AGID | Agenzia per l'Italia digitale

LINEE GUIDA

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"

9 novembre 2026

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"

Informazioni sul documento

Tipologia: Linee Guida

Versione: 1.0

Stato: Definitivo

Data: 9 novembre 2026

Codice identificativo: Linee Guida per l'adozione di IA nella Pubblica Amministrazione

Struttura responsabile: Direzione Generale

Riferimento normativo: DPCM 12 maggio 2024, recante "Piano triennale nella Pubblica Amministrazione 2024-2026"

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 S.p.A.
- 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 S.p.A.
- Unione Province d'Italia

Sommario

Introduzione 7

1. Ambito di applicazione 8

1.1 Ambito soggettivo 8

1.2 Ambito oggettivo 8

2. Riferimenti e sigle 9

2.1 Note di lettura del documento 9

2.2 Struttura 10

2.3 Riferimenti normativi 10

2.4 Linee Guida di riferimento 17

2.5 Acronimi 19

2.6 Termini e definizioni 23

3. L'Intelligenza Artificiale 24

3.1 Ciclo di vita di un sistema di IA 28

3.2 I ruoli nella catena del valore dell'IA 31

3.3 Classificazione dei sistemi di IA sulla base del rischio 33

3.4 Principi per l'adozione dell'IA nella Pubblica Amministrazione 34

4. Modello organizzativo di adozione dell'IA 40

4.1 Strategia per l'IA 41

4.2 Analisi del contesto e delle caratteristiche della PA 43

4.3 Obiettivi e ambiti prioritari di applicazione 45

4.4 Norme tecniche 47
4.5 Casi d'uso 49
4.5.1 Requisiti 49
4.5.2 Indicatori di prestazione 50
4.6 Governance 51
4.7 Gestione del rischio 52
4.8 Valutazione d'impatto 53
4.9 Piano operativo 53
4.10 Risorse, competenze, comunicazione 54
4.11 Implementazione 55
4.12 Monitoraggio e valutazione 56
4.13 Miglioramento continuo 56
5. Conformità delle soluzioni di IA 57
5.1 Responsabilità lungo la catena del valore dell'IA 58
5.2 Monitoraggio del ciclo di vita delle soluzioni di IA 59
5.3 Misure di sorveglianza 61
5.4 Conservazione della documentazione 62
6. Governance etica dell'IA 63
7. Comunicazione 68
7.1 Misure di trasparenza 68
7.2 Obblighi di informativa 69
8. Formazione e sviluppo delle competenze 73

8.1 Indicazioni operative per le PA 85

9. Gestione e qualità dei dati 92

9.1 Tipologie di dati 95

9.2 Caratteristiche dei dati 100

9.3 Processi e governance dei dati 104

10. Protezione dei dati personali 117

11. Sicurezza cibernetica 124

11.1 Tassonomie di attacco 125

11.1.1 Evasion attacks 126

11.1.2 Poisoning attacks 126

11.1.3 Privacy attacks 127

11.1.4 Abuse attacks 128

11.2 Gestione del rischio cibernetico 128

11.3 Asset 131

11.4 Minacce 132

11.5 Obiettivi di sicurezza 133

11.6 Gestione integrata della sicurezza dei sistemi di IA 137

Introduzione

Con riferimento all'utilizzo dei sistemi di Intelligenza Artificiale (IA) da parte delle Pubbliche Amministrazioni (PA), anche ai fini dell'articolo 2, comma 1, del decreto legislativo 7 marzo 2005, n. 82, recante il "Codice dell'amministrazione digitale" (CAD), il decreto del Presidente del Consiglio dei ministri 12 gennaio 2024, recante il "Piano triennale per l'informatica nella Pubblica Amministrazione 2024-2026", prevede l'adozione delle Linee Guida per l'adozione, lo sviluppo e il procurement di sistemi di Intelligenza Artificiale nella Pubblica Amministrazione.

Le Linee Guida per l'adozione, lo sviluppo e il procurement 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".

Le Linee Guida rispettano la normativa vigente in materia di Intelligenza Artificiale sia a livello unionale sia nazionale, con particolare riferimento al Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'IA 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) e alla Legge 23 settembre 2025, n. 132 recante "Disposizioni e deleghe al Governo in materia di Intelligenza Artificiale".

1. Ambito di applicazione

1.1 Ambito soggettivo

Le presenti Linee Guida per l'adozione di sistemi 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 definiscono i criteri e le modalità per l'adozione dei sistemi di Intelligenza Artificiale, con particolare riguardo agli aspetti di conformità al quadro normativo vigente e agli impatti organizzativi derivanti dal loro utilizzo.

Esse si applicano a tutti i sistemi e ai modelli di Intelligenza Artificiale così come definiti all'articolo 3 del Regolamento (UE) 2024/1689 e richiamati all'articolo 2 della Legge 132/2025.

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.

Le presenti Linee Guida assumono come riferimento il quadro normativo dell'Unione europea e nazionale applicabile all'Intelligenza Artificiale, incluse le definizioni, i principi, gli orientamenti della Commissione Europea, le norme tecniche armonizzate e le Linee Guida previste dal Regolamento (UE) 2024/1689.

Le norme tecniche, le Linee Guida e le migliori pratiche definite a livello internazionale sono richiamate, ove pertinenti, quale supporto all'applicazione del quadro europeo e nazionale, purché compatibili con esso e non in contrasto con le disposizioni vigenti.

2.2 Struttura

Considerata la velocità con la quale le tecnologie basate sull'Intelligenza Artificiale evolvono, le Linee Guida devono garantire un adattamento costante ai cambiamenti imposti dall'incessante rivoluzione digitale. Per tale motivo sono resi disponibili sul sito istituzionale dell'Agenzia strumenti pratici di supporto ovvero documenti descrittivi dell'attività operativa di applicazione delle Linee Guida.

L'elenco aggiornato degli strumenti delle Linee Guida IA è disponibile sul sito istituzionale.

Questo consentirà un periodico aggiornamento di tali documenti per tener conto dell'evoluzione normativa e tecnologica e delle buone pratiche emergenti.

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 (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'Intelligenza Artificiale).
[OMNIBUS_AI]	Regolamento (UE) 2026/1744 del Parlamento europeo e del Consiglio, dell'8 luglio 2026, che modifica i regolamenti (UE) 2024/1689, (UE) 2018/1139 e (UE) 2023/1230 per quanto

	riguarda la semplificazione dell'attuazione di regole armonizzate sull'intelligenza artificiale (Omnibus digitale sull'IA).
[Data Act]	Regolamento (UE) 2023/2854 del Parlamento europeo e del Consiglio, del 13 dicembre 2023, 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 (UE) 2022/868 del Parlamento europeo e del Consiglio, del 30 maggio 2022, relativo alla governance europea dei dati e che modifica il regolamento (UE) 2018/1724 (regolamento sulla governance dei dati).
[CRA]	Regolamento (UE) 2024/2847 del Parlamento europeo e del Consiglio, del 23 ottobre 2024, relativo a requisiti orizzontali di cibersicurezza per i prodotti con elementi digitali e che modifica i regolamenti (UE) n. 168/2013 e (UE) 2019/1020 e la direttiva (UE) 2020/1828 (Cyber Resilience Act).
[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.
[GDPR]	Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016, 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 2012/1025]	Regolamento (UE) n. 1025/2012 del Parlamento europeo e del Consiglio, del 25 ottobre 2012, sulla normazione europea, che modifica le direttive 89/686/CEE e 93/15/CEE del Consiglio nonché le direttive 94/9/CE, 94/25/CE, 95/16/CE, 97/23/CE, 98/34/CE, 2004/22/CE, 2007/23/CE, 2009/23/CE e 2009/105/CE del Parlamento europeo e del Consiglio e che abroga la decisione 87/95/CEE del Consiglio e la decisione n. 1673/2006/CE del Parlamento europeo e del Consiglio.
[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 cibersicurezza nell'Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva (UE) 2018/1972 e che abroga la direttiva (UE) 2016/1148.
[CER]	Direttiva (UE) 2022/2557 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 (Critical Entities Resilience).

[Dir. UE 2016/680]	Direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio.
[Reg. CE 2008/765]	Regolamento (CE) n. 765/2008 del Parlamento europeo e del Consiglio, del 9 luglio 2008, che pone norme in materia di accreditamento e vigilanza del mercato per quanto riguarda la commercializzazione dei prodotti e che abroga il regolamento (CEE) n. 339/93 (New Legislative Framework).
[Dec. CE 2008/768]	Decisione n. 768/2008/CE del Parlamento europeo e del Consiglio, del 9 luglio 2008, relativa a un quadro comune per la commercializzazione dei prodotti e che abroga la decisione 93/465/CEE (New Legislative Framework).
[Reg. UE 2019/1020]	Regolamento (UE) 2019/1020 del Parlamento europeo e del Consiglio, del 20 giugno 2019, sulla vigilanza del mercato e sulla conformità dei prodotti e che modifica la direttiva 2004/42/CE e i regolamenti (CE) n. 765/2008 e (UE) n. 305/2011 (New Legislative Framework).

[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.[1]
[L. IA]	Legge 23 settembre 2025, n. 132 recante "Disposizioni e deleghe al Governo in materia di intelligenza artificiale"
[D.Lgs. 179/2026]	Decreto legislativo 7 ottobre 2026, n. 179 recante "Adeguamento della normativa nazionale alle disposizioni del regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, in materia di poteri delle Autorità nazionali e di utilizzo dell'intelligenza artificiale nella formazione".
[L. 90/2024]	Legge 28 giugno 2024, n. 90 recante "Disposizioni in materia di rafforzamento della cybersicurezza nazionale e di reati informatici."
[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'articolo 1, comma 1 della legge 11 febbraio 2019, n. 12.
[Accessibilità]	Legge 9 gennaio 2004, n. 4 e s.m.i. recante "Disposizioni per favorire e semplificare l'accesso degli utenti e, in particolare, delle persone con disabilità agli strumenti informatici".

[1] La Direttiva (UE) 2024/2853 entrerà in vigore e dovrà essere recepita con legge nazionale entro il 9 dicembre 2026.

[Codice privacy]	Decreto legislativo 30 giugno 2003, n. 196 e s.m.i. recante "Codice in materia di protezione dei dati personali, recante disposizioni per l'adeguamento dell'ordinamento nazionale al regolamento (UE) n. 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, 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".
[CAD]	Decreto legislativo 7 marzo 2005, n. 82 e s.m.i. recante "Codice dell'amministrazione digitale".
[D.Lgs. 36/2006]	Decreto legislativo 24 gennaio 2006, n. 36 e s.m.i. 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".
[Codice dei contratti]	Decreto legislativo 31 marzo 2023, n. 36 e s.m.i. recante "Codice dei contratti pubblici in attuazione dell'articolo 1 della legge 21 giugno 2022, n. 78, recante delega al Governo in materia di contratti pubblici".
[D.Lgs. 51/2018]	Decreto legislativo 18 maggio 2018, n. 51 recante "Attuazione della direttiva (UE) 2016/680 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati".

[D.Lgs. 200/2021]	Decreto legislativo 8 novembre 2021, n. 200 recante "Attuazione della 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)".
[D.Lgs. 82/2022]	Decreto legislativo 27 maggio 2022, n. 82 recante "Attuazione della direttiva (UE) 2019/882 (EAA) del Parlamento europeo e del Consiglio, del 17 aprile 2019, sui requisiti di accessibilità dei prodotti e dei servizi".
[D.Lgs. 134/2024]	Decreto legislativo 4 settembre 2024, n. 134 recante "Attuazione della direttiva (UE) 2022/2557 (CER) 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".
[D.Lgs. 138/2024]	Decreto legislativo 4 settembre 2024, n. 138 recante "Recepimento della direttiva (UE) 2022/2555 (NIS 2), relativa a misure per un livello comune elevato di cibersicurezza nell'Unione, recante modifica del regolamento (UE) n. 910/2014 e della direttiva (UE) 2018/1972 e che abroga la direttiva (UE) 2016/1148".
[DPR 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 (PIAO).
[Piano Triennale]	DPCM 12 gennaio 2024, recante "Piano triennale per l'informatica nella pubblica amministrazione 2024-2026".

[PT agg. 2025] DPCM 3 dicembre 2024 recante "Aggiornamento 2025 del Piano triennale per l'informatica nella pubblica amministrazione 2024-2026".

[PT agg. 2026] DPCM 4 settembre 2025 recante "Aggiornamento 2026 del Piano triennale per l'informatica nella pubblica amministrazione 2024-2026".

[COE FCAI] Convenzione quadro del Consiglio d'Europa sull'intelligenza artificiale e i diritti umani, la democrazia e lo Stato di diritto.

2.4 Linee Guida di riferimento

Di seguito sono elencate le Linee Guida emesse da AgID ai sensi dell'articolo 71 del CAD e altra documentazione regolamentare, che sono richiamati, anche indirettamente, nel presente documento. Le Linee Guida AgID sono consultabili sul sito istituzionale, dove sono pubblicati anche i relativi aggiornamenti in conseguenza dell'evoluzione tecnologica o della necessità di adeguamento alla normativa di riferimento.

[LG_OPENDATA] Linee Guida Open Data.

[LG_RIUSO] Linee Guida su acquisizione e riuso di software per le Pubbliche Amministrazioni.

[LG_ACCESS] Linee Guida sull'accessibilità degli strumenti informatici.

[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".
[REG_CLOUD]	Regolamento unico per le infrastrutture e i servizi cloud per la PA n. 21007 adottato da ACN il 27 giugno 2024.
[AI_ACT_PROHIB]	Orientamenti della Commissione relativi alle pratiche di Intelligenza Artificiale vietate ai sensi del regolamento (UE) 2024/1689 (AI Act).
[AI_ACT_GPAI]	C(2025) 7719 Orientamenti della Commissione relativi all'ambito di applicazione degli obblighi dei fornitori di modelli di IA per finalità generali stabiliti dal regolamento (UE) 2024/1689 (AI Act).
[GPAI CoP]	Codice di buone pratiche dell'IA per finalità generali, 10 luglio 2025.
[AI_ACT_TRANSP]	Orientamenti sugli obblighi di trasparenza per i fornitori e gli operatori di determinati sistemi di IA.
[TRANSP_COP]	Codice di buone pratiche sulla trasparenza dei contenuti generati dall'IA.

[HUDERIA]	Consiglio d'Europa, Comitato dei Ministri, Metodologia per la valutazione dei rischi e degli impatti dei sistemi di Intelligenza Artificiale dal punto di vista dei diritti umani, della democrazia e dello Stato di diritto (HUDERIA – Human Rights, Democracy and Rule of Law Impact Assessment), approvata il 26 febbraio 2025.
[COBRA]	Consiglio d'Europa, Comitato dei Ministri, Modello HUDERIA per l'analisi dei rischi basata sul contesto (COBRA – Context-Based Risk Analysis): modello per l'analisi del rischio contestuale, approvato il 25 febbraio 2026.
[DigComp 3.0]	European Commission, Joint Research Centre (2025), DigComp 3.0 Quadro europeo delle competenze digitali, quinta edizione (traduzione in italiano).

2.5 Acronimi

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

[PA]	Pubblica Amministrazione: nel contesto delle presenti Linee guida, si intendono genericamente tutti i soggetti di cui all'articolo 2, comma 2, del CAD.
[ACN]	Agenzia per la Cybersicurezza Nazionale.
[AgID]	Agenzia per l'Italia Digitale.
[AI HLEG]	High-Level Expert Group on Artificial Intelligence, Gruppo indipendente di esperti ad alto livello sull'Intelligenza Artificiale istituito dalla Commissione Europea.

[AI RMF]	Artificial Intelligence Risk Management Framework, quadro per la gestione del rischio dell'IA del NIST.
[ALTAI]	Assessment List for Trustworthy Artificial Intelligence, lista di valutazione per un'IA affidabile.
[API]	Application Programming Interface, interfaccia di programmazione delle applicazioni.
[CAD]	Codice dell'amministrazione digitale.
[CEI]	Comitato Elettrotecnico Italiano.
[CEN]	Comitato europeo di normazione.
[CENELEC]	Comitato europeo di normazione elettrotecnica.
[CIO]	Chief Information Officer.
[DDoS]	Distributed Denial of Service, attacco distribuito di negazione del servizio.
[DPIA]	Valutazione d'impatto sulla protezione dei dati ex articolo 35 GDPR e articolo 27 Dir. 2016/680.
[EAA]	European Accessibility Act, direttiva (UE) 2019/882.
[EDPB]	European Data Protection Board, Comitato europeo per la protezione dei dati.
[ENISA]	Agenzia dell'Unione europea per la cibersicurezza.
[FRIA]	Valutazione d'impatto sui diritti fondamentali per i sistemi di IA ad alto rischio ex articolo 27 AI Act.

