comandi o domande anche in linguaggio naturale. Questa classe di funzionalità è utile o per produrre nuovi testi in diverse lingue e per rispondere in tempo reale alle richieste di informazione.

F.3.A.a. Generazione di risposte a domande frequenti: utilizzo di chatbot multilingue per fornire informazioni immediate su procedure amministrative, scadenze o servizi disponibili, migliorando l'accessibilità ai cittadini.

F.3.A.b. Traduzione di documenti: traduzione in tempo documenti ufficiali per facilitare il multilinguismo e agevolare l'interazione con organizzazioni estere.

F.3.A.c. Sintesi di documenti: creazione di versioni sintetizzate di documenti per renderli più comprensibili e fruibili dagli utenti finali.

F.3.A.d. Generazione documenti: produzione di documenti (es. atti, provvedimenti) adattati alle specifiche esigenze degli utenti finali.

F.3.A.e. Assistente virtuale: implementazione di assistenti virtuali disponibili in più lingue per supportare l'inclusione e rispondere alle esigenze di cittadini italiani e stranieri.

F3.B. Generazione di immagini, video e audio. L'IA può generare immagini, video o audio sulla base di descrizioni testuali o comandi vocali, offrendo alle amministrazioni strumenti per produrre contenuti multimediali di supporto ai servizi pubblici**.**

F.3.B.a. Creazione di video educativi o informativi: generazione di video basati su testi descrittivi per campagne di sensibilizzazione su temi specifici (es.: salute, ambiente, beni culturali).

F.3.B.b. Produzione di immagini per materiale promozionale: creazione di grafiche e immagini per brochure, siti web o campagne promozionali.

F.3.B.c. Sintesi vocale per contenuti accessibili: generazione di contenuti audio per migliorare l'accessibilità di documenti e servizi digitali, come audioguide o letture di testi pubblici.

F.3.B.d. Simulazioni per la formazione: generazione di scenari visivi o audio realistici per formare il personale della PA in ambiti specifici (es.: protezione civile, sicurezza sul lavoro).

F3.C. Generazione di modelli 3D. L'IA può creare modelli 3D a partire da specifiche tecniche, requisiti e vincoli forniti dall'utente. Questa capacità è particolarmente utile per la progettazione e visualizzazione di progetti fisici o ambienti virtuali, supportando settori come l'architettura e l'urbanistica.

F.3.C.a. Progettazione urbanistica e riqualificazione: creazione di modelli 3D di aree urbane per pianificare interventi di riqualificazione o nuovi progetti infrastrutturali.

F.3.C.b. Visualizzazione di progetti edilizi: generazione di rappresentazioni 3D di edifici o spazi per facilitare la comunicazione tra progettisti, amministrazioni e cittadini.

F.3.C.c. Simulazioni ambientali: realizzazione di modelli 3D per simulare impatti ambientali, come il comportamento del traffico, la gestione delle risorse idriche o la resilienza agli eventi climatici estremi.

F.4 Suggerimenti e raccomandazioni. L'IA può essere impiegata per fornire suggerimenti personalizzati e raccomandazioni basate sui dati specifici degli utenti o su similitudini rilevate con altri utenti. Tale finalità può essere utilizzata sia ad uno esterno, verso cittadini e imprese, sia ad uno interno, verso i dipendenti della PA

F.4.A. Suggerimenti e raccomandazioni (utenti). L'IA analizza le informazioni fornite dagli utenti (in modo diretto o indiretto) o dall'ambiente (tramite sensori) per offrire suggerimenti personalizzati. Questo approccio è utile per aiutare i cittadini e le imprese a trovare rapidamente servizi o informazioni rilevanti, semplificando la navigazione dei portali pubblici e ottimizzando il supporto.

F.4.A.a. Orientamento ai servizi pubblici: fornire suggerimenti personalizzati sui servizi pubblici, basandosi sul profilo dell'utente.

F.4.A.b. Suggerimenti di opportunità: identificare e fornire suggerimenti di opportunità (lavoro, formazione) agevolazioni e incentivi in base al profilo dell'utente.

F.4.A.c. Esperto virtuale - Supporto alle richieste amministrative: fornire indicazioni specifiche su procedure, documenti o prerequisiti necessari per completare una richiesta amministrativa.

F.4.B. Suggerimenti e raccomandazioni (operatori della PA). L'IA può fornire un sistema di suggerimenti mirati per agevolare le attività degli operatori della PA nel corso dell'attività ordinaria sia nelle procedure amministrative che riguarda cittadini ed imprese, sia nel rispetto dei regolamenti interni, migliorando l'efficienza e la precisione nelle operazioni quotidiane.

F.4.B.a. Esperto Virtuale -Supporto nella gestione delle procedure amministrative: fornire indicazioni sui passi della procedura, suggerire norme applicabili, informazioni necessarie all'iter e relative fonti, al fine di garantendo la conformità normativa e semplificare la gestione dell'iter.

F.4.B.b. Esperto virtuale - Supporto alla conformità normativa: suggerire aggiornamenti normativi o regolamenti pertinenti alle attività svolte dall'operatore, assicurando che le procedure siano in linea con la normativa vigente.

F.4.B.c. Esperto virtuale - Gestione documentale: fornire suggerimenti per la classificazione, protocollazione e gestione dei documenti garantendo una gestione più rapida e ordinata degli archivi.

F.4.B.d. Identificazione di priorità operative: segnalare scadenze critiche per evitare ritardi nelle operazioni, suggerendo azioni preventive o prioritarie per rispettare i termini previsti.

F.5 Decisioni basate sui dati. L'IA supporta le decisioni della PA basandosi su evidenze concrete, grazie alla capacità di analizzare dati e fornire informazioni utili e accurate.

F.5.A Identificazione di pattern e realizzazione di previsioni a partire dai dati. L'IA è utilizzata per analizzare grandi quantità di dati e identificare modelli, regolarità o anomalie che potrebbero non essere facilmente visibili all'analisi umana. Questa capacità permette di fare previsioni su comportamenti futuri di variabili specifiche, offrendo alle amministrazioni strumenti avanzati per anticipare cambiamenti e reagire in modo proattivo.

F.5.A.a. Previsioni sulla domanda di servizi pubblici: analizzare dati storici e prevedere picchi o variazioni nella domanda di servizi pubblici, consentendo una pianificazione più efficiente delle risorse.

F.5.A.b. Pianificazione predittiva: ottimizzare o distribuire in modo proattivo risorse in base a esigenze future (es.: trasporti, salute).

F.5.A.c. Individuazione di anomalie nei dati finanziari o contabili: identificare schemi o transazioni anomale nei dati finanziari o contabili, contribuendo a rilevare potenziali irregolarità, inefficienze o frodi.

F.5.A.d. Gestione delle emergenze: prevedere possibili eventi critici, come disastri naturali o sovraccarichi infrastrutturali, per migliorare la preparazione e la risposta tempestiva alle emergenze.

F.5.B. Ottimizzazione. L'IA permette di ottimizzare determinate funzioni obiettivo attraverso la modifica delle variabili, rispettando specifici vincoli operativi o di risorse. Questo è particolarmente utile per le amministrazioni che devono massimizzare l'efficienza delle proprie operazioni pur rispettando limiti di budget o altre restrizioni.

F.5.B.a. Allocazione ottimale delle risorse umane: pianificazione strategica delle attività del personale per garantire l'efficienza operativa e la copertura dei servizi.

F.5.B.b. Ottimizzazione del budget: definizione delle priorità di spesa e allocazione delle risorse finanziarie per ottenere il massimo impatto con i fondi disponibili, rispettando vincoli normativi e obiettivi strategici.

F.5.B.c. Ottimizzazione dei percorsi: identificazione dei tragitti più efficienti per migliorare la puntualità, ridurre i costi operativi e ottimizzare la frequenza dei mezzi (es. trasporto pubblico, raccolta rifiuti).

F.5.B.d. Gestione ottimale delle risorse energetiche negli edifici pubblici: regolazione automatica dei consumi energetici per ridurre sprechi e rispettare obiettivi di sostenibilità.