[GDPR]	General Data Protection Regulation, regolamento (UE) 2016/679.
[GPAI]	General-Purpose AI model, modello IA per finalità generali.
[IA]	Intelligenza Artificiale.
[IAM]	Identity and Access Management, gestione delle identità e degli accessi.
[ICT]	Information and Communication Technology, tecnologie dell'informazione e della comunicazione.
[IEC]	International Electrotechnical Commission.
[IEEE]	Institute of Electrical and Electronics Engineers.
[IoT]	Internet of Things, Internet delle cose.
[ISO]	International Organization for Standardization.
[JRC]	Joint Research Centre, Centro comune di ricerca della Commissione Europea.
[JSON]	JavaScript Object Notation.
[JTC]	Joint Technical Committee.
[KPI]	Key Performance Indicator, indicatore chiave di prestazione.
[MFA]	Multi-Factor Authentication, autenticazione a più fattori.
[ML]	Machine Learning, apprendimento automatico.
[MLOps]	Machine Learning Operations.
[NCSC]	National Cyber Security Centre (Regno Unito).

[NIST]	National Institute of Standards and Technology (Stati Uniti).
[NLF]	New Legislative Framework, quadro normativo dell'Unione europea per l'armonizzazione della legislazione sui prodotti.
[NLP]	Natural Language Processing, elaborazione del linguaggio naturale.
[OCSE]	Organizzazione per la cooperazione e lo sviluppo economico (OECD).
[PDCA]	Plan-Do-Check-Act.
[PIAO]	Piano integrato di attività e organizzazione di cui al DPR 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.
[RIA]	Responsabile dell'Intelligenza Artificiale.
[RPD]	Responsabile della protezione dei dati.
[RTD]	Responsabile per la Transizione Digitale.
[SDG]	Sustainable Development Goals.
[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.
[TRL]	Technology Readiness Level.
[TTP]	Tactics, Techniques and Procedures, tattiche, tecniche e procedure.
[UIA]	Ufficio del Responsabile dell'Intelligenza Artificiale.
[UNESCO]	United Nations Educational, Scientific and Cultural Organization.
[UNI]	Ente Italiano di Normazione.
[URI]	Uniform Resource Identifier.
[UTD]	Ufficio per la Transizione Digitale.
[XML]	eXtensible Markup Language.

2.6 Termini e definizioni

Per i termini e le definizioni utilizzati nel presente documento si rimanda alla normativa unionale e allo strumento "*Termini e definizioni*" delle Linee Guida AgID sull'Intelligenza Artificiale pubblicato sul sito istituzionale dell'Agenzia.

Lo strumento raccoglie i principali concetti tecnici, normativi e operativi relativi all'IA. I termini e le definizioni sono allineati, ove applicabile, alla normativa europea e agli standard internazionali di riferimento.

3. L'Intelligenza Artificiale

L'Intelligenza Artificiale (IA) è un insieme di tecnologie in grado di trasformare e potenziare attività economiche e sociali, al fine di migliorare 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*[2] *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"*[3].

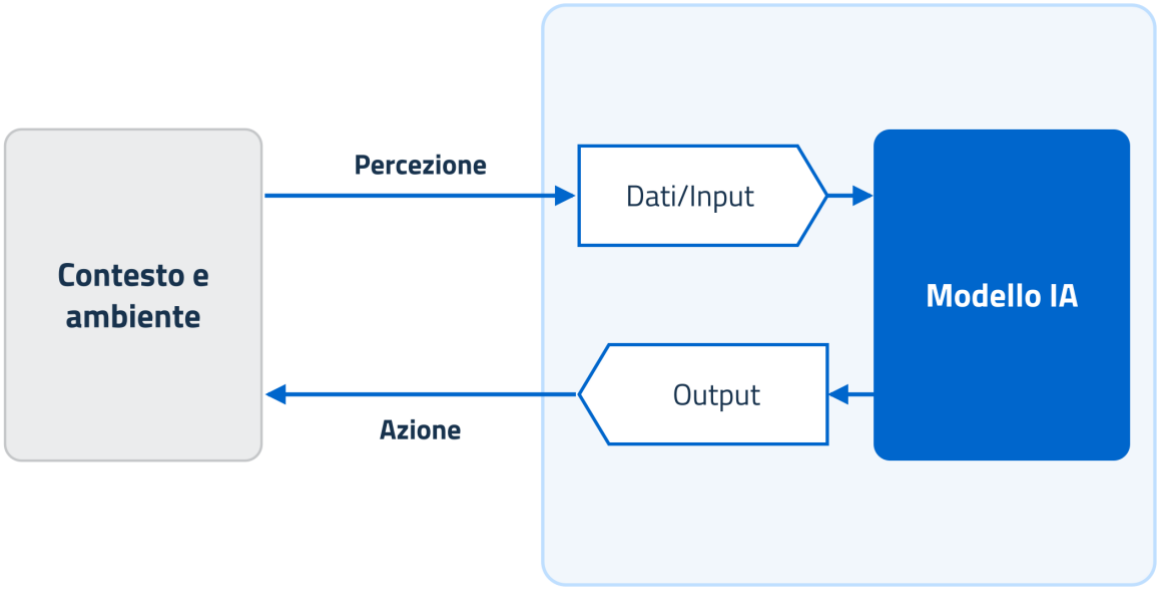


Figura 1 – Sistema di IA[4]

[2] Il testo dell'AI Act in italiano riporta "automatizzato" come traduzione del termine inglese "machine-based".

[3] AI Act articolo 3 definizione 1). Ai fini dell'AI Act si tengano in considerazione, inoltre, le "Guidelines on the definition of an artificial intelligence system established by AI Act".

[4] L'immagine è tratta da OECD AI Principles overview. La definizione di AI system dell'AI Act è stata allineata alla definizione OCSE nel corso del trilogo di negoziazione sull'AI Act.

La figura 1 illustra l'architettura funzionale di un sistema di IA. Una caratteristica distintiva dei più diffusi sistemi di IA generativa è la capacità inferenziale[5] del modello di IA[6] che ne costituisce il componente principale; tale componente, addestrato 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.

A differenza dei sistemi deterministici, i sistemi di IA generativa, in particolare quelli basati su modelli di apprendimento automatico, producono risultati che riflettono una logica statistico-probabilistica: le previsioni, le raccomandazioni o gli output prodotti derivano da inferenze apprese dal modello e sono pertanto soggetti a margini di incertezza.

Tale caratteristica richiede particolare attenzione nella valutazione dell'affidabilità e della robustezza dei risultati, soprattutto in contesti ad alto impatto, in cui gli effetti delle decisioni

[5] 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.

[6] 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).

automatizzate possono incidere sui diritti fondamentali, sull'equità sociale e/o sull'accesso a servizi pubblici essenziali.

I sistemi di IA possono operare con livelli di autonomia variabili, che spaziano da un controllo umano diretto a un funzionamento indipendente, come ampiamente descritto nel par. 3.2 delle Linee Guida per lo sviluppo di sistemi di Intelligenza Artificiale nella Pubblica Amministrazione (di seguito Linee Guida sviluppo).

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

Con "sistemi di IA ad alto rischio" si fa riferimento ai sistemi di IA che, per le finalità e i contesti di utilizzo previsti dall'AI Act, possono determinare rischi significativi per la salute, la sicurezza e i diritti fondamentali, nonché per la democrazia, lo Stato di diritto e la protezione dell'ambiente[7].

Con "modello IA per finalità generali" (GPAI) si intende: "*un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'auto-supervisione su larga scala, che sia caratterizzato da una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato*"[8].

L'articolo 51 dell'AI Act introduce una specifica categoria di modelli di IA per finalità generali (General-Purpose AI models - GPAI) "a rischio sistemico", caratterizzati da "capacità ad alto

[7] AI Act considerando 1 e articolo 1.

[8] AI Act, articolo 3, punto 63.

impatto" o da un impatto significativo sul mercato interno. Tali modelli sono soggetti a un regime regolatorio specifico previsto dall'AI Act.

A norma dell'articolo 56 dell'AI Act, la Commissione Europea ha definito un Codice di buone pratiche, da adottare su base volontaria da parte dei fornitori di modelli GPAI con rischio sistemico, in attesa dell'elaborazione di una richiesta formale di standardizzazione agli organismi europei di normazione per la definizione di standard specifici per tali modelli (cfr. articolo 40 dell'AI Act).

I sistemi di IA possono essere soggetti a *bias*[9], 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'articolo 10 dell'AI Act prevede, per i sistemi ad alto rischio, specifici requisiti di governance dei dati, finalizzati a prevenire bias e distorsioni. Per un approfondimento sulla governance dei dati, si rimanda al cap. 9 del presente documento.

[9] ISO/IEC TR 24027 «Artificial intelligence - Bias in AI systems and AI aided decision making».

3.1 Ciclo di vita di un sistema di IA

Le presenti Linee Guida adottano come modello di riferimento il ciclo di vita di un sistema di IA definito dall'OCSE, fermo restando la coerenza con il Regolamento AI Act e con la normativa in materia di protezione dei dati personali. Di seguito sono descritte le fasi che caratterizzano tale ciclo di vita, dalla pianificazione iniziale fino alla sua dismissione.

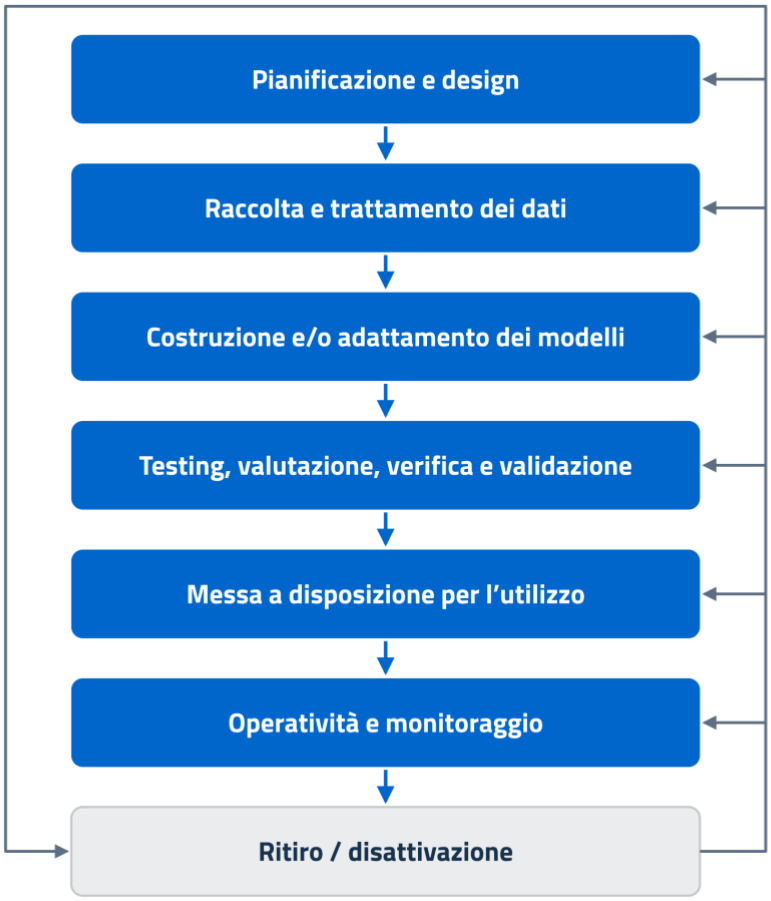


Figura 2 – Ciclo di vita di un sistema di IA[10]

- **Pianificazione e design:** l'organizzazione individua l'obiettivo che si prefigge di raggiungere mediante l'introduzione di un sistema basato su tecniche di IA, comprendendo il contesto nel quale il sistema dovrà operare (cfr. par. 4.2) e i dati che verranno utilizzati dal sistema che si sta progettando. La fase comprende la definizione dei

[10] OECD AI Principles overview.

fabbisogni e la raccolta dei requisiti espressi dagli stakeholder, la traduzione di questi ultimi in requisiti tecnici misurabili e il design del modello logico sottostante, in coerenza con il modello logico di riferimento introdotto dalle Linee Guida sviluppo (cfr. cap. 3).

- **Raccolta e trattamento dei dati:** i dati (espressi 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, integrati e trasformati secondo le esigenze, nel rispetto della normativa vigente in materia di protezione dei dati personali. Tali operazioni devono seguire le indicazioni incluse nel capitolo sulla gestione e la qualità dei dati (cfr. cap. 9). 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 individua uno o più modelli di IA adeguati agli obiettivi del sistema e ai dati disponibili. Scegliere il set di modelli implica anche la scelta degli algoritmi di apprendimento[11] con i quali potrebbe essere necessario effettuare il fine-tuning dei modelli stessi. In alternativa l'organizzazione può scegliere l'approccio del "trasferimento dell'apprendimento", utilizzando come punto di partenza un modello già addestrato su altri dati e adattandolo al proprio caso d'uso mediante ulteriore addestramento o *fine-tuning*. Questo approccio può essere scelto quando l'organizzazione non dispone di dati sufficienti per l'addestramento, di competenze interne o di capacità computazionali adeguate, ovvero può essere classificata come un operatore base, in coerenza con la classificazione introdotta nelle Linee Guida sviluppo al par. 3.5.

[11] Le tre categorie più comuni di algoritmi sono l'apprendimento supervisionato, non supervisionato e con rinforzo.

- **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 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**: 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 degli output. 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 soggetto fornitore *(provider)* e soggetto attuatore (*deployer)* identificate dall'AI Act[12] nel contesto generale dei sistemi di IA. Per una descrizione degli obblighi specifici applicabili a fornitore, attuatore e agli altri ruoli individuati dall'AI Act e la definizione di sistemi IA ad alto rischio, si faccia riferimento al cap. 5. In particolare, si possono considerare le seguenti definizioni introdotte dall'AI Act:

- *Provider*: "*una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che sviluppa un sistema di IA o un modello di IA per finalità generali o che fa sviluppare un sistema di IA o un modello di IA per finalità generali e immette tale sistema o modello sul mercato o mette in servizio il sistema di IA con il proprio nome o marchio, a titolo oneroso o gratuito*";
- *Deployer*: "*una persona fisica o giuridica, un'autorità pubblica, un'agenzia o un altro organismo che utilizza un sistema di IA sotto la propria autorità, tranne nel caso in cui il sistema di IA sia utilizzato nel corso di un'attività personale non professionale*".

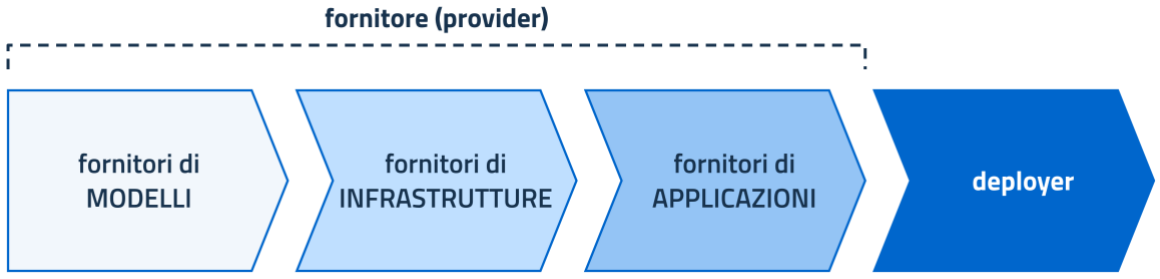


Figura 3 – Catena del valore dell'IA

[12] AI Act articolo 3 definizioni 3) e 4).

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 della pipeline dati, networking, hardware, servizi cloud, piattaforme MLOps.

- **Fornitori di applicazioni**

  Soggetti che sviluppano e/o forniscono applicazioni per scopi specifici adattando modelli di IA esistenti. 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.

- ***Deployer***

  Soggetti che utilizzano sistemi di IA per rispondere a esigenze operative o strategiche. I *deployer* possono integrare tali sistemi nei propri processi, personalizzarli o adattarli (ad esempio tramite fine-tuning su dati proprietari), con l'obiettivo di ottimizzare le prestazioni in funzione del contesto di applicazione. Ai fini dell'AI Act se un *deployer* modifica in modo significativo il sistema IA o lo utilizza sotto il proprio nome o marchio, assume le responsabilità del fornitore (cfr. articolo 25 AI Act).