F.5.B.e. Ottimizzazione della gestione delle code nei servizi: pianificazione e gestione delle code e lista d'attesa per garantire la massima efficienza nell'assegnazione dei posti disponibili e riducendo i tempi di attesa per i cittadini (es.: asili nido, strutture sanitarie).

F.6. Orchestrazione di processi. L'IA può coordinare in modo dinamico e intelligente le diverse fasi di un processo amministrativo, gestendo l'automazione delle attività tramite workflow flessibili e adattivi (non deterministici). L'IA permette di orchestrare un flusso di lavoro, gestendo automaticamente le attività e ottimizzando l'interazione tra gli attori del processo. L'orchestrazione dinamica e adattiva consente alla PA di rispondere a condizioni specifiche o a imprevisti, adattando i processi per soddisfare al meglio le esigenze dei cittadini e migliorando l'efficacia complessiva.

F.6.a. Orchestrazione delle richieste di assistenza: gestione intelligente e dinamico delle richieste di assistenza che coinvolgono più unità organizzative interne e/o esterne.

F.6.b. Automazione dei processi di approvazione: implementazione di flussi adattivi per l'approvazione di pratiche amministrative, con regole flessibili che si adattano a condizioni specifiche.

F.6.c. Ottimizzazione dei flussi per risorse umane e logistica: orchestrazione dei processi legati alla gestione del personale e delle attività logistiche, migliorando la pianificazione e l'utilizzo delle risorse disponibili.

H. Procedure di governance

Il documento fornisce un riferimento alle PA per definire le procedure di governance utili per l'adozione dell'IA

La governance dell'IA rappresenta la capacità organizzativa di controllare la formulazione e l'implementazione di una strategia tecnologica[74]. Per una gestione efficace, si suggerisce di suddividerla in tre pratiche principali:

- **pratiche procedurali**: definiscono il modo in cui le PA eseguono le funzioni operative per una gestione efficace dell'IA.
- **pratiche strutturali**: identificano i ruoli, le responsabilità e i decisori chiave all'interno dell'organizzazione.
- **pratiche relazionali**: includono il coordinamento tra gli stakeholder e le interazioni tra attori interni ed esterni all'ente.

Queste pratiche sono attuate a tre livelli:

- **strategico**: riguarda le decisioni di lungo termine sull'adozione dell'IA, gli obiettivi e i vincoli da considerare, nonché le collaborazioni necessarie. Le decisioni sono prese dai vertici dell'ente.
- **tattico**: si concentra sulle scelte di medio termine, coinvolgendo specifiche unità organizzative e definendo come le soluzioni di IA debbano essere integrate nei processi operativi.
- **operativo**: riguarda l'implementazione concreta e quotidiana delle soluzioni di IA, con un focus su risultati tangibili.

L'incrocio tra le tre dimensioni della governance (procedurale, strutturale e relazionale) e i tre livelli organizzativi (strategico, tattico e operativo) genera una matrice figura di riferimento (Figura G.1.), che supporta le PA nel:

- assegnare responsabilità e azioni specifiche a ogni livello di governance.
- identificare eventuali aree di miglioramento o lacune nella gestione dell'IA.
- garantire il corretto allineamento tra le decisioni strategiche, tattiche e operative.

[74] JRC della Commissione Europea "Competences and governance practices for AI in the public sector" https://publications.jrc.ec.europa.eu/repository/handle/JRC138702

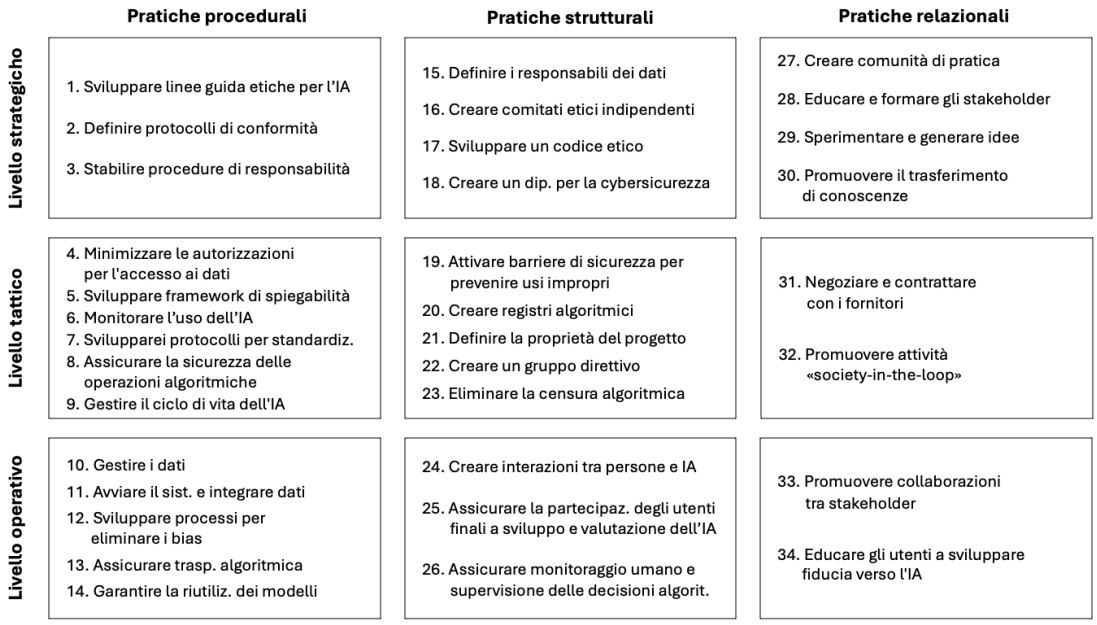

Figura H.1. Framework per le pratiche di organizzazione e governance dell'IA in ambito pubblico (fonte: JRC della Commissione Europea)

H.1. Pratiche relative a procedure

Le PA DOVREBBERO adottare procedure adeguate a garantire un utilizzo sicuro, etico e conforme dell'IA. Di seguito si riportano alcune procedure che le PA DOVREBBERO integrare nella governance dell'IA.

Livello Strategico

1. **Linee guida etiche per l'IA**. Stabilire processi formalizzati che garantiscano che le soluzioni di IA rispettino i più alti standard etici (cfr. cap. 7).
2. **Protocolli di conformità.** Garantire l'allineamento delle soluzioni IA ai requisiti normativi e regolatori applicabili[75].
3. **Procedure di responsabilità.** Definire ruoli e azioni in caso di malfunzionamenti o impatti indesiderati, assicurando la continuità operativa anche mediante sistemi alternativi all'IA.

Livello Tattico

4. **Gestione degli accessi ai dati.** Minimizzare le autorizzazioni e stabilire criteri di accesso basati sul principio della necessità.
5. **Framework di spiegabilità.** Implementare processi formalizzati per spiegare i risultati del sistema di IA alle parti interessate con il necessario livello di informazione.
6. **Monitoraggio dell'IA.** Definire sistemi per il controllo continuo delle performance e dell'impatto dei sistemi di IA.

[75] Riguardo la conformità ai requisiti dell'AI Act da parte dei sistemi IA ad alto rischio si tega in considerazione l'art 6 del suddetto Regolamento che prevede la valutazione della conformità di terza parte prima dell'immissione sul mercato del prodotto di IA.

7. **Standardizzazione dei sistemi IA.** Adottare la normativa tecnica allo stato dell'arte riguardo lo sviluppo, l'implementazione e i test delle soluzioni IA (cfr. 4.4., Allegato **Errore. L'origine riferimento non è stata trovata.**).
8. **Sicurezza delle operazioni algoritmiche.** Prevenire manipolazioni e usi impropri degli algoritmi, garantendo affidabilità e protezione.
9. **Gestione del ciclo di vita dell'IA.** Pianificare aggiornamenti e manutenzione continua per assicurare efficacia e conformità nel tempo.