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:

- **Sistemi di IA vietati**[13]*:* sono i sistemi dal rischio inaccettabile e, pertanto, vietati. L'articolo 5 dell'AI Act individua i sistemi IA per i quali l'AI Act vieta l'immissione sul mercato in quanto possono provocare danni significativi alle persone. Si citano ad esempio: sistemi 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; sistemi di IA che introducano meccanismi di *social scoring* provocando un trattamento pregiudiziale o sfavorevole di determinate classi di individui.
- **Sistemi di IA ad alto rischio:** sono i sistemi che possono determinare rischi significativi per la salute, la sicurezza e i diritti fondamentali, nonché per la democrazia, lo Stato di diritto e la protezione dell'ambiente. 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'articolo 6 e l'Allegato III) anche in relazione alle possibili deroghe (cfr. articolo 6 comma 3). 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.

[13] Gli obblighi relativi agli usi vietati dell'IA di cui all'articolo 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)".*

- **Sistemi di IA a rischio limitato**: sono sistemi che presentano rischi reputati minori rispetto ai precedenti, per i quali sono previsti gli obblighi di trasparenza dell'articolo 50 dell'AI Act. In caso di malfunzionamento, si contraddistinguono per una potenziale dannosità limitata, con un impatto limitato su interessi pubblici e diritti fondamentali.
- **Sistemi di IA a rischio minimo***:* sono sistemi dal rischio irrilevante. In caso di malfunzionamento, la relativa dannosità è molto limitata, potendo avere un impatto trascurabile su interessi pubblici e diritti fondamentali.

Per tutti i sistemi di IA, 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**, della **Convenzione europea dei diritti dell'uomo** e della **Convenzione quadro del Consiglio d'Europa sull'Intelligenza Artificiale** valutando l'impatto dei sistemi di IA sui diritti tutelati;
- tenere conto della **Dichiarazione europea sui diritti e i principi digitali per il decennio digitale**, assicurando l'allineamento delle applicazioni IA ai valori europei;
- ispirarsi agli **Orientamenti etici per un'IA affidabile** dell'AI HLEG, promuovendo trasparenza, equità e responsabilità nell'uso dell'IA.

Il Capo V dell'AI Act definisce, inoltre, 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 i sistemi di IA perseguendo, in via prioritaria ma non esclusiva, le seguenti finalità:

- 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.

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 (cfr. articolo 1 e 2 AI Act).

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, nel rispetto del considerando 69 e dell'articolo 2, par. 7, dell'AI Act (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, qualunque sia il livello di automazione previsto[14], 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, contribuendo al contempo alla diminuzione della disuguaglianza culturale. Le PA adottano misure preventive per evitare la riproduzione o l'amplificazione di *bias* presenti nella società.

P.7 **Trasparenza, spiegabilità e documentazione.** Le PA adottano sistemi di IA garantendo la trasparenza e comprensibilità delle loro decisioni e del funzionamento secondo il principio di proporzionalità tra trasparenza ed efficacia. Le PA si assicurano che i sistemi di IA siano sviluppati e implementati in modo da consentire un adeguato livello di comprensione del loro funzionamento. Le PA si assicurano che i sistemi di IA siano adeguatamente documentati durante il loro ciclo di vita. Gli articolo 11 e 13 dell'AI Act rafforzano tale principio per i sistemi ad alto rischio.

P.8 **Trasparenza e informazione.** Le PA informano gli utenti sull'interazione con sistemi di IA, rendendoli consapevoli delle capacità e dei limiti di tali sistemi. L'articolo 50 dell'AI Act rafforza tale principio per i sistemi destinati ad interagire direttamente con le persone fisiche (cfr. cap. 7).

Qualità e affidabilità dei sistemi di IA

P.9 **Qualità dei dati.** Le PA adottano sistemi di IA garantendo una gestione dei dati etica e trasparente. Questo include l'utilizzo di dati di elevata qualità, la definizione di politiche

[14] Si vedano, sul punto, le "Linee Guida per lo sviluppo di sistemi di Intelligenza Artificiale nella Pubblica Amministrazione".

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. L'articolo 10 dell'AI Act rafforza tale principio per i sistemi ad alto rischio (cfr. cap. 9).

P.10 **Accuratezza**. Le PA garantiscono che i risultati e le decisioni prodotti dai sistemi di Intelligenza Artificiale siano accurati e coerenti con gli obiettivi prefissati. A tal fine, devono essere adottate misure per il monitoraggio continuo delle prestazioni, al fine di individuare e correggere tempestivamente eventuali errori o deviazioni. L'articolo 15 dell'AI Act rafforza tale principio per i sistemi ad alto rischio.

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. L'articolo 15 dell'AI Act rafforza tale principio per i sistemi ad alto rischio.

P.12 **Sicurezza cibernetica.** Le PA garantiscono la cybersicurezza 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. L'articolo 15 dell'AI Act rafforza tale principio per i sistemi ad alto rischio.

P.13 **Sorveglianza umana.** Le PA garantiscono un livello adeguato di sorveglianza umana dei sistemi di IA. Ai fini della sorveglianza 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. L'articolo 14 dell'AI Act rafforza tale principio per i sistemi ad alto rischio.

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. L'articolo 12 dell'AI Act rafforza tale principio per i sistemi ad alto rischio.

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, coinvolgendo attivamente i portatori di interesse, per sperimentare e integrare soluzioni tecnologiche basate sull'IA, creando un ecosistema favorevole all'innovazione e all'adozione di tecnologie digitali. Le PA prendono in considerazione l'impiego di sistemi e modelli di IA open source, valutandone il potenziale in termini di trasparenza, riuso, interoperabilità, autonomia tecnologica e sostenibilità economica. Le PA promuovono forme di collaborazione e aggregazione, anche stabile o sotto forma di reti, tra amministrazioni, per la condivisione di competenze, infrastrutture, dati, soluzioni tecnologiche, capacità di sviluppo e di procurement.

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 e consapevole delle opportunità, dei rischi 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, anche in forma aggregata o condivisa, così come promosso anche nelle Linee Guida sviluppo e nelle Linee Guida per il procurement di IA nella Pubblica Amministrazione (di seguito Linee Guida procurement), al fine di sostenere il processo di trasformazione digitale, rafforzare il controllo sui sistemi e sui processi digitali, migliorare la resilienza e l'autonomia operativa dell'amministrazione, garantendo adeguati livelli di sicurezza. In caso di aggregazione e condivisione, le PA DEVONO porre la massima attenzione sulla protezione dei dati personali eventualmente coinvolti, nel rispetto dei principi di cui all'articolo 5 GDPR.

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 organizzativo di adozione dell'IA

Le PA DEVONO 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 DEVONO implementare un modello organizzativo 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. Tale norma costituisce un riferimento metodologico per l'impostazione di un sistema di gestione dell'IA e deve essere applicata in modo coerente con il quadro normativo vigente.

Nei casi previsti dall'AI Act, le PA DEVONO adottare il sistema di gestione della qualità richiesto dal Regolamento IA e tenere conto delle pertinenti norme tecniche armonizzate, ove disponibili (cfr. par. 4.4).

Le norme tecniche non armonizzate, inclusa la ISO/IEC 42001:2023, possono concorrere a dimostrare l'adozione di adeguate misure organizzative, gestionali e di controllo, ma non conferiscono la presunzione di conformità ai requisiti dell'AI Act.

La Figura 4 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 di *deployer* di sistemi di IA. In particolare, le PA POSSONO, per l'adozione di sistemi di IA di rischio limitato o minimo, accorpare le fasi del modello, adeguandole alla bassa complessità del progetto di adozione.

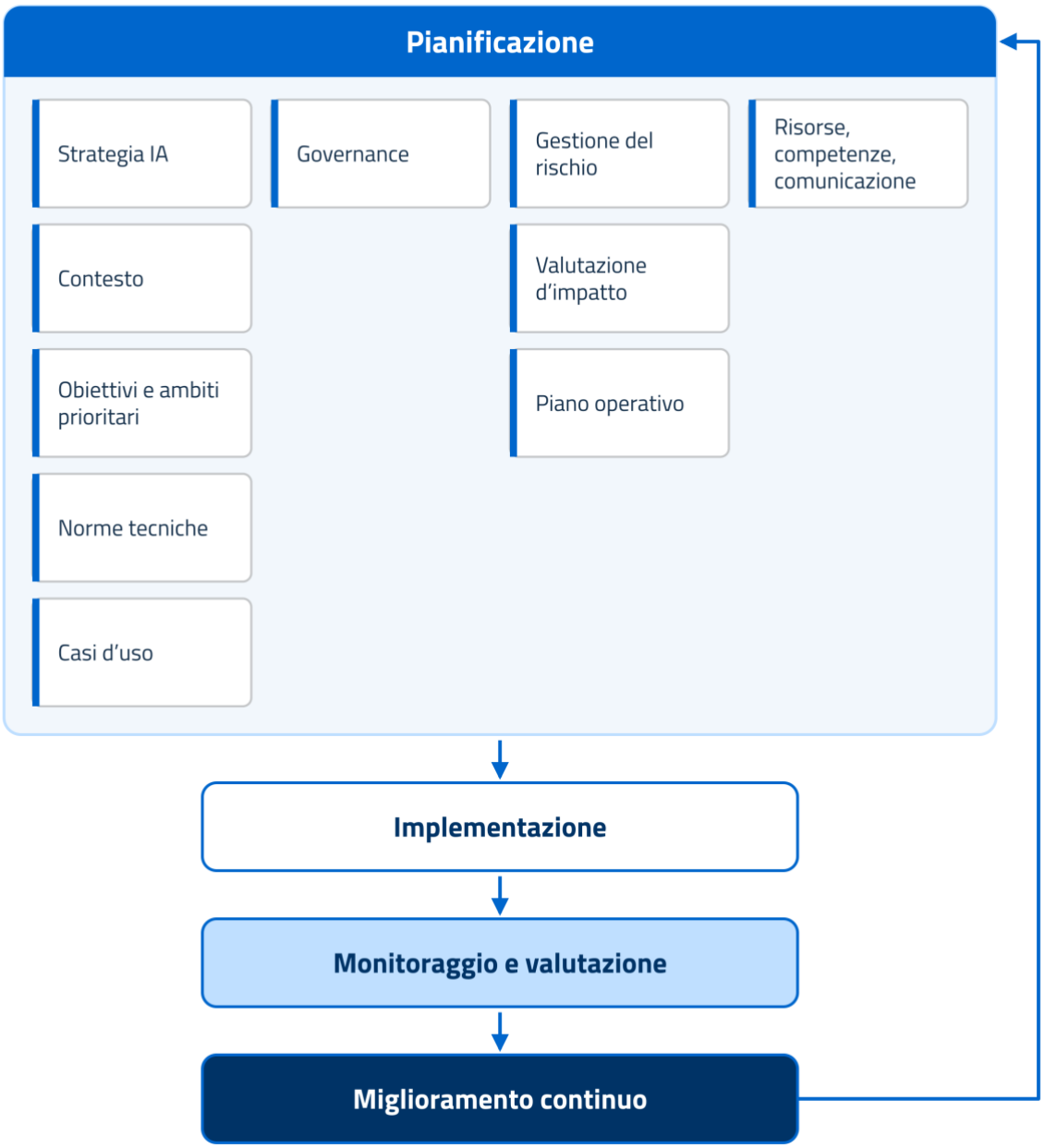


Figura 4 – Rappresentazione delle fasi del modello di adozione dell'IA.

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 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.

La strategia DEVE contenere le misure atte a garantire la trasparenza sulle modalità di impiego dei sistemi di IA da parte della PA.

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

- migliorare la qualità dei dati, ivi comprese le basi documentali, adottando una policy di data governance efficace (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. par. 4.5);
- valutare le esperienze e sperimentazioni già effettuate da altre PA;
- privilegiare l'impiego di sistemi e modelli IA *open source* o *open weight*, privilegiando un modello di riferimento basato su un orchestratore centralizzato in grado di garantire l'uso e l'integrazione di una molteplicità di modelli di IA, in funzione del caso d'uso e dello specifico fabbisogno della PA, come definito al cap. 3 delle Linee Guida sviluppo;
- avviare sperimentazioni e successiva adozione, anche in forma associata con altre PA, partendo dai casi d'uso di minore complessità (*quick win*) e adottando una politica di aggregazione soprattutto tra PA definite "*operatori base*", così come definiti nelle Linee Guida sviluppo. La PA DEVE definire la strategia per l'IA in coerenza con il PIAO, la strategia sulla data governance e il Piano Triennale per l'informatica nella Pubblica Amministrazione definito da AgID.

La PA DEVE implementare la propria strategia dell'IA affidandone il coordinamento, unitamente alla governance sui dati, al Responsabile per la Transizione Digitale (RTD) e al suo ufficio (UTD),

come previsto dal piano triennale per l'informatica nella PA, coinvolgendo il Responsabile della protezione dei dati (RPD) e il Referente per la cybersicurezza[15] o analoga figura.

Il Responsabile per la Transizione Digitale (RTD) collabora con il Responsabile della protezione dei dati (RPD) nel rispetto dei compiti attribuiti a quest'ultimo dagli articoli 38 e 39 del GDPR.

In contesti ad elevata intensità di utilizzo dell'Intelligenza Artificiale, la PA PUÒ istituire una figura dedicata al coordinamento dell'IA, il Responsabile dell'Intelligenza Artificiale (RIA), al quale PUÒ essere assegnato un proprio ufficio (UIA), che opera seguendo le direttive del RTD.

La PA DEVE monitorare e aggiornare periodicamente la strategia al fine di:

- verificarne l'effettiva applicazione e applicabilità;
- rilevare eventuali scostamenti rispetto agli obiettivi prefissati;
- valutare la necessità di aggiornamenti in relazione all'evoluzione degli obiettivi e delle tecnologie.

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, in coerenza al principio di aggregazione promosso nelle Linee Guida sviluppo e nelle Linee Guida procurement o adottando strategie e modelli definiti da enti sovraordinati o della stessa tipologia.

I fattori esterni includono i requisiti normativi come quelli, ad esempio, definiti da AI Act, DGA, GDPR, NIS 2, CAD e Legge 132/2025, nonché le Linee Guida e le decisioni adottate dalle autorità

[15] Referente per la cybersicurezza designato ai sensi dell'articolo 8, comma 2, della legge 28 giugno 2024, n. 90.

competenti, l'evoluzione della normativa tecnica e dell'offerta tecnologica, con particolare riferimento alla disponibilità di nuovi servizi e soluzioni di IA.

Con riferimento ai trattamenti che comportano rischi elevati ai sensi dell'AI Act, le PA DEVONO, durante l'intero ciclo di vita del sistema di IA, valutare le possibili ricadute derivanti dall'utilizzo dei sistemi di IA, i correlati rischi per i diritti e le libertà fondamentali delle persone e le misure idonee alla loro adeguata mitigazione.

I fattori interni comprendono i requisiti relativi a: struttura organizzativa (incluse le dimensioni dell'ente), stakeholder interni (es. personale dell'ente), procedimenti amministrativi, 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.

In tale contesto, si sottolinea l'importanza dell'attività di classificazione dei dati e servizi digitali delle PA, prevista dal Regolamento unico per le infrastrutture e i servizi cloud per la PA n. 21007 di ACN, connessa ai livelli minimi di sicurezza, affidabilità, capacità elaborativa, risparmio energetico delle infrastrutture digitali per la PA e alle caratteristiche di qualità, sicurezza, performance, scalabilità, interoperabilità e portabilità dei servizi cloud per la PA[16].

Dall'analisi del contesto e delle caratteristiche della PA derivano le azioni della strategia per l'IA[17] (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;

[16] 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"

[17] OECD Observatory of Public Sector Innovation.

- 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, sicuro e non discriminante, definendone le modalità di implementazione nel rispetto dei diritti fondamentali.

4.3 Obiettivi e ambiti prioritari di applicazione

Le PA sono incoraggiate a sperimentare l'introduzione di sistemi basati su tecniche di IA individuando ambiti e casi d'uso con un impatto proporzionale alla disponibilità di infrastrutture, tecnologie, competenze e disponibilità finanziarie. 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.

Tale analisi DEVE includere una valutazione dei costi, coerente con le indicazioni contenute nelle Linee Guida procurement, e dei benefici economici, considerando il potenziale risparmio di risorse e il miglioramento della qualità e dell'efficacia 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'articolo 68 del CAD[18].

[18] 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.

Le PA possono adottare soluzioni di IA in contesti ritenuti rilevanti per il miglioramento dei servizi, dell'efficienza operativa o della capacità decisionale, come ad esempio

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:
    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. **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.
    c. **Miglioramento della gestione della conoscenza**: le PA utilizzano l'IA per estrarre informazioni e concetti rilevanti da testi non strutturati, migliorando l'accessibilità, l'arricchimento e la fruibilità del patrimonio informativo.
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 possono impiegare 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 può consentire 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 possono utilizzare 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 possono adottare soluzioni di IA per rendere i servizi pubblici accessibili e conformi all'articolo 53 del CAD e alla normativa e alle Linee Guida vigenti in materia, 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 possono adottare 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[19] agli organismi europei di standardizzazione CEN e CENELEC di sviluppare norme tecniche armonizzate[2021] ai sensi del Regolamento UE 1025/2012, che definiscono approcci concreti per soddisfare tali requisiti. In base a quanto stabilito dall'AI Act, 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.

Gli standard tecnici elaborati a livello internazionale da organismi quali ISO e IEC possono costituire un utile riferimento metodologico e operativo per i diversi aspetti connessi alla progettazione, allo sviluppo, all'adozione e alla gestione dei sistemi di Intelligenza Artificiale.

Negli ambiti disciplinati dall'AI Act, le PA POSSONO applicare le pertinenti norme tecniche armonizzate ove disponibili ai fini della presunzione di conformità dei sistemi di IA ad alto rischio, come previsto dall'articolo 40 dell'AI Act.

Le norme tecniche e gli standard internazionali non armonizzati possono concorrere a dimostrare l'adozione di adeguate misure tecniche, organizzative e gestionali. La loro applicazione, tuttavia, non conferisce la presunzione di conformità prevista dall'articolo 40 dell'AI Act, riconosciuta, nei limiti dei requisiti e degli obblighi coperti, alle norme armonizzate o a parti di esse i cui riferimenti siano pubblicati nella Gazzetta ufficiale dell'Unione europea.

[19] 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.

[20] 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, sorveglianza umana, accuratezza, robustezza e cybersicurezza.

[21] Si veda l'articolo 113 dell'AI Act. 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".

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.

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[22]. I casi d'uso[23] POSSONO essere implementati in forma associata da più PA, anche attraverso gli spazi di sperimentazione e sviluppo e le sandbox regolamentari previste dall'AI Act.

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

4.5.1 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

[22] 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

[23] 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.

presente le aspettative delle parti interessate. Le PA DEVONO applicare, oltre ai *Principi* indicati nelle presenti Linee Guida (cfr. par. 3.4), le seguenti categorie di requisiti:

- conformità all'AI Act (cfr. cap. 5)
- gestione e qualità dei dati (cfr. cap. 9)
- conformità alla normativa in materia di protezione dei dati personali (cfr. cap. 10)
- sicurezza cibernetica (cfr. cap. 11).

Nell'identificazione dei requisiti, le PA DEVONO tenere in considerazione anche il livello di rischio del sistema di IA da implementare (cfr. par. 4.7): nel caso di sistemi ad alto rischio, si applicano i requisiti obbligatori previsti dall'AI Act.

4.5.2 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 lungo l'intero ciclo di vita del caso d'uso, al fine di valutare l'evoluzione delle prestazioni del sistema di IA.

I KPI POSSONO misurare sia prestazioni tecniche sia l'impatto sui processi organizzativi 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;
- conformità normativa;
- costo e sostenibilità.

4.6 Governance

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), secondo le modalità individuate nella strategia (cfr. par. 4.1), se del caso assicurando uno stretto coordinamento con il RPD.

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

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

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

Fermi restando gli obblighi di conformità all'AI Act, le PA POSSONO adottare un codice etico e di condotta per l'IA. Tale codice DOVREBBE 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 di salute, sicurezza, Stato di diritto e diritti fondamentali degli individui.

Le PA DEVONO adottare un sistema di gestione del rischio adeguato alle caratteristiche, alle finalità e al contesto di utilizzo dei sistemi di Intelligenza Artificiale. Nei casi disciplinati dall'AI Act, tale sistema DEVE essere conforme ai requisiti previsti dal Regolamento stesso.

Al fine della definizione e dell'attuazione del sistema di gestione del rischio, le PA POSSONO fare riferimento alle norme tecniche UNI ISO 31000[24] e alla ISO/IEC 23894[25]. Nei casi previsti dall'AI Act, le PA POSSONO applicare le pertinenti norme tecniche armonizzate europee ove disponibili.

Resta fermo che le norme tecniche non armonizzate possono concorrere a dimostrare l'adozione di adeguate misure di gestione del rischio, ma non conferiscono la presunzione di conformità ai requisiti dell'AI Act[26].

Le PA DEVONO effettuare la gestione dei rischi specifici derivanti dall'utilizzo di sistemi di IA, tenendo conto anche dei rischi per i diritti e le libertà fondamentali degli interessati derivanti dai connessi trattamenti di dati personali, nel rispetto dei principi di minimizzazione dei dati, integrità e riservatezza, protezione dei dati *by design* e *by default* e degli obblighi di sicurezza previsti dalla normativa in materia di protezione dei dati personali e dall'AI Act.

Occorre considerare che i sistemi di IA possono integrare modelli IA per finalità generali (GPAI). In tali casi, a valle dell'analisi del rischio, la classificazione iniziale del sistema potrebbe variare, rendendo necessario riclassificare come "ad alto rischio" anche sistemi inizialmente valutati

[24] UNI ISO 31000:2018 Gestione del rischio - Linee Guida

[25] ISO/IEC 23894:2023 Information technology - Artificial intelligence - Guidance on risk management

[26] Le norme in Europa

come a rischio inferiore. In presenza di tale riclassificazione, si applicano integralmente le prescrizioni specifiche previste dall'AI Act per i sistemi ad alto rischio.

4.8 Valutazione d'impatto

L'AI Act definisce requisiti rigorosi per la valutazione d'impatto dei sistemi IA "ad alto rischio" (FRIA, cfr. cap. 5) che, qualora sia già rispettato uno qualsiasi degli obblighi di cui all'articolo 27 dell'AI Act mediante la valutazione d'impatto sulla protezione dei dati personali (DPIA), integra tale valutazione.

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

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 minimo". Resta fermo lo svolgimento della valutazione d'impatto sulla protezione dei dati personali ai sensi dell'articolo 35 GDPR e dell'articolo 27 Dir. (UE) 2016/680 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 procurement), lo sviluppo (cfr. Linee Guida sviluppo) 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, in coerenza con la descrizione dello stack tecnologico dell'IA introdotto nelle Linee Guida sviluppo costituito da cinque strati: energia, chip, infrastrutture, modelli IA e pipeline dati, strato applicativo. 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 condotta.

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. Nella documentazione operativa, utile a garantire l'efficacia e il miglioramento continuo del sistema, rientrano, a titolo esemplificativo: manuali, procedure, valutazioni di impatto, report di monitoraggio, registro dei reclami.

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;
- a valutare l'impatto sui diritti fondamentali;
- a valutare la necessità di effettuare una valutazione d'impatto sulla protezione dei dati personali ove rilevante in considerazione dello specifico sistema di IA di cui viene pianificato l'impiego (art. 35 e 36 GDPR);
- a 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 e alla L. 132/2025, nonché, in quanto applicabile, alla normativa unionale e nazionale in materia di protezione dei dati personali.

Le PA DEVONO fare riferimento alla definizione di "sistema di IA" contenuta nell'articolo 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.

Le PA DEVONO procedere alla classificazione del sistema di IA utilizzato o sviluppato stabilendone il livello di rischio, coerentemente con quanto previsto dall'articolo 6 dell'AI Act (cfr. par. 3.3).

Le PA DEVONO riconoscere il ruolo che ricoprono secondo l'AI Act, ad esempio: fornitore o *deployer* (cfr. par. 3.2), nonché, in quanto applicabile, di titolare o responsabile del trattamento dei dati personali ai sensi del GDPR.

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 Responsabilità lungo la catena del valore dell'IA

Le responsabilità e gli obblighi lungo la catena del valore dell'IA devono essere coerenti con l'AI Act, che prevede obblighi ben definiti a seconda del ruolo all'interno della catena, come definito negli articoli 22-23-24-25-26 dell'AI Act. In dettaglio, come espresso nell'articolo 25:

Qualsiasi *deployer* o soggetto terzo è considerato quale fornitore di un sistema di IA ad alto rischio ai fini dell'AI Act ed è quindi soggetto agli obblighi del fornitore (articolo 16 AI Act), nelle seguenti circostanze:

- se appone il proprio nome o marchio su un sistema di IA ad alto rischio già immesso sul mercato o messo in servizio, fatti salvi accordi contrattuali che prevedano una diversa ripartizione degli obblighi al riguardo;
- se apporta una modifica sostanziale a un sistema di IA ad alto rischio già immesso sul mercato o già messo in servizio in modo tale che resti un sistema di IA ad alto rischio a norma dell'articolo 6;
- se modifica la finalità prevista di un sistema di IA, anche un sistema IA per finalità generali, che non sia stato classificato come ad alto rischio e che è già stato immesso sul mercato o messo in servizio in modo tale che il sistema di IA interessato diventi un sistema di IA ad alto rischio a norma dell'articolo 6 dell'AI Act.

5.2 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 dei cittadini e degli utenti in generale 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 e dal ruolo svolto dalla PA (fornitore o *deployer*).

Fermi restando gli obblighi definiti dall'AI Act per i sistemi ad "alto rischio", le PA, anche per i sistemi di IA classificati come a "rischio limitato" o "minimo", DEVONO:

- verificare che i dati siano trattati in conformità alle normative vigenti, con particolare attenzione alla protezione dei dati personali, ivi compresi i profili relativi al potenziale trasferimento dei dati personali verso Paesi terzi (art. 44 ss. GDPR), 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.

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 e i loro impatti sui servizi istituzionali della PA e sui diritti fondamentali dei cittadini e degli utenti cui è rivolto il sistema di IA;
- **verifiche e test operativi:** sperimentazione dei sistemi IA in condizioni reali, per valutarne prestazioni, affidabilità ed eventuali criticità;
- **acquisizione e analisi dei reclami:** raccolta strutturata delle segnalazioni provenienti da utenti, cittadini o operatori della PA, con analisi qualitativa e quantitativa degli elementi di criticità evidenziati;

- **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'articolo 27 dell'AI Act, una valutazione dell'impatto sui diritti fondamentali (FRIA), eventualmente preceduta (quando rilevante) dalla valutazione d'impatto sulla protezione dei dati personali (DPIA) - che può essere integrata dalla FRIA ai sensi dell'articolo 27, par. 4, dell'AI Act -, 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** successivamente 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 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** la definizione, l'attuazione e il rispetto del sistema di gestione dei rischi, finalizzato a individuare, valutare e mitigare i rischi pertinenti connessi all'utilizzo dei sistemi di IA per la salute, la sicurezza e i diritti fondamentali. Il sistema di gestione dei rischi DEVE essere periodicamente riesaminato e aggiornato al fine di garantirne la continua efficacia. Le decisioni e le azioni significative adottate nell'ambito di tale processo DEVONO essere adeguatamente motivate e documentate;
- **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'articolo 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 i log** e disporre di documentazione tecnica contenente le informazioni necessarie per valutare la conformità del sistema ai requisiti dei sistemi di IA ad alto rischio.

5.3 Misure di sorveglianza

Le PA DEVONO adempiere gli obblighi riguardanti le misure di sorveglianza umana dei sistemi IA ad alto rischio previsti all'articolo 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 sorveglianza umana, integrando interfacce adeguate e strumenti di controllo;
- la PA *deployer* DEVE adottare procedure e strumenti per assicurare il monitoraggio, l'interpretazione e, dove necessario, l'intervento umano sulle decisioni generate dall'IA.

Le PA *deployer* affidano la sorveglianza umana dei sistemi di IA ad alto rischio al RTD e al suo ufficio (UTD), in stretta collaborazione con il Responsabile della protezione dei dati (RPD) e il Referente per la cybersicurezza 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.

5.4 Conservazione della documentazione

Le Pubbliche Amministrazioni DEVONO assicurare la conservazione della documentazione relativa ai sistemi di IA ad alto rischio, in conformità alla normativa nazionale ed europea applicabile. La documentazione conservata deve garantire la tracciabilità dei processi automatizzati e delle attività svolte mediante il sistema di IA, attraverso la conservazione dei dati, dei metadati e delle informazioni necessarie alle attività di verifica, controllo e audit, nel rispetto della normativa in materia di protezione dei dati personali.

L'AI Act (articolo 18) 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 a ogni altra normativa unionale e nazionale - anche settoriale - applicabile, le PA POSSONO definire e adottare principi, codici etici o codici di condotta per l'utilizzo responsabile dei sistemi di IA, anche integrandoli nei codici di comportamento e negli altri strumenti organizzativi già adottati dall'amministrazione. Tali codici possono essere sviluppati anche in collaborazione con altre amministrazioni o adottando modelli definiti da enti sovraordinati o della stessa tipologia.

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'articolo 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.

I codici etici e di condotta, in quanto strumenti finalizzati a promuovere una cultura condivisa di responsabilità nell'uso dell'IA, tendono 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.

Le PA POSSONO sviluppare codici etici e di condotta, ai sensi dell'articolo 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, inclusi soggetti privati, organizzazioni rappresentative della società civile, università, centri di ricerca, associazioni di fornitori, autorità competenti.

Nell'individuazione di eventuali requisiti o misure aggiuntive, le PA DEVONO considerare:

- le soluzioni tecniche disponibili e le buone pratiche maturate nei settori di riferimento;
- i principi etici elaborati o in corso di elaborazione a livello nazionale, europeo e internazionale secondo quanto previsto dal medesimo articolo 95 dell'AI Act.

Le PA DEVONO considerare, ai fini dell'elaborazione, dell'aggiornamento e del consolidamento dei propri codici etici e di condotta in materia di Intelligenza Artificiale, le pertinenti fonti normative e gli atti di indirizzo adottati a livello europeo e internazionale, tra cui:

- **Consiglio d'Europa**, *Convenzione quadro sull'Intelligenza Artificiale e i diritti umani, la democrazia e lo Stato di diritto*, nonché la metodologia HUDERIA per la valutazione dei rischi e degli impatti dei sistemi di IA e il relativo modello per l'analisi dei rischi basata sul contesto (*Context-Based Risk Analysis – COBRA*);
- **Gruppo indipendente di esperti ad alto livello sull'Intelligenza Artificiale istituito dalla Commissione Europea (AI HLEG)**, *Orientamenti etici per un'IA affidabile* (*Ethics Guidelines for Trustworthy AI*) e *Lista di valutazione per un'Intelligenza Artificiale affidabile* (*Assessment List for Trustworthy Artificial Intelligence – ALTAI*).

Le PA POSSONO inoltre tenere conto, ove compatibili con tale quadro, dei principi, delle raccomandazioni, degli standard, delle metodologie, delle migliori pratiche e dei riferimenti etici e culturali elaborati dalle organizzazioni internazionali, dalle istituzioni e dagli organismi tecnici competenti.

Tra i riferimenti rilevanti possono essere richiamati, a titolo esemplificativo e non esaustivo:

- **Organizzazione delle Nazioni Unite per l'educazione, la scienza e la cultura (UNESCO)**, *Raccomandazione sull'etica dell'Intelligenza Artificiale* e relativi strumenti operativi, tra cui la metodologia per la valutazione (*Readiness Assessment Methodology*) e gli strumenti per la valutazione dell'impatto etico.

- **Organizzazione per la cooperazione e lo sviluppo economico (OCSE - OECD)**, *Raccomandazione del Consiglio sull'Intelligenza Artificiale* e relativi *Principi dell'OCSE sull'Intelligenza Artificiale*; *Governing with Artificial Intelligence*; *OECD Due Diligence Guidance for Responsible AI*.
- **Nazioni Unite**, rapporto *Governing AI for Humanity*, elaborato dall'Organo consultivo di alto livello sull'Intelligenza Artificiale del Segretario generale, nonché le iniziative di cooperazione internazionale rappresentate dal *Dialogo globale sulla governance dell'Intelligenza Artificiale* e dal *Gruppo scientifico internazionale indipendente sull'Intelligenza Artificiale*.
- **G7 Hiroshima AI Process**, con particolare riferimento agli *International Guiding Principles for Organizations Developing Advanced AI Systems* e all'*International Code of Conduct for Organizations Developing Advanced AI Systems*.
- **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems**, *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*.
- **IEEE Standards Association**, standard e pratiche raccomandate della famiglia IEEE 7000 per l'integrazione delle considerazioni etiche nella progettazione, nello sviluppo e nella valutazione dei sistemi autonomi e intelligenti, tra cui IEEE 7000 e IEEE 7010.
- **Papa Leone XIV**, Lettera enciclica *Magnifica Humanitas. Sulla custodia della persona umana nel tempo dell'Intelligenza Artificiale*.
- **Pontificia Accademia per la Vita**, con il supporto della Fondazione RenAIssance, *Rome Call for AI Ethics*.
- **Pontificia Commissione per lo Stato della Città del Vaticano**, decreto DCCII recante *Linee Guida in materia di Intelligenza Artificiale*.