Livello Operativo

10. **Gestione dei dati.** Stabilire le procedure a cui i dati devono essere sottoposti quando utilizzati nei cicli di vita dei progetti di IA[76] (cfr. cap. 9).
11. **Avvio e integrazione dei sistemi IA.** Adottare protocolli di integrazione di sistemi e dati, al fine di trasferire e unire efficacemente insiemi di dati provenienti da fonti diverse, fermo restando il rispetto della protezione dei dati personali.
12. **Eliminazione dei bias.** Sviluppare metodologie per identificare e correggere eventuali pregiudizi nei dati e nei modelli di IA.
13. **Trasparenza algoritmica.** Adottare processi in grado di spiegare perché gli algoritmi producono determinati risultati e quale sia la logica sottostante ai loro esiti
14. **Riusabilità dei modelli IA.** Definire procedure per la riusabilità dei modelli chiarendo i contesti nei quali un modello può essere riutilizzato.

H.2. Pratiche strutturali

Le PA DOVREBBERO garantire che le responsabilità e le autorità relative all'IA siano chiaramente assegnate e comunicate all'interno dell'ente. Di seguito si riportano alcune procedure che le PA DOVREBBERO integrare nella governance dell'IA:

Livello Strategico

15. **Definizione dei responsabili dei dati.** Designare figure incaricate della gestione sicura e conforme dei dataset utilizzati nei sistemi IA.
16. **Comitati etici indipendenti.** Istituire organismi autonomi per valutare l'impatto etico e sociale dell'IA e prevenire conflitti di interesse.
17. **Codice etico per l'IA.** Elaborare e aggiornare un codice etico che integri principi di trasparenza, equità e responsabilità nell'uso dell'IA.

[76] Come specificato nel report del comitato ad hoc per lo sviluppo dell'IA del Consiglio d'Europa "Artificial Intelligence in Public Sector".

18. **Struttura per la cybersicurezza.** Istituire un'unità dedicata alla protezione dei sistemi IA da rischi informatici e minacce alla sicurezza[77].

Livello Tattico

19. **Barriere di sicurezza.** Implementare misure di protezione per prevenire usi impropri e ridurre i rischi operativi.
20. **Registri algoritmici.** Creare database che documentino gli algoritmi IA in uso, garantendo trasparenza e tracciabilità.
21. **Proprietà dei progetti IA.** Definire chiaramente ruoli e responsabilità nei progetti IA, con particolare attenzione ai dati e allo sviluppo al fine identificare le responsabilità delle decisioni dei sistemi di IA.
22. **Gruppo direttivo per l'IA.** Costituire un organo tecnico a supporto del responsabile dell'IA dell'ente incaricato affiancarlo nelle decisioni che riguardano i sistemi di IA durante il loro intero ciclo di vita.
23. **Censura algoritmica.** Eliminare limitazioni arbitrarie all'accesso alle informazioni generate o gestite dall'IA.

Livello Operativo

24. **Interazioni tra persone e IA.** Stabilire linee guida per regolare il rapporto tra agenti artificiali e utenti umani.
25. **Coinvolgimento degli utenti finali.** Garantire la partecipazione degli utenti nel processo di sviluppo e valutazione delle soluzioni IA.
26. **Monitoraggio e supervisione umana.** Assicurare un controllo continuo sulle decisioni algoritmiche per mitigare o prevenire azioni indesiderate.

H.3. Pratiche relazionali

Le pratiche relazionali includono aspetti complementari a quelli procedurali e strutturali. Di seguito si riportano alcune procedure che le PA DOVREBBERO integrare nella governance dell'IA.

Livello Strategico

27. **Creare comunità di pratica.** Costituire o partecipare a centri di competenza, reti informali o collaborazioni tra PA per condividere esperienze e best practice sull'IA, facilitando l'apprendimento e il supporto reciproco, soprattutto per enti di dimensioni più piccole.

[77] L'art. 8 della legge 28 giugno 2024, n. 90, Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici, ha introdotto l'obbligo di individuare una struttura per la cybersicurezza per determinate PA e soggetti pubblici.