Tali riferimenti hanno natura, valore giuridico e ambito di applicazione differenti e devono pertanto essere utilizzati in funzione integrativa e di supporto, senza pregiudicare l'applicazione delle disposizioni vincolanti dell'Unione europea e nazionali. Le PA ne valutano la pertinenza rispetto alle proprie finalità istituzionali, al contesto di utilizzo dei sistemi di IA e ai rischi per i diritti fondamentali, tenendo conto del loro grado di aggiornamento.

Nel redigere un codice etico, le PA DEVONO perseguire una serie di finalità ulteriori rispetto a quelle già previste dall'AI Act. In particolare, tali codici dovrebbero orientare la condotta dei soggetti coinvolti nello sviluppo, nel procurement e nell'adozione dei sistemi di IA, individuando principi e valori di riferimento volti a promuovere il benessere sociale e ambientale attraverso l'impiego dell'IA, ponendo l'accento su obiettivi etici quali:

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

L'adozione e le revisioni dei codici etici DOVREBBERO essere accompagnate da adeguate forme di consultazione pubblica e di coinvolgimento degli stakeholder rilevanti, incluse le Università e i relativi network di competenza, esperti, rappresentanti del settore pubblico e privato, nonché associazioni impegnate nella tutela dei diritti dei cittadini.

Le PA DEVONO garantire la revisione periodica dei codici etici adottati in materia di IA, al fine di assicurarne l'aggiornamento rispetto all'evoluzione tecnologica, normativa e organizzativa.

Nell'ambito dell'adozione e della revisione dei codici etici, le PA DEVONO altresì tenere conto, ove pertinenti, degli obiettivi di sviluppo sostenibile delle Nazioni Unite (*Sustainable Development Goals – SDGs*) e degli indirizzi dell'Unione europea in materia di sostenibilità e transizione verde.

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.

L'articolo 50 dell'AI Act prevede che i sistemi ad alto rischio e determinati sistemi di IA 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;
- **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 anche adeguata, esaustiva e accessibile informativa agli interessati sull'utilizzo di sistemi di IA nell'ambito del processo amministrativo (incluso l'utilizzo dell'output del sistema IA nel processo decisionale), sulle caratteristiche generali del trattamento effettuato attraverso tali sistemi, sulle 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 anche attraverso l'uso di un linguaggio semplice e comprensibile e, ove valutato opportuno dalla PA alla luce delle caratteristiche del trattamento, anche attraverso la messa a disposizione degli interessati di un estratto della DPIA.

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 - costituiti da tutti i soggetti a vario titolo coinvolti, comprese le persone giuridiche eventualmente destinatarie di decisioni derivanti dall'utilizzo di IA - 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 art. 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:** indicare gli elementi architetturali descritti nelle Linee Guida sviluppo;
  - **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 su cui si basano le decisioni, a vari livelli di automazione, prese dal sistema di IA, nel rispetto della normativa vigente anche in materia di protezione dei dati personali e, in particolare, dell'articolo 22 del GDPR. È essenziale che l'informativa comprenda:
  - **modalità di funzionamento del sistema:** indicando se il sistema 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:** le Pubbliche Amministrazioni POSSONO adottare decisioni basate su trattamenti automatizzati, compresa la profilazione, che producano effetti giuridici o incidano in modo analogo significativamente sull'interessato, esclusivamente nei casi e alle condizioni previsti dalla normativa vigente e, in particolare, in caso di coinvolgimento di dati personali, dall'articolo 22 del GDPR. In tali casi DEVONO essere previste misure adeguate a tutela dei diritti, delle libertà e dei legittimi interessi dell'interessato, inclusi, ove applicabili, il diritto di ottenere l'intervento umano, di esprimere la propria opinione e di contestare la decisione. Le PA DEVONO pertanto:
  - **garantire la trasparenza del processo decisionale:** fornire informazioni chiare ed esaustive sulla logica utilizzata e sui principali elementi della decisione automatizzata che riguarda l'interessato;
  - **descrivere i criteri e i dati utilizzati:** indicare le principali categorie di dati considerate, i criteri determinanti e il modo in cui tali elementi hanno inciso sull'esito della decisione, senza pregiudicare i diritti e le libertà altrui, la sicurezza dei sistemi e la tutela dei segreti commerciali.
- **Diritto di opposizione e intervento umano nel processo decisionale:** è fondamentale che le PA informino i propri utenti della facoltà di opporsi a decisioni automatizzate e di richiedere l'intervento umano nel processo decisionale, qualora questo possa avere effetti significativi sui loro diritti, ai sensi dell'AI Act e del GDPR. 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, prediligendo l'uso di un linguaggio chiaro, semplice e privo di tecnicismi.

8. Formazione e sviluppo delle competenze

La PA, al fine di cogliere in modo consapevole i vantaggi dell'IA, DEVE sviluppare un contesto gestionale, organizzativo, culturale e tecnologico adeguato, nel quale riveste un ruolo chiave l'acquisizione e lo sviluppo delle competenze 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, formando il proprio personale per poter gestire i diversi livelli di automazione.

La PA DEVE sviluppare competenze specifiche per governare e disciplinare l'utilizzo dell'IA all'interno dell'organizzazione, garantendo il rispetto della cornice normativa in materia di IA nonché dei 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.

Secondo quanto previsto dalla Strategia per l'Intelligenza Artificiale 2024-2026, la capacità di offrire percorsi di formazione continua e *up-skilling* sull'IA può rendere il settore pubblico all'avanguardia nell'innovazione tecnologica e, al contempo, più attrattivo per i talenti, contribuendo a trattenere le migliori risorse e a rafforzare la competitività del sistema amministrativo nel suo complesso.

L'articolo 4 dell'AI Act prevede che i fornitori e i *deployer* di sistemi di IA adottino misure volte a sostenere lo sviluppo dell'alfabetizzazione in materia di IA del personale e delle altre persone che si occupano del funzionamento e dell'utilizzo dei sistemi per loro conto, tenendo conto delle conoscenze tecniche, dell'esperienza, dell'istruzione e della formazione degli interessati, nonché del contesto di utilizzo e delle persone o dei gruppi di persone sui quali i sistemi di IA sono utilizzati.

Ne deriva che le PA DEVONO definire misure formative e organizzative differenziate e proporzionate in relazione al ruolo svolto, alle competenze possedute, alle caratteristiche dei sistemi utilizzati e ai relativi contesti di impiego.

Come richiamato anche nell'AI Toolkit predisposto dall'OCSE[27], il raggiungimento di un adeguato livello di maturità nell'adozione dell'IA presuppone la disponibilità di competenze diversificate ai vari livelli dell'organizzazione, dalle competenze di base a quelle specialistiche, nonché la definizione di interventi di *upskilling* e formazione differenziati e mirati.

In linea generale, è pertanto necessario mappare in maniera sistematica e completa le figure professionali presenti nelle PA, con l'obiettivo di identificare chiaramente il livello di fabbisogno formativo espresso e le responsabilità legate alla governance dell'IA. Questo processo permetterà di definire i piani formativi in maniera mirata, garantendo che il personale disponga delle competenze necessarie per i sistemi IA. La mappatura dovrebbe restituire almeno l'evidenza di un livello di fabbisogno di partenza, che include le competenze digitali di base, compresa la cosiddetta *AI literacy* (alfabetizzazione sull'IA), distinto da un livello di fabbisogno più avanzato o specialistico.

Per la definizione e lo sviluppo delle competenze digitali di base, comprese quelle connesse all'IA, le PA POSSONO fare riferimento al DigComp 3.0 – *European Digital Competence Framework*, che integra trasversalmente le competenze relative all'Intelligenza Artificiale, inclusa l'IA generativa, nel quadro complessivo delle competenze digitali. Tale riferimento è coerente anche con quanto previsto dall'articolo 4, paragrafo 3, dell'AI Act, secondo cui le raccomandazioni del Consiglio per l'IA in materia di alfabetizzazione devono tenere conto dei quadri europei delle competenze.

[27] OECD/UNESCO (2024), «G7 Toolkit for Artificial Intelligence in the Public Sector», OECD Publishing, Paris.

Il concetto di *AI literacy* racchiude un insieme di conoscenze e abilità atte a garantire la capacità di interagire in maniera efficace, consapevole e responsabile con le applicazioni di IA, presenti in maniera sempre più pervasiva in molti aspetti della vita sociale e del lavoro (anche pubblico).

L'*AI literacy*, pertanto, implica non solo la comprensione di base dell'esistenza dell'IA e della sua capacità di influenzare i processi di lavoro (*AI awareness*) ma anche la consapevolezza di *come* e *perché* si esplica tale influenza, includendo la conoscenza della varietà di implicazioni (operative, normative, etiche) associate all'IA.

In questa prospettiva l'*AI literacy* declinata nella Pubblica Amministrazione si esplica, in primo luogo, nella capacità dei dipendenti pubblici 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.

A questa si aggiunge la consapevolezza delle implicazioni nella trasposizione nel contesto lavorativo pubblico delle modalità di interazione con applicazioni di IA ricorrenti nella vita privata (es. motori di ricerca basati su IA, applicazioni di IA generativa basate su licenze individuali acquisite sia a titolo oneroso che gratuitamente, applicazioni *embedded* nei social media, ecc.).

Ulteriori elementi chiave dell'*AI literacy* sono, tra gli altri, la padronanza dei principi etici legati all'uso dell'IA, inclusa la trasparenza nell'ambito di decisioni prese in maniera automatizzata e la capacità di 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.

Le PA DEVONO garantire che tutto il personale acquisisca un adeguato livello di alfabetizzazione in materia di IA, quale componente ormai imprescindibile delle competenze digitali di base, al fine di consentire un utilizzo consapevole, corretto e responsabile delle opportunità offerte dai sistemi di IA.

Per quanto riguarda le competenze specialistiche, intendendo come tali quelle necessarie alle attività che influenzano le prestazioni dell'IA all'interno di una organizzazione, lo standard ISO/IEC 42001 (cfr. cap. 4) sul sistema di gestione dell'IA stabilisce requisiti specifici relativi alle competenze del personale coinvolto in tali attività, 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. 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.

Le PA e, in particolare, gli Uffici per la trasformazione digitale (UTD) DEVONO rafforzare e/o sviluppare le competenze specialistiche connesse all'IA attraverso il ricorso a politiche mirate di *upskilling* volte anche 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 gestione dell'IA richiede un insieme multidisciplinare di conoscenze e competenze, che comprende, tra gli altri, ambiti quali l'informatica, l'ingegneria, la matematica, il diritto delle tecnologie, la fisica, la comunicazione pubblica, la sociologia, l'antropologia e la filosofia, con particolare riferimento alla logica e all'etica.

Le PA DEVONO identificare e rafforzare le competenze tecniche e di dominio necessarie in relazione agli specifici processi interessati dall'adozione dell'IA, quali, a titolo esemplificativo, la formazione, il reclutamento e il procurement, nonché agli specifici ambiti applicativi, quali, tra gli altri, la sanità e i trasporti.

Lo standard ISO/IEC 42001 raccomanda di considerare la necessità di competenze diversificate in funzione delle diverse fasi del ciclo di vita dell'IA, così come per garantire l'adeguata rappresentatività con riferimento ai dati usati per l'addestramento dei modelli di *machine learning*.

L'adozione dell'IA non determina necessariamente la creazione di nuove strutture organizzative ma richiede innanzitutto l'evoluzione e il rafforzamento delle competenze di profili professionali già presenti nella Pubblica Amministrazione.

Le PA DEVONO promuovere e assicurare, per il personale che già ricopre ruoli nell'ambito delle tecnologie dell'informazione, lo sviluppo e l'acquisizione di specifiche competenze in materia di IA attraverso percorsi mirati di *upskilling*, specializzazione e aggiornamento professionale continuo.

In funzione delle caratteristiche dell'amministrazione, dei sistemi adottati e del relativo livello di complessità, alcuni profili possono progressivamente specializzarsi fino ad assumere funzioni prevalentemente o esclusivamente dedicate all'IA. A titolo esemplificativo e non esaustivo, tra i

profili professionali delle tecnologie dell'informazione[28] interessati da tale evoluzione possono rientrare:

- **Responsabile dei sistemi informativi**, con funzioni assimilabili a quelle del ***Chief Information Officer (CIO),*** ruolo ricoperto nella PA dal Responsabile per la Transizione Digitale. È responsabile dell'indirizzo e della governance complessiva dei sistemi informativi e dell'allineamento delle scelte tecnologiche agli obiettivi dell'organizzazione; con l'adozione dell'IA le relative competenze si estendono alla strategia, alla governance e alla gestione dei rischi connessi ai sistemi di IA. Nella PA questo ruolo assume anche quello di Responsabile dell'IA (cfr. par. 4.1).
- **Enterprise Architect**, responsabile della definizione e dell'evoluzione dell'architettura complessiva dei sistemi informativi e del relativo allineamento con processi, dati, applicazioni e infrastrutture; l'adozione dell'IA richiede di integrare modelli, servizi e componenti di IA nell'architettura dell'amministrazione.
- **Systems Architect**, responsabile della progettazione dell'architettura tecnica dei sistemi e dell'integrazione delle relative componenti; con l'IA deve considerare anche requisiti di scalabilità, prestazioni, sicurezza, interoperabilità e integrazione dei modelli e delle piattaforme di IA.
- **Project Manager**, responsabile del coordinamento e della conduzione dei progetti ICT in termini di obiettivi, attività, tempi, risorse, rischi, qualità e rapporti con i fornitori; nei progetti di IA deve integrare le specificità del relativo ciclo di vita, i requisiti di governance e conformità e i rischi propri dei sistemi di IA.

[28] Per l'individuazione e la descrizione dei profili professionali operanti nell'ambito delle tecnologie dell'informazione, possono costituire utili riferimenti la UNI 11621-2:2021, Attività professionali non regolamentate – Profili di ruolo professionale per l'ICT – Parte 2: Profili europei di ruolo professionale per l'ICT di seconda generazione, e la UNI 11621-4:2024, Attività professionali non regolamentate – Profili di ruolo professionale per l'ICT – Parte 4: Profili di ruolo professionale relativi alla sicurezza delle informazioni.

- **Data Specialist**, figura che presidia la gestione, la qualità e la disponibilità dei dati; l'utilizzo dell'IA richiede un rafforzamento delle competenze relative alla preparazione, trasformazione, qualità, tracciabilità e gestione dei dati utilizzati dai sistemi di IA.
- **Developer**, figura responsabile dello sviluppo e dell'implementazione delle applicazioni software; l'evoluzione verso sistemi basati sull'IA richiede competenze specifiche relative ad algoritmi, modelli di machine learning e deep learning, IA generativa ed elaborazione del linguaggio naturale.
- **Information Security Specialist**, figura che si identifica o affianca il Referente per la cybersicurezza e che presidia la sicurezza dei sistemi informativi, analizzando minacce e vulnerabilità e individuando le relative misure di prevenzione e mitigazione; l'IA richiede competenze ulteriori rispetto alle vulnerabilità e alle minacce specifiche dei sistemi di IA.

Oltre all'evoluzione dei profili professionali già operanti nell'ambito delle tecnologie dell'informazione, lo sviluppo, l'adozione e la gestione dei sistemi di IA richiedono competenze specialistiche riconducibili a specifici profili professionali. In questo ambito la UNI 11621-8[29] può costituire un utile riferimento per l'individuazione e la descrizione dei profili professionali operanti nel settore dell'Intelligenza Artificiale, secondo il modello della serie UNI 11621 e dell'European e-Competence Framework (e-CF)[30].

Tra i profili professionali maggiormente rilevanti ai fini delle Linee Guida IA, a titolo esemplificativo e non esaustivo, rientrano:

- **Responsabile dell'IA** o **Chief AI Officer**, responsabile della definizione e supervisione della strategia di IA, della relativa governance e della gestione dei rischi, assicurando

[29] UNI 11621-8:2026, *Attività professionali non regolamentate – Profili di ruolo professionale per l'ICT – Parte 8: Profili di ruolo professionale relativi all'Intelligenza Artificiale (IA)*.

[30] UNI EN 16234-1:2020, *e-Competence Framework (e-CF) – Framework comune europeo per i professionisti ICT in tutti i settori – Parte 1: Framework (modello di riferimento)*.

l'integrazione sicura ed efficace delle soluzioni di IA nei processi dell'organizzazione, la conformità normativa, la trasparenza, la tracciabilità e la sorveglianza umana. Nelle PA, in coerenza con il modello organizzativo definito nelle presenti Linee Guida, tali funzioni sono ricondotte al Responsabile per la Transizione Digitale (cfr. par. 4.1); nei contesti ad elevata intensità di utilizzo dell'Intelligenza Artificiale, la PA può istituire una figura dedicata al coordinamento dell'IA, il Responsabile dell'Intelligenza Artificiale (RIA), al quale può essere assegnato un proprio ufficio (UIA), che opera secondo le direttive del RTD.