28. **Educare e formare gli stakeholder interni ed esterni.** Promuovere percorsi di formazione per gli addetti della PA e gli stakeholder esterni coinvolti, garantendo competenze adeguate allo sviluppo, l'uso e la supervisione dell'IA (cfr. cap. 8).
29. **Sperimentare e generare idee.** Sviluppare iniziative di sperimentazione e innovazione, anche mediante gli Spazi di sperimentazione e sviluppo per l'IA definiti dal Piano triennale per l'informatica nella PA, avvalendosi della collaborazione con università, centri di ricerca e con il settore privato, per accelerare l'adozione e il miglioramento delle applicazioni IA nella PA.
30. **Promuovere il trasferimento di conoscenze.** Favorire lo scambio di competenze e risultati tra progetti IA, sviluppando canali formali e informali per la condivisione di esperienze tra individui e dipartimenti sia all'interno che all'esterno dell'organizzazione, al fine di apprendere dai progetti e ridurre i tempi di consegna.

Livello Tattico

31. **Negoziare e contrattare con i fornitori.** Stabilire criteri chiari per la selezione e gestione dei fornitori di soluzioni IA, garantendo trasparenza e allineamento agli obiettivi strategici e normativi della PA (cfr. Linee guida per il procurement dell'IA).
32. **Promuovere attività *society-in-the-loop*.** Coinvolgere gli utenti finali nella progettazione e nell'evoluzione delle soluzioni IA, raccogliendo feedback e suggerimenti attraverso consultazioni pubbliche e strumenti di partecipazione.

Livello Operativo

33. **Promuovere collaborazioni tra stakeholder.** Facilitare il dialogo e la cooperazione tra enti pubblici, imprese e cittadini per migliorare la governance e l'efficacia delle applicazioni IA nella PA.
34. **Educare gli utenti a sviluppare fiducia verso l'IA.** Attuare iniziative di sensibilizzazione e trasparenza per rendere i cittadini consapevoli del funzionamento delle soluzioni IA e aumentarne la fiducia e l'accettazione.

I. Indicatori di prestazione

Il documento riporta i principali indicatori di prestazione (KPI) raccomandati nel ciclo di adozione di soluzioni di IA nella PA, con l'obiettivo di fornire uno strumento operativo nella misurazione, valutazione e monitoraggio delle prestazioni e dell'efficacia delle soluzioni IA.

Per ciascun KPI sono riportati:

- **ambito del KPI**: identifica l'area di valutazione della soluzione AI (ad es., user experience, equità).
- **KPI**: specifica l'indicatore di prestazione individuato per l'ambito di riferimento.
- **descrizione**: offre una breve definizione dell'indicatore, chiarendo il suo ruolo e il suo scopo nella valutazione della soluzione.
- **formula**: descrive il calcolo matematico dell'indicatore, fornendo un metodo standardizzato per misurare le prestazioni.
- **SLA**: indica il livello di servizio atteso, utile per definire obiettivi di prestazione chiari e verificabili. Gli SLA possono variare a seconda della fase di sviluppo della soluzione di IA e del TRL (es.: POC, sistema in produzione). Gli SLA riportati nella successiva tabella fanno riferimento ad un sistema operativo (TRL 9).

I KPI riportati di seguito, e i relativi SLA, sono stati definiti sulla base delle seguenti fonti: ISO/IEC 25010:2023 (Systems and Software Quality Requirements and Evaluation - SQuaRE), ISO/IEC 27001:2022 (Information Security Management Systems - Requirements), ISO/IEC 38500:2024 (Corporate Governance of Information Technology), ITIL (Information Technology Infrastructure Library).

Affidabilità e stabilità

Disponibilità (Uptime - UP)

- *descrizione*: misurazione in % dell'affidabilità operativa della soluzione AI, con particolare riferimento al tempo medio di inattività e alle garanzie di continuità del servizio.
- *formula:* (tempo di uptime) / (tempo totale) x 100
- *SLA:* ≥ 99.5%

Tolleranza agli errori (TLE)

- *descrizione.* frequenza e natura degli errori riscontrati (errori critici, errori non bloccanti), con definizione dei livelli di rischio accettabili.
- *formula:* (N. errori critici) / (N. totale operazioni) x 100
- *SLA:* ≤ 1% di errori critici per mese operativo

Accuratezza e precisione dei ri*sultati*

Tasso di precisione (TP)

- *descrizione:* percentuale di previsioni o output corretti rispetto al numero totale di elaborazioni, con soglia minima stabilita per l'accettazione del sistema.
- *formula:* (previsioni corrette o output corretti) / (N. totale previsioni) x 100

- *SLA:* ≥ 95%

Scalabilità e flessibilità

Capacità di scalabilità (CS)

- descrizione: numero massimo di richieste o utenti simultanei supportati senza degrado delle performance.
- *formula:* (N. utenti simultanei supportati) / (N. utenti richiesti) x 100
- *SLA:* ≥ 100% (supporto totale del carico richiesto)

Flessibilità di adattamento (FLA)

- *descrizione:* valutazione della capacità della soluzione di integrare futuri sviluppi e moduli aggiuntivi con un impatto minimo sui sistemi esistenti.
- formula: (moduli integrabili senza degrado) / (N. totale moduli) x 100
- *SLA:* ≥ 95% di moduli integrabili senza degrado delle performance

Conformità normativa

Conformità alle normative GDPR e/o AI Act

- *descrizione:* assicurazione che la soluzione sia conforme alla normativa sulla protezione dei dati (GDPR e/o AI Act) e altre normative rilevanti.
- *formula:* (requisiti GDPR e/o AI Act conformi) / (totale requisiti GDPR e/o AI Act) x 100
- *SLA:* 100% (conformità completa)

Indice di conformità (Compliance Risk Index - CRI)

- *descrizione:* valuta il livello di conformità normativa stimata della soluzione. Un CRI alto assicura che la soluzione rispetti gli obblighi normativi, riducendo il rischio legale e operativo.
- *formula:* (N. requisiti normativi soddisfatti) / (totale requisiti normativi rilevati) x 100
- *SLA:* CRI ≥ 90% per soluzioni altamente conformi

Sicurezza

Misure di sicurezza (MS)

- *descrizione:* tipo e livello delle misure di sicurezza implementate per prevenire accessi non autorizzati, protezione dei dati e resilienza agli attacchi informatici.
- *formula:* (N. di attacchi bloccati) / (N. di attacchi totali) x 100
- *SLA:* ≥ 98% di attacchi bloccati

Tempo di risposta e prestazioni operative

Tempo medio di risposta (RT)

- *descrizione:* tempo medio richiesto per l'elaborazione delle richieste, con riferimento specifico a scenari di picco e continuità delle prestazioni.
- *formula:* (somma dei tempi di risposta per ogni richiesta) / (N. richieste totali)
- *SLA:* ≤ 200 ms per richiesta in scenari di utilizzo normale

Efficienza operativa (EO)

- *descrizione:* quantificazione del risparmio di tempo generato dalla soluzione in merito alla risposta ad un determinato servizio.

- *formula:* ((tempo impiegato per l'esecuzione della determinata attività - tempo impiegato dalla soluzione IA per eseguire la medesima attività) / tempo impiegato per l'esecuzione della determinata attività) x 100
- *SLA:* ≥ 50% (metà del tempo risparmiato)

Capacità di adattamento a variazioni di carico (AVC)

- *descrizione:* valutazione della risposta del sistema a carichi variabili e gestione delle risorse nei casi di elevato utilizzo.
- *formula:* (performance sotto carico elevato) / (performance normale) x 100
- *SLA:* ≥ 90% delle performance normali in scenari di elevato utilizzo

Sostenibilità economica e ROI

Costi di implementazione e manutenzione (CIM)

- *descrizione:* analisi dettagliata dei costi iniziali e ricorrenti associati alla soluzione.
- *formula:* (reddito operativo) / (capitale investito operativo netto) x 100
- *SLA:* determinata rispetto al budget massimo disponibile per la PA

Livelli di automazione

Tasso di automazione (Automation rate - AR)