- **AI Product Manager**, figura che assicura che prodotti e servizi basati su sistemi di IA rispondano alle esigenze degli utenti e dell'organizzazione, traducendo gli obiettivi strategici in requisiti e priorità e supervisionandone l'intero ciclo di vita, dalla definizione alla progettazione, dal rilascio al monitoraggio successivo, con attenzione a qualità, sicurezza, sostenibilità, conformità e impatti.
- **AI Prompt Engineer**, figura specializzata nella progettazione, validazione e manutenzione di prompt, template e catene di istruzioni per sistemi di IA generativa, con l'obiettivo di migliorare qualità, precisione, sicurezza e affidabilità degli output, riducendo allucinazioni, bias e rischi associati all'interazione con i modelli generativi. Tale figura è l'esperto che istruisce l'IA generativa alla produzione di output efficaci.
- **AI Algorithm Engineer**, figura specializzata nella progettazione, sviluppo, ottimizzazione e validazione di algoritmi e modelli di IA, nonché nella definizione delle relative architetture algoritmiche. Traduce requisiti funzionali e vincoli tecnici in specifiche algoritmiche e valuta robustezza, accuratezza, scalabilità, spiegabilità ed efficienza delle soluzioni.
- **AI Deep Learning Engineer**, esperta in reti neurali profonde, *deep learning* e programmazione, in grado di progettare e sviluppare soluzioni per problemi complessi che richiedono elevate capacità computazionali, quali, ad esempio, *computer vision*, riconoscimento vocale ed elaborazione avanzata del linguaggio.

- **AI Data Engineer**, figura specializzata nella progettazione, sviluppo e gestione delle pipeline e delle infrastrutture dati necessarie ai sistemi di IA, assicurando che i dati siano raccolti, trasformati e resi disponibili in modo efficiente, affidabile, tracciabile e coerente con i requisiti di qualità. La qualità dei dati costituisce infatti un elemento essenziale per l'affidabilità dei risultati prodotti dai sistemi di IA.
- **AI Data Scientist**, figura con competenze in analisi dei dati, statistica e *machine learning*, che analizza e interpreta grandi volumi di dati, sviluppa e valida modelli analitici e predittivi e trasforma i risultati delle analisi in informazioni utili ai processi decisionali. Opera con attenzione alla qualità dei dati, alla spiegabilità, all'equità e alla valutazione dei bias.
- **AI Security Specialist**, figura che si identifica o affianca il Referente per la cybersicurezza, è specializzata nella protezione dei sistemi di IA da vulnerabilità, minacce e attacchi, anche specifici dell'IA, e nell'applicazione dei principi di sicurezza *by design* lungo il ciclo di vita del sistema. Contribuisce alla robustezza, resilienza e conformità delle soluzioni di IA e collabora con i team di sviluppo, data science e governance.
- **AI Machine Learning Engineer**, figura che progetta, implementa e industrializza sistemi di *machine learning*, trasformando modelli e algoritmi in soluzioni produttive, scalabili e affidabili. Presidia l'intero ciclo di vita del modello, dalla preparazione dei dati e dall'addestramento fino al *deployment*, al monitoraggio e all'eventuale *retraining*, anche attraverso pratiche MLOps.
- **AI Natural Language Processing Engineer**, figura specializzata nella progettazione, sviluppo e implementazione di sistemi in grado di comprendere, interpretare e generare linguaggio umano, utilizzando tecniche di linguistica computazionale, *machine learning*, *deep learning* e modelli linguistici. Presidia l'intero ciclo di vita delle soluzioni NLP, con particolare attenzione all'accuratezza, ai *bias* linguistici e culturali e all'integrazione nei processi dell'organizzazione.

Accanto ai profili ICT, l'adozione dell'IA incide anche su professionalità già presenti nell'amministrazione, chiamate ad ampliare e aggiornare le proprie competenze rispetto alle specifiche implicazioni giuridiche, organizzative, comunicative e di tutela dei diritti connesse ai sistemi di IA. Tali figure mantengono la propria specializzazione di origine, integrandola con conoscenze e capacità specifiche necessarie a supportare un'adozione responsabile dell'IA.

- **Specialista in protezione dei dati personali**, esperto che può eventualmente identificarsi con il Responsabile della Protezione dei Dati (RPD), in grado di fornire adeguata consulenza specialistica alla PA nella costante verifica del rispetto della normativa in materia di protezione dei dati personali nell'utilizzo di sistemi di IA lungo l'intero ciclo di vita del sistema, con particolare riferimento ai rischi per i diritti e le libertà della persona.
- **Giurista informatico**, figura che integra competenze giuridiche e tecnologiche e supporta l'amministrazione nell'interpretazione e nell'applicazione della normativa connessa all'uso delle tecnologie digitali. Presidia, tra gli altri, gli aspetti relativi alla conformità normativa, alla protezione dei dati personali, alla sicurezza informatica, alla proprietà intellettuale, alla contrattualistica e alla gestione delle controversie. Con l'adozione dell'IA deve estendere le proprie competenze alle specifiche disposizioni applicabili ai sistemi di IA, alla redazione e revisione dei relativi contratti e licenze, nonché all'analisi e alla valutazione dei rischi giuridici e dei profili di responsabilità.
- **Comunicatore pubblico**, figura che presidia i processi di comunicazione interna ed esterna dell'amministrazione. Con l'adozione dell'IA deve sviluppare competenze specifiche per diffondere conoscenza e cultura dell'IA all'interno dell'organizzazione e verso cittadini e stakeholder, assicurando una comunicazione chiara, trasparente e comprensibile sulle modalità di utilizzo dei sistemi di IA e sui relativi impatti.

L'adozione dell'IA può determinare o rafforzare l'esigenza di professionalità non ancora stabilmente presenti nelle amministrazioni, caratterizzate da competenze trasversali necessarie a governare gli impatti etici, organizzativi e sociali dei sistemi di IA.

- **Organizational Designer,** figura che analizza e progetta assetti organizzativi, ruoli, processi e modalità di interazione tra persone e tecnologie. Con l'adozione dell'IA assume particolare rilievo nella definizione di modelli organizzativi capaci di integrare efficacemente l'interazione tra persone e sistemi di IA, assicurando chiarezza delle responsabilità, qualità del lavoro, sostenibilità organizzativa e adeguati meccanismi di sorveglianza umana.
- **Change Manager**, figura che presidia i processi di cambiamento organizzativo e supporta l'amministrazione nell'introduzione di nuove tecnologie e modalità operative. Con l'adozione dell'IA deve sviluppare la capacità di valutarne gli impatti su processi, ruoli, competenze e assetti organizzativi, definendo e attuando strategie di accompagnamento al cambiamento, gestione delle resistenze e adozione consapevole delle nuove soluzioni.
- **AI Ethicist**, esperto incaricato di condurre valutazioni etiche su tecnologie, applicazioni e pratiche di Intelligenza Artificiale, identificando, analizzando e prendendo decisioni sulle priorità valoriali, definendo Linee Guida e fornendo raccomandazioni per rendere operativi i valori morali in modo da tradurli in standard legali, in particolare su temi legati alla discriminazione, alla responsabilità, alla trasparenza e alla privacy. Una figura professionale dotata di competenze su analisi concettuale e norme, principi e ragionamento morali, in grado di facilitare discussioni e stimolare il confronto per migliorare le capacità decisionali nel dominio dell'Intelligenza Artificiale, sensibilizzando chi realizza e utilizza gli strumenti di Intelligenza Artificiale, il quale deve restarne pienamente responsabile. La figura in questione dovrebbe inoltre seguire strettamente gli sviluppi

della ricerca nel dominio dell'IA e impegnarsi nella loro divulgazione per creare un "vocabolario etico" condiviso.

I profili professionali sopra individuati hanno carattere esemplificativo e non esaustivo e POSSONO essere integrati, arricchiti, adattati o ulteriormente specializzati dalle PA in funzione delle proprie esigenze organizzative e operative, degli specifici ambiti di applicazione dell'IA, nonché dell'evoluzione tecnologica, normativa e delle competenze professionali.

Le PA DEVONO individuare e sviluppare, anche attraverso percorsi di reclutamento, formazione, *upskilling* e specializzazione, le professionalità e le competenze necessarie alla progettazione, adozione, gestione e supervisione dei sistemi di IA, assicurando adeguati percorsi di sviluppo professionale per dirigenti e specialisti chiamati a ricoprire ruoli connessi all'IA.

Le PA DEVONO assicurare che i dirigenti pubblici, oltre che i Responsabili per la Transizione Digitale e le relative strutture, acquisiscano e mantengano un livello di conoscenze e competenze in materia di IA adeguato alle rispettive responsabilità, al fine di consentire l'assunzione di decisioni consapevoli e informate in merito all'adozione e all'impiego dei sistemi di IA. Tali competenze DEVONO consentire di valutarne l'opportunità di utilizzo, i benefici attesi, i rischi e gli impatti sui processi e sui servizi pubblici, nonché di assicurare il rispetto del quadro normativo applicabile e degli obiettivi di trasformazione digitale dell'amministrazione.

In particolare, i dirigenti pubblici necessitano di comprendere le proposte di progetti avanzate dagli esperti ed essere in grado di valutarne le opportunità, i rischi, le implicazioni e gli impatti[31], in coerenza con quanto definito nel cap. 4 delle presenti Linee Guida, 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

[31] ARISA (2023). «AI Skills Needs Analysis».

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'IA, alla stregua degli specialisti, ove responsabili dell'adozione dell'IA, i dirigenti pubblici 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[32].

8.1 Indicazioni operative per le PA

La definizione di attività formative per tutto il personale, la formazione continua specialistica per l'Ufficio per la Transizione Digitale (UTD), l'individuazione e selezione delle competenze necessarie alla gestione delle attività sono 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 priorità strategiche di ciascuna amministrazione e delle esigenze formative, la definizione degli obiettivi a livello organizzativo e individuale, 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, quali il Piano integrato dell'attività e organizzazione (PIAO) e il Piano della

[32] Si veda sul punto il Decreto del Ministro per la Pubblica Amministrazione del 28 settembre 2022 di adozione delle Linee Guida 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.

performance, come recentemente richiamato dalla Direttiva del Ministro per la Pubblica Amministrazione del 14 gennaio 2025 in materia di formazione[33].

Al fine di dotarsi dell'adeguato sistema di competenze necessario per approcciare l'adozione di applicazioni che si basano sull'IA, le PA DEVONO 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 o attraverso il ricorso a risorse nazionali e più in generale attraverso un'azione sinergica degli enti preposti alla formazione per il pubblico impiego e delle stesse amministrazioni.

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.

Le PA DEVONO promuovere l'*AI literacy* di tutti i dipendenti pubblici, riconosciuta nell'ambito della Direttiva del Ministro per la Pubblica Amministrazione sulla formazione 2025 a tutti gli effetti una competenza digitale di base, attraverso la definizione e il monitoraggio di specifici obiettivi di performance individuali connessi allo sviluppo di tali competenze. In particolare, le amministrazioni DEVONO, al minimo, favorire il conseguimento degli obiettivi formativi definiti dalla Direttiva sulla formazione del Ministro della PA del 24 marzo 2023[34] 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 progressivamente pubblicati a catalogo. Le amministrazioni POSSONO altresì

[33] Direttiva del Ministro per la Pubblica Amministrazione "Valorizzazione delle persone e produzione di valore pubblico attraverso la formazione. Principi, obiettivi e strumenti" del 14 gennaio 2025

[34] Direttiva del Ministro per la 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

rafforzare ulteriormente l'AI literacy dei propri dipendenti avvalendosi di interventi formativi mirati su temi specifici, realizzati internamente o attraverso risorse e professionalità esterne variamente definite.

Le PA DEVONO garantire la massima diffusione delle Linee Guida sull'IA di 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, dei RPD, dei Referenti per la cybersicurezza 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 PA DEVONO promuovere e favorire la partecipazione dell'UTD alla community promossa da AgID e dedicata ai Responsabili per la Transizione Digitale (progetto ReTe Digitale) che per sua natura rappresenta lo spazio privilegiato per la condivisione, collaborazione e confronto sull'IA. Le PA 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 PA possono avvalersi oltre che della formazione disponibile sulla piattaforma Syllabus, dell'offerta formativa promossa dalla Scuola Nazionale dell'Amministrazione e dai suoi poli formativi territoriali, da Formez PA, dalle Università, anche attraverso il programma "PA 110 e lode" finanziato dal Dipartimento della funzione pubblica della PCM e, più in generale, dai soggetti pubblici e privati accreditati per la formazione continua al personale della Pubblica Amministrazione.

Le PA che attuano progetti di trasformazione digitale basati sull'AI DEVONO altresì 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 PA o in determinate tipologie di PA (si pensi, a mero titolo di esempio, 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, spaziando tra ambiti di carattere tecnico-specialistico, giuridico, etico e le soft skills. In aggiunta alle soluzioni fin qui evidenziate, per far fronte a tali interventi, le singole PA possono sviluppare specifici progetti formativi anche di carattere innovativo (basati, tra gli altri, su seminari e conferenze, hackathon, attività laboratoriali, programmi di mentorship), coinvolgendo gli operatori pubblici e privati che compongono il sistema dell'offerta formativa rivolta alla PA e avvalendosi, nel breve periodo, di risorse promosse nell'ambito di progettualità previste dal PNRR, come nel caso dell'iniziativa

PerformaPA, in aggiunta alle risorse europee (es. fondi strutturali), nazionali o regionali, dedicate a vario titolo alla formazione.

Le PA coinvolte in attività di sviluppo, acquisizione e gestione dell'IA DEVONO favorire l'accesso di tutto il personale a percorsi di formazione sui temi dell'IA e della protezione dei dati personali, considerato che "*il diritto alla vita privata e alla protezione dei dati personali deve essere garantito durante l'intero ciclo di vita del sistema di IA*" (cfr. considerando 69 AI Act), anche al fine di rendere effettiva la "sorveglianza umana" da parametrare in considerazione del livello di rischio per i diritti umani (cfr. cons. 75 e 91 AI Act).

Le PA coinvolte in attività di sviluppo, acquisizione e gestione dell'IA DEVONO, inoltre, favorire l'accesso a percorsi di acquisizione e mantenimento di competenze tecnico-specialistiche e certificazioni professionali, rivolti all'UTD e in generale ai professionisti IT coinvolti nello sviluppo, collaudo, monitoraggio ed evoluzione dell'IA, a partire da una rilevazione sistematica e periodica dei fabbisogni di competenze, anche avvalendosi della collaborazione con Università (es. progetto PA 110 e lode), dei Poli di innovazione digitale, degli istituti di alta formazione e centri di competenza a livello nazionale e internazionale.

Le PA DEVONO, altresì, prevedere, nei propri piani formativi, percorsi relativi alla cybersicurezza dell'IA e concernenti la capacità di rispondere a minacce, incidenti e crisi cibernetiche nel contesto dell'IA nonché competenze diffuse in materia di protezione dei dati personali.

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

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 Ministero dell'Università e della 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, poli di innovazione digitale, laboratori e centri di ricerca, nonché con i 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 PA e, in particolare, gli enti locali e le amministrazioni erogatrici di servizi POSSONO contribuire a promuovere l'AI *awareness* e AI *literacy* dei cittadini destinatari di tali servizi, affinché riescano ad interagire in modo efficace, consapevole e sicuro con soluzioni di IA.

Le PA DEVONO mettere i cittadini nelle condizioni di conoscere le potenzialità, i limiti e gli impatti delle tecnologie di IA, al fine di favorire un accesso ai servizi pubblici consapevole, sicuro e rispettoso dei propri diritti. A tal fine, le PA, e in particolare gli enti territoriali, DEVONO

promuovere iniziative di informazione, sensibilizzazione e facilitazione rivolte ai cittadini, anche valorizzando, ove pertinenti e disponibili, le iniziative finanziate dal PNRR, quali il Servizio civile digitale e la Rete dei servizi di facilitazione digitale, nonché attivando forme di collaborazione con le amministrazioni e gli altri soggetti competenti, al fine di accrescere la consapevolezza sui temi dell'Intelligenza Artificiale e sulle modalità di accesso ai servizi che si avvalgono di tali tecnologie[35].

Le PA POSSONO prevedere, promuovere e attivare progetti specifici, nonché campagne informative e formative mirate, anche attraverso il coinvolgimento delle realtà sociali ed economiche presenti nei rispettivi territori.

[35] 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.

9. Gestione e qualità dei dati

L'AI Act nel considerando 67 evidenzia l'importanza di implementare una politica finalizzata al miglioramento della qualità dei dati quale prerequisito essenziale per lo sviluppo e l'adozione di sistemi di IA affidabili[36]. 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*". Ne consegue che la governance e la corretta gestione dei dati costituiscono presupposti essenziali per assicurarne disponibilità, qualità e affidabilità. Sebbene lo sviluppo e l'addestramento dei sistemi di IA richiedano spesso la disponibilità di quantità significative di dati, è soprattutto la qualità di tali dati, unitamente a una loro corretta gestione, a incidere sulla precisione, sull'affidabilità e sulla robustezza dei risultati prodotti dal sistema.

A tale proposito, il rapporto tecnico UNI CEI CEN/CLC/TR 18115:2025 "*Governance dei dati e qualità per l'Intelligenza Artificiale nel contesto europeo*" fornisce un quadro esaustivo sugli standard connessi all'IA, con una attenzione particolare al tema dei dati che alimentano i sistemi di Intelligenza Artificiale e al loro ciclo di vita, spaziando dalla governance dei dati in singoli progetti fino al livello di una governance mondiale nel contesto di fenomeni senza confini quali ad esempio quelli del clima, dell'energia o dei rischi pandemici.

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

[36] 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.

sono tutti aspetti di cui bisogna tener conto quando si parla di IA. Tale quadro comprende, tra l'altro:

- l'adozione di una Strategia per l'Unione dei 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 Direttiva Open Data).

Il tema della gestione dei dati, in ogni fase del processo connesso all'adozione, il procurement e sviluppo di soluzioni di IA (training, documenti di ancoraggio, output, ecc.) 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 rilevanza le regole di interoperabilità, in particolare semantica, definite con riferimento al quadro normativo indicato e le strategie di *licensing* connesse all'acquisizione, elaborazione, rilascio/pubblicazione e riutilizzo dei dataset pubblici.

Ferma restando la possibilità della Commissione Europea di elaborare iniziative, anche di natura settoriale, per agevolare la riduzione degli ostacoli tecnici che ostruiscono lo scambio transfrontaliero di dati per lo sviluppo dell'IA, anche per quanto riguarda l'interoperabilità semantica e tecnica dei diversi tipi di dati (cfr. considerando 165 dell'AI Act), le PA DEVONO seguire le indicazioni già definite in Linee Guida e altri documenti vincolanti, sia a livello europeo che nazionale, in termini di interoperabilità.

Il requisito fondamentale è quello ribadito nelle Linee Guida per l'interoperabilità tecnica secondo cui "*la comunicazione tra soggetti DEVE utilizzare modelli dati condivisi, in modo da razionalizzare e uniformare la rappresentazione dell'informazione quale presupposto per favorire l'interoperabilità tra soggetti differenti*". Alla luce di questo, la regola che, in generale, deve guidare è che (i) i dati debbano essere disponibili attraverso sistemi interoperabili e che (ii) bisogna fare riferimento a modelli dati, ontologie e vocabolari controllati esistenti per verificare se i concetti siano già provvisti di entità, proprietà e, ove presenti, URI ampiamente adottati, specie se in ambito europeo. Un esempio è fornito dal framework di interoperabilità relativo al dominio dei dati territoriali definito con la Direttiva cosiddetta INSPIRE (Direttiva 2007/2/CE) e ai conseguenti Regolamenti di esecuzione su metadati e specifiche tecniche dei dati e servizi, che è indicato nel Regolamento di esecuzione (UE) 2023/138 anche come riferimento per molte delle serie di dati di elevato valore. Lo stesso D.Lgs. 36/2006, all'articolo 6, comma 9, indica che, in caso di dati territoriali e del monitoraggio ambientale, si applicano le regole tecniche definite dalla Direttiva INSPIRE.

Riguardo alle strategie di licensing, 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)[37], diventa una imprescindibile necessità, evitando la parcellizzazione – specie per i dati di elevato 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 ottimale riutilizzabilità e la interpretabilità nella forma più trasparente anche

[37] Adottate da AgID con la Determinazione del Direttore Generale n. 183/2023 ai sensi dell'articolo 12 del D.Lgs. 36/2006

da parte dei nuovi sistemi di Intelligenza Artificiale già disponibili o in corso di realizzazione. È fortemente raccomandato utilizzare dataset istituzionali per l'addestramento dei sistemi di IA, poiché essi assicurano l'affidabilità della fonte e la solidità del sistema che alimenta i dati.

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 di 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.

Anche considerato che i dataset utilizzati in fase di addestramento, convalida e prova devono essere pertinenti, sufficientemente rappresentativi e, nella misura in cui è possibile, esenti da errori e completi nell'ottica della finalità prevista dai sistemi di IA (articolo 10, par. 3, AI Act), in quanto rilevante, particolare cura DEVE essere prestata alla corretta attuazione dei principi di protezione dei dati personali di cui all'articolo 5 del GDPR.

9.1 Tipologie di dati

Come detto in premessa, i dati - anzitutto i dati personali per le più penetranti garanzie previste dal diritto (primario e secondario) dell'Unione e dal diritto nazionale - costituiscono l'aspetto più rilevante e allo stesso tempo più critico nell'adozione di IA per la PA e non solo, 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 e, allo stesso modo, la PA può utilizzare dati provenienti da fonti esterne all'organizzazione, sempre nel rispetto dei presupposti e delle garanzie richieste dalla normativa in materia di protezione dei dati personali e dell'AI Act.

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. articolo 3, paragrafo 1, punto 29, AI Act);
- **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. articolo 3, paragrafo 1, punto 30, AI Act);
- **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. articolo 3, paragrafo 1, punto 32, AI Act).

Con riferimento alle fonti di dati, possono essere distinti:

- **dati endogeni**, cioè dati generati internamente 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 esogeni**, acquisiti tramite procedure di appalto 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") e che seguano le regole tecniche definite a livello europeo e nazionale (ad es. le indicazioni sui formati aperti e leggibili meccanicamente presenti nelle Linee Guida Open Data). 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. articolo 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 IA 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, richiamati anche nell'AI Act (v. considerando 111, articolo 10, par. 5, articolo 59), 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 protezione dei dati personali. 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 è stata avviata una regolamentazione specifica è quello dei dati sanitari, per il quale è stato recentemente adottato il Regolamento (UE) 2025/327 su tale spazio.

9.2 Caratteristiche dei dati

Come richiamato dalla Strategia Italiana per la IA 2024-2026, 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 costituire, preservare tecnicamente e giuridicamente e pubblicare dataset di qualità ed effettivamente riutilizzabili.

Come azione propedeutica, le PA DEVONO 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. A questo si aggiungono gli aspetti relativi al trattamento dei dati personali o alla proprietà intellettuale nel caso di utilizzo di dati aperti per l'addestramento, la convalida e la prova dei sistemi di IA per i quali, oltre alla trattazione in altre sezioni di questo documento, si richiamano le indicazioni presenti nelle Linee Guida Open Data.

Tale azione è anche in linea con il principio 2.1 del documento di raccomandazioni OCSE[38] che spinge i governi a investimenti anche in set di dati aperti che siano rappresentativi e rispettino la

[38] Recommendation of the Council on Artificial Intelligence

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.

I sistemi di IA apprendono e producono analisi affidabili se hanno a disposizione una adeguata quantità di dati di elevata qualità. DEVE essere garantito, altresì, il rispetto della normativa in materia di protezione dei dati personali e di 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 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 derivanti da specifiche norme, come l'accessibilità (di cui alla legge n. 4/2004, al D.Lgs. 82/2022 e alle relative Linee Guida AgID) o la protezione dei dati correlata alle indicazioni derivanti dal GDPR e dal Codice privacy.

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.

Oltre alle caratteristiche di qualità indicate nelle Linee Guida Open Data (**accuratezza**, **coerenza**, **completezza** e **attualità**) e a quelle derivanti da specifici obblighi normativi (**accessibilità**, **disponibilità** e **riservatezza**), considerando i requisiti di qualità individuati dall'AI Act, per i sistemi di IA DEVONO essere garantite anche le caratteristiche di qualità dei dati di seguito riportate e derivanti dagli standard ISO/IEC 25012 e ISO/IEC 5259-2, con l'aggiunta di ulteriori a seconda del contesto.

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).

In tema di qualità di dati per l'IA, la Strategia per l'Unione dei dati prevede che la Commissione presenterà una richiesta di normazione per l'elaborazione di una norma europea sulla qualità dei dati che riguardi la completezza, la coerenza, la provenienza, la chiarezza semantica e la governance, in modo che imprese, autorità di regolamentazione e ricercatori dispongano di

parametri di riferimento condivisi per serie di dati affidabili. Le indicazioni in questo documento sono in linea con la direzione indicata nella Strategia europea; tuttavia, sarà necessario presidiare l'iniziativa di standardizzazione europea e procedere all'eventuale allineamento di queste indicazioni alla norma che sarà adottata.

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 vengono alimentati da grandi quantità di dati per addestrare modelli complessi. Analogamente, i dati includono una varietà di esempi rappresentativi delle possibili situazioni che i sistemi di IA si trovano a valutare.

Un altro aspetto rilevante consiste nel "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.

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 IA, 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 IA, la PA DEVE 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, norme e adempimenti quali quelli di apertura, trasparenza, accountability e del perseguimento di una 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. La PA DEVE definire

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 gestire una data governance utile al sistema di IA in fase di sperimentazione, il percorso delineato dalle Linee Guida Open Data richiamate può essere opportunamente adattato in modo che sia utile ad attività specifiche per l'IA, come meglio precisato nel seguito e rappresentato nella figura 5.

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 considerate 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, protezione dei dati personali);
- **Fase 4**. Lecita 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 5. Le fasi identificate con celle con sfondo grigio sono quelle relative ad attività e operazioni finalizzate esclusivamente all'IA.

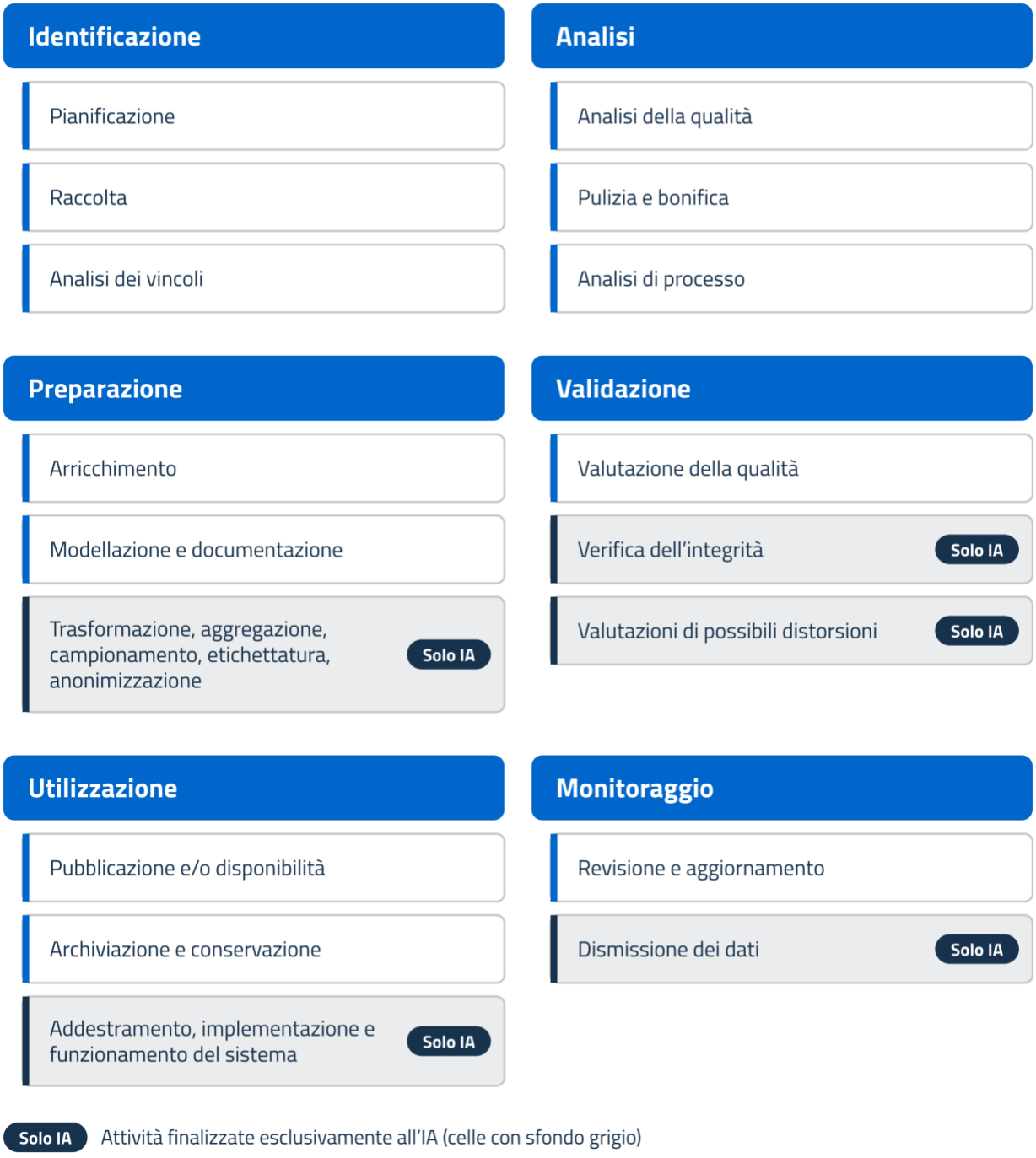


Figura 5 – Modello di processo di gestione dei dati.

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 e il rispetto dei principi di protezione dei dati personali. 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). Propedeutico a ciò (vedi ISO/IEC 8183, fase 2) è, tra l'altro, l'individuazione dell'architettura dati più opportuna (ad es. Data Lake, Data Mesh o Data Fabric) e della conseguente tecnologia, anche facendo riferimento ad approcci innovativi per la gestione dei dati.

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, *data masking*, 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 indicati dall'AI Act. Ulteriori validazioni specifiche, da investigare in audit periodici, potrebbero essere richieste nei singoli domini.

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, aggiornamenti e audit 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) in caso di dati non più utilizzati dal sistema (ad es. cancellazione sicura, archiviazione, rielaborazione). Lo Standard indica, tra l'altro, che in questa fase alcune categorie di dati dovrebbero essere conservate per scopi di auditing (ad esempio, i dati di log per dimostrare la conformità), così come la cancellazione dei dati può avvenire per esplicite condizioni della licenza o per applicazione della normativa sulla protezione dei dati personali.

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.

Nella figura 6 viene rappresentata la correlazione tra le fasi del ciclo di vita dell'IA con le fasi del processo di gestione dei dati. Fermo restando che la valutazione della qualità dei dati deve accompagnare tutto il ciclo di vita degli stessi, viene anche proposto, analogamente allo Standard ISO/IEC 5259-1:2024, durante quali fasi del ciclo di vita effettuare la misurazione delle caratteristiche di qualità considerate nel modello definito nel presente documento.