- *descrizione:* indica la percentuale di processi automatizzati, riducendo il carico manuale.
- *formula:* (numero di task automatizzati) / (numero di task totali) x 100
- *SLA:* ≥ 50% di task automatizzati

Riduzione dell'intervento umano (Reduction in Human Intervention - RHI)

- *descrizione:* misura la riduzione della necessità di interventi umani grazie all'AI.
- *formula:* ((numero di interventi umani prima - interventi umani dopo) / numero di interventi umani prima) x 100
- *SLA:* ≥ 30% di riduzione

Eco-sostenibilità

Tasso di efficienza energetica (Energy efficiency rate - EER)

- *descrizione:* misura il risparmio energetico ottenuto con la soluzione AI rispetto ai metodi precedenti.
- *formula:* ((consumo energetico attuale - consumo con soluzione AI) / consumo energetico attuale) x 100
- *SLA:* ≥ 20% di riduzione

Etica

Tasso di equità (Equity rate - ER)

- *descrizione:* misura il grado di equità dell'algoritmo AI, verificando che non vi siano discriminazioni basate su dati protetti e/o sensibili (es. genere, età).
- *formula:* (numero di decisioni non discriminatorie) / (numero di decisioni totali) x 100
- *SLA:* ≥ 95%, suggerendo un'alta equità

User experience

Tasso di spiegabilità (Explainability rate - XR)

- *descrizione:* misura la capacità dell'algoritmo di fornire decisioni comprensibili e spiegabili per gli utenti e gli operatori. Un XR elevato assicura che le risposte AI siano trasparenti e motivate, favorendo la fiducia nella tecnologia.
- *formula:* (numero di decisioni spiegabili con chiarezza) / (numero di decisioni totali) x 100
- *SLA:* ≥ 90%, garantendo che la maggior parte delle decisioni sia chiaramente spiegabile.

Punteggio dell'esperienza utente (User experience score - UES)

- *descrizione:* rappresenta la facilità d'uso e la soddisfazione degli utenti che interagiscono con la soluzione AI. Il punteggio può derivare da sondaggi o valutazioni degli utenti su parametri come la semplicità d'interazione, la velocità di risposta e la chiarezza delle informazioni.
- *formula:* (punteggio medio dei feedback utente) / (punteggio massimo) x 100
- *SLA:* ≥ 80% di soddisfazione, suggerendo una user experience positiva.

Formazione

Tasso di personale formato (Training coverage rate - TCR)

- *descrizione:* misura la percentuale di personale della PA formato per utilizzare efficacemente la soluzione AI. Assicura che il personale sia adeguatamente preparato per l'adozione della tecnologia.
- *formula:* (numero di dipendenti formati) / (numero totale di dipendenti coinvolti) x 100
- *SLA:* ≥ 90% del personale target formato.

Supporto fornito alla formazione continua (Continuous learning support rate - CLSR)

- *descrizione:* valuta il grado di supporto fornito alla formazione continua per il personale, monitorando la disponibilità e l'accesso a sessioni di aggiornamento o moduli avanzati.
- *formula:* (numero di sessioni di formazione continua offerte) / (numero minimo di sessioni pianificate per SLA) x 100
- *SLA:* ≥ 100%, garantendo un supporto formativo continuativo.

Adeguamento e Manutenzione

Tempo medio di risposta (Mean time to response MTTR)

- *descrizione:* misura il tempo medio di risposta in caso di malfunzionamento o richiesta di assistenza. Un MTTR basso garantisce che i problemi vengano affrontati tempestivamente.
- *formula:* (tempo totale di risposta) / (numero di interventi richiesti)
- *SLA:* ≤ 4 ore, assicurando risposte tempestive in caso di malfunzionamenti.

Frequenza di aggiornamento e manutenzione (Frequency of updates and maintenance - FUM)

- *descrizione:* valuta la frequenza degli aggiornamenti e della manutenzione della soluzione AI, assicurando che l'algoritmo rimanga efficace e allineato alle normative.
- *formula:* (numero di aggiornamenti effettuati) / (numero di aggiornamenti pianificati) x 100

- *SLA:* ≥ 100%, con aggiornamenti completi e puntuali.