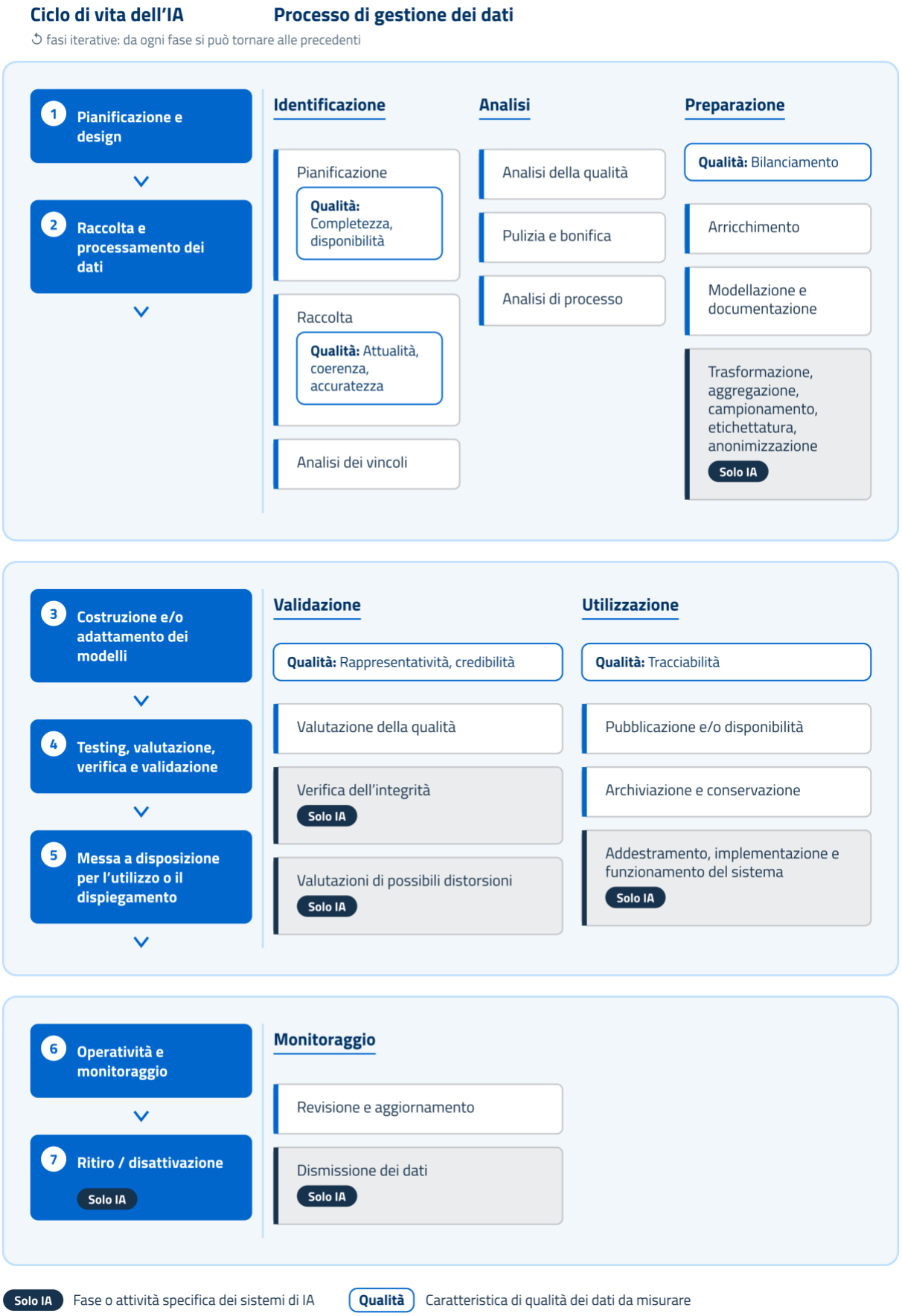


Figura 6 – Correlazione tra le fasi del ciclo di vita dell'IA con le fasi del processo di gestione dei dati.

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 Digitale (RTD), 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 (RPD) laddove siano coinvolti dati personali.

Nel caso in cui le PA ricorrano a sistemi di Intelligenza Artificiale erogati mediante servizi cloud, esse DEVONO garantire che il trattamento dei dati avvenga nel rispetto delle normative applicabili, con particolare riferimento alla protezione dei dati personali e alla sicurezza delle informazioni. Le PA DEVONO, inoltre, individuare chiaramente i fornitori e gli eventuali subfornitori coinvolti nella catena di erogazione del servizio e assicurare adeguate garanzie affinché i fornitori non utilizzino i dati trattati per finalità proprie o incompatibili con quelle dell'amministrazione. A tal fine le PA DEVONO tenere conto delle indicazioni operative contenute nelle Linee Guida procurement.

L'individuazione di apposite strutture di coordinamento o gruppi di lavoro esistenti o da istituire, 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*". 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, la governance dei dati nei singoli Enti, anche in relazione ai dati per l'IA, DOVREBBE essere 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 articolo 50-ter, dati per l'IA).

In questo contesto, è rilevante anche il tema della collaborazione istituzionale trattato anche in diverse edizioni del 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 e/o provinciali), 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 primo trattamento dei dati.

Questo può tradursi nella promozione di iniziative di aggregazione 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 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, individuate nelle Linee Guida sviluppo quali *operatori avanzati*, 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 (ad esempio *operatori base o esperti*, vedi Linee Guida sviluppo), 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, implementando così una architettura coerente con quella definita nelle Linee Guida sviluppo al cap. 3;

- 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 possono essere utilizzati per lo svolgimento e/o il supporto di attività che comportano il trattamento di dati anche personali: lo sviluppo, l'acquisto e l'adozione di tali sistemi da parte delle PA, pertanto, DEVONO intervenire nel rispetto del diritto fondamentale alla protezione dei dati personali, come esplicitamente previsto dall'articolo 2, paragrafo 7, dell'AI Act.

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

L'AI Act stesso chiarisce che il diritto dell'Unione in materia di protezione dei dati personali, della vita privata e della riservatezza delle comunicazioni si applica ai dati personali trattati in relazione ai diritti e agli obblighi stabiliti dall'AI Act ed esplicita, inoltre, che non è pregiudicata l'applicazione del vigente diritto dell'Unione sul trattamento dei dati personali, inclusi i compiti e i poteri delle autorità di controllo indipendenti competenti a monitorare la conformità con tali strumenti e che parimenti sono impregiudicati gli obblighi dei fornitori e dei deployer dei sistemi di IA nel loro ruolo di titolari o responsabili del trattamento di dati personali.

La PA, in particolare, DEVE verificare il rispetto dei principi enunciati all'articolo 5 del GDPR - compresa la responsabilizzazione della PA stessa che ne farà uso in qualità di titolare del trattamento, ai sensi del paragrafo 2 della medesima disposizione normativa, documentando le valutazioni poste alla base delle proprie scelte in merito all'adozione di IA -, nel rispetto della normativa unionale e nazionale in materia di protezione dei dati personali e dei provvedimenti e

pareri emessi dal Garante europeo della protezione dei dati, dal Comitato europeo per la protezione dei dati 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, la PA DEVE effettuare un'analisi di tale sistema sotto lo specifico profilo della protezione dei dati personali, come previsto dalla normativa vigente in materia di protezione dei dati personali al fine di garantire e dimostrare il rispetto di quanto segue [39] :

- il trattamento dei dati personali mediante il sistema di IA avviene in modo lecito, corretto e trasparente nei confronti dell'interessato; in particolare la PA DEVE valutare e individuare i necessari presupposti di liceità e le imprescindibili garanzie di trasparenza e di tutela dei diritti 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; in particolare, nell'utilizzo di strumenti di IA nello svolgimento delle proprie attribuzioni istituzionali, la PA DEVE richiamare le specifiche finalità di interesse pubblico nell'alveo delle attribuzioni che la legge le affida;
- è sempre individuata un'idonea base giuridica - anche integrata con la Legge 23 settembre 2025, n. 132 - del trattamento di dati personali mediante il sistema di IA, sia nella fase di sviluppo o acquisto dello stesso sia in quella, successiva e distinta, di utilizzo, anche con riferimento all'eventuale ulteriore uso dei dati per scopi diversi da quelli della raccolta, anche ai sensi dell'articolo 6, paragrafo 4, del GDPR;

[39] Sugli aspetti oggetto di trattazione, si vedano più estesamente i provvedimenti dello European Data Protection Board e del precedente Article 29 Working Party nonché 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.

- 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; in particolare, la PA DEVE valutare l'effettiva necessità di ricorrere a dati personali e, in caso non sia possibile procedere con l'utilizzo di dati anonimi o dati sintetici, predisporre la minimizzazione degli stessi;
- i dati personali trattati a mezzo del sistema di IA sono esatti e, se necessario, aggiornati, adottando la PA tutte le misure ragionevoli per garantire l'esercizio dei diritti da parte degli interessati 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 e sono adeguatamente individuate e predisposte le misure concernenti modalità e tempi di conservazione dei dati personali trattati con sistemi di IA;
- 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, anche specificamente correlate all'utilizzo di strumenti di IA, adeguate a proteggere i dati da 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 adeguatamente tutto il personale in merito ai rischi connessi all'utilizzo di sistemi di IA 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 articoli 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, in seguito, periodicamente - anche in caso di adozione di sistemi di IA non classificabili ad alto rischio e pertanto non soggetti all'obbligo di valutazione d'impatto sui diritti fondamentali di cui all'articolo 27 dell'AI Act -, la PA effettua una valutazione d'impatto sulla protezione dei dati personali ai sensi dell'articolo 35 del GDPR o 27 della Direttiva (UE) 2016/680, delle Linee Guida dell'Article 29 Working Party[40] e dei provvedimenti del Garante per la protezione dei dati personali, 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; la portata dell'impatto negativo del sistema di IA sui diritti fondamentali fra cui la protezione dei dati personali, invero, è di particolare rilevanza ai fini della classificazione di un sistema di IA tra quelli ad alto rischio;
- le categorie particolari di dati personali e i dati personali relativi a condanne penali e reati, come individuati agli articoli 9-10 del GDPR, sono trattati a mezzo di sistemi di IA previa individuazione di una idonea base giuridica e di misure di sicurezza tecniche e organizzative appropriate, specifiche e documentate nel rispetto del principio di responsabilizzazione, non solo nel doveroso rispetto di quanto stabilito dalla normativa

[40] 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.

unionale e nazionale in materia di protezione dei dati personali ma anche ai sensi dello stesso AI Act, che prevede regole specifiche sulla protezione delle persone fisiche con riguardo al trattamento dei dati personali.

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'articolo 4, n. 7) del GDPR e, pertanto, di avere la responsabilità (anche in ottica di responsabilizzazione, ai sensi dell'articolo 5, paragrafo 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.

In tale contesto la PA DEVE valorizzare il ruolo e assicurare il coinvolgimento del Responsabile della protezione dei dati (RPD) nell'ambito di ogni fase del ciclo di vita di un sistema di IA (cfr. paragrafo 3.1. delle presenti Linee guida), nel rispetto di quanto previsto dal GDPR in merito alla posizione e ai compiti del RPD, anche individuando forme di collaborazione con il Responsabile per la Transizione Digitale.

Con specifico riferimento ai trattamenti effettuati per finalità di cura e di ricerca scientifica, le presenti Linee Guida rinviano alle specifiche disposizioni previste dall'AI Act, dal GDPR, dal Codice privacy, dalla citata Legge 23 settembre 2025, n. 132, con particolare focus sugli articoli 8 e 9, rispettivamente rubricati "*Ricerca e sperimentazione scientifica nella realizzazione di sistemi di Intelligenza Artificiale in ambito sanitario*" e "*Disposizioni in materia di trattamento di dati personali*", nonché alla specifica disciplina di settore.

Si ricorda, inoltre, che il considerando 25 dell'AI Act, nell'escludere dal suo ambito di applicazione "*i sistemi e i modelli di IA specificamente sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici*", sottolinea che "*In ogni caso, qualsiasi attività di ricerca e sviluppo*

dovrebbe essere svolta conformemente alle norme etiche e professionali riconosciute nell'ambito della ricerca scientifica e dovrebbe essere condotta conformemente al diritto dell'Unione applicabile".

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:

"*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)"[41].

[41] Si veda il "*Decalogo per la realizzazione di servizi sanitari nazionali attraverso sistemi di Intelligenza Artificiale*". 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.

Su specifici aspetti di protezione dei dati personali trattati nel contesto dei modelli di IA, si è espresso altresì il Comitato europeo per la protezione dei dati[42], 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[43].

La PA DEVE valutare se il modello di IA possa o meno considerarsi anonimo, con particolare riferimento alla probabilità di ottenere o ricostruire, intenzionalmente o meno, i dati personali utilizzati per addestrare il modello, anche attraverso il mero utilizzo del sistema di IA.

Anche con riferimento a quanto sopra, inoltre, in caso di ricorso a responsabili del trattamento che utilizzino strumenti di IA, la PA DEVE esplicitare, anche mediante la designazione di cui all'articolo 28 del GDPR, il divieto di trattamento dei dati personali per addestrare o riaddestrare i modelli da rendere disponibili a terzi, nel rispetto del principio di limitazione della finalità e degli obblighi gravanti sui responsabili.

[42] Si veda il parere n. 28 del 17 dicembre 2024

[43] Si veda il parere n. 28/2024 dell'EDPB.

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[44], 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[45] 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.

In aggiunta ai validi riferimenti presentati nel presente capitolo (es.: NIST, ENISA), le PA DEVONO tenere in considerazione le norme tecniche prodotte dal CEN/CENELEC JTC 21 "Artificial Intelligence", a seguito della richiesta di standardizzazione della Commissione Europea (cfr. par. 4.4), in quanto rappresentano standard tecnici armonizzati idonei a garantire la presunzione di conformità ai requisiti di cybersicurezza previsti dall'articolo 15 dell'AI Act.

---

[44] Agenzia per l'Italia Digitale (AgID), Piano Triennale per l'informatica nella Pubblica Amministrazione - edizione 2024-2026, aggiornamento 2025.

[45] La definizione è tratta dall'articolo pubblicato sul sito di MITRE ATLAS. MITRE ATLAS è una *knowledge base* delle tattiche e tecniche avversarie relative ai sistemi di IA.

11.1 Tassonomie di attacco

In aggiunta ai tradizionali attacchi cyber[46] 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[47] 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[48].

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

[46] 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.

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

[48] 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[49].

Come possibili strategie di mitigazione, è possibile applicare l'**adversarial 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[50].

Questi attacchi possono essere distinti in:

[49] 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.

[50] 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.

- **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 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 par. 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 cybersicurezza 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) sviluppato dal *National Institute of Standards and Technology (NIST)*.

Il framework è disegnato per assistere organizzazioni e individui, definiti come *AI Actors*[51], 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.

[51] 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.

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 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.

Per supportare le organizzazioni nell'uso del framework, il NIST ha sviluppato un playbook 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 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.

11.4 Minacce

Ogni fase del ciclo di vita è caratterizzata da una o più minacce[52]. In accordo a quanto indicato nel documento di ENISA "Artificial Intelligence Cybersecurity Challenges" 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[53] 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 (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);

52 Una medesima minaccia può far riferimento a più fasi del ciclo di vita.

53 Il par. 11.1 tratta le varie categorie di attacchi associati a questo tipo di minacce.

- **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 IA, 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 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*[54] 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.

[54] National Cyber Security Centre (NCSC), Guidelines for secure AI system development.

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 cybersicurezza 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[55] 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

[55] Per l'elenco delle categorie di asset si faccia riferimento al par. 11.3.

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.

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 cybersicurezza 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 contesto 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 cap. 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 cybersicurezza adeguate ai connessi rischi sulle infrastrutture ICT utilizzate in ogni parte del ciclo di vita dei sistemi di IA. Per le misure di cybersicurezza applicabili alle infrastrutture ICT, ferma restando la normativa applicabile, 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:* **pianificazione e design**

pratiche di sicurezza: stabilire criteri di sicurezza e protezione dei dati personali già in questa fase per garantire l'obiettivo rispetto ai requisiti normativi e operativi.

1.2 *fase:* **raccolta e processamento dei dati**

pratiche di sicurezza: applicare tecniche di protezione dei dati personali, come l'anonimizzazione, la pseudonimizzazione e la crittografia. Verificare la sicurezza delle sorgenti esterne e l'affidabilità dei fornitori di dati. Garantire che i dati, specialmente quelli personali, siano trattati in modo conforme ai principi di minimizzazione dei dati e integrità. Proteggere l'accesso durante il trattamento di tali dati. Assicurare che i dati siano protetti contro modifiche non autorizzate durante questa fase.

1.3 *fase*: **costruzione e/o addestramento dei modelli**

pratiche di sicurezza: garantire che le informazioni sensibili non siano esposte e che i dati trattati siano protetti contro accessi non autorizzati. Assicurarsi che il modello abbia una robustezza adeguata contro attacchi adversarial (le vulnerabilità possono variare in base al modello, valutare di integrare i dettagli nelle sezioni apposite). Assicurarsi che i dati utilizzati per l'addestramento siano privi di contaminazioni o bias. Controllare che le modifiche agli iperparametri non espongano il modello a nuove vulnerabilità. Proteggere il processo di tuning da accessi non autorizzati. 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.

1.4 *fase*: **messa a disposizione per l'utilizzo o il dispiegamento**

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 messo a disposizione.

1.5 *fase*: **operatività e monitoraggio**

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. Effettuare un'analisi delle prestazioni del modello in ambienti controllati per assicurarsi che continui a operare in modo sicuro e affidabile.

1.6 *fase*: **ritiro e disattivazione**

pratiche di sicurezza: definire e adottare procedure per lo smaltimento sicuro dei componenti ritirati e/o disattivati quali ad esempio la sanitizzazione dei dati del modello.

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*).



Agenzia per l'Italia Digitale

Sede: Via Liszt 21, 00144 Roma

Telefono: +39 06 852641

PEC: protocollo@pec.agid.gov.it

Sito web: www.agid.gov.it

Codice fiscale: 97735020584

Il documento

Linee Guida per l'adozione di IA nella Pubblica Amministrazione · versione 1.0

Documento redatto con il modello dei documenti ufficiali AgID, versione 1.0.

Torna al sommario