MARZO 2024

Innovare con Intelligenza

Linee guida per un'IA responsabile nei settori pubblico e privato

Stefano da Empoli, Silvia Compagnucci, Alessandro D'Amato, Maria Rosaria Della Porta, Enrica Lipilini, Domenico Salerno, Daniela Suarato

I progressi nel campo dell'intelligenza artificiale hanno condotto i governi e le organizzazioni sovranazionali attraverso commissioni ad hoc e fora internazionali a interessarsi sempre di più del tema per coglierne le relative opportunità, mitigandone al contempo i possibili rischi in termini di etica, privacy, sicurezza e, più in generale, dell'impatto sociale dell'IA e delle decisioni intraprese con l'ausilio – più o meno ampio – di questa tecnologia.

- A maggio 2023, l'*OECD.AI Policy Observatory* ha registrato ben 930 iniziative a livello di policy provenienti da oltre 70 giurisdizioni.
- I documenti programmatici e le linee guida analizzate condividono molti punti di contatto, specialmente con riguardo ai principi etici e di sicurezza da applicare all'IA, oltre che in termini di cooperazione internazionale e collaborazione multi-stakeholder.
- Le iniziative sul piano internazionale esaminate nel presente studio evidenziano una divergenza di vedute con riferimento ai modelli *open-source*, al rafforzamento della regolamentazione, alla misurazione dell'impatto dell'IA in termini di sostenibilità e di necessità di risorse energetiche, nonché per l'individuazione della responsabilità in caso di danni causati da decisioni assunte sulla base di algoritmi di IA.
- In un contesto generale di crescente attenzione sulle due sponde del Pacifico sono molto diversi gli approcci seguiti e gli intenti perseguiti dai governi ed in particolare da Cina e Stati Uniti.
- L'UE sta affrontando le sfide regolamentari associate ai sistemi di IA, con l'ambizione, attraverso l'AI Act, di definire un set di regole in grado di assurgere a modello per lo sviluppo e le applicazioni di IA nel contesto globale.
- Al fine di incentivare il *public procurement* delle moderne soluzioni di IA sono state promosse importanti iniziative normative sia a livello internazionale, eurounitario che nazionale.
- Accanto agli incentivi previsti per il settore pubblico, per migliorare e rendere più efficienti i processi che aumentano la competitività di un'organizzazione e dell'intero sistema Paese, anche in ambito privato è necessario delineare strumenti idonei a promuovere l'approvvigionamento di sistemi e servizi di intelligenza artificiale.

SOMMARIO

SOMMARIO 2

EXECUTIVE SUMMARY 3

PRIMA PARTE: REGOLAMENTAZIONE E DIALOGO TRA GOVERNI IN TEMA DI INTELLIGENZA ARTIFICIALE 13

1. LINEE GUIDA E DOCUMENTI PROGRAMMATICI IN MATERIA DI IA DEI PRINCIPALI ORGANISMI INTERNAZIONALI 13

1.1. Road to Global Digital Compact 2024: il primo rapporto dell'AI Advisory Body ONU su un quadro globale di governance dell'IA 13

1.2. OCSE: dai principi del 2019 alle prime considerazioni sull'IA generativa 16

1.3. Il dialogo tra governi: i Principi guida e il Codice di condotta internazionale per i sistemi avanzati di IA del G7 nell'ambito dell'Hiroshima Process 21

1.4. Una strategia comune per l'IA: la Dichiarazione di Bletchley Park 24

1.5. World Bank: opportunità e rischi dell'IA nel settore pubblico 26

1.6. OMS: dai principi etici e di governance del 2021 agli aspetti regolatori dell'IA per la sanità 30

1.7. Analisi dei punti di contatto e di contrasto delle iniziative in materia di IA a livello internazionale 34

2. CONSIGLIO D'EUROPA: VERSO LA PRIMA CONVENZIONE INTERNAZIONALE IN MATERIA DI IA 37

3. STATI UNITI E CINA: IL DIVERSO APPROCCIO ALL'IA SULLE DUE SPONDE DEL PACIFICO 41

4. UNIONE EUROPEA: L'ANALISI DELLA PRIMA REGOLAMENTAZIONE ORGANICA AL MONDO SULL'INTELLIGENZA ARTIFICIALE 52

5. ITALIA: VERSO LA NUOVA STRATEGIA NAZIONALE PER L'INTELLIGENZA ARTIFICIALE 57

SECONDA PARTE: L'ADOZIONE E LA DIFFUSIONE DELL'IA NEL SETTORE PUBBLICO E PRIVATO 60

6. IL MERCATO DELL'INTELLIGENZA ARTIFICIALE: STATUS QUO E PROSPETTIVE FUTURE 60

7. LE APPLICAZIONI IA A SUPPORTO DELLE IMPRESE E DELLA PA 64

7.1. Il livello di adozione dell'IA nei vari settori economici e la diffusione delle tecnologie intelligenti nelle imprese italiane ed europee 64

7.2. Gli impatti dell'adozione delle tecnologie IA sulla pubblica amministrazione 70

7.3. Public procurement: dalle linee guida internazionali alla proposta di clausole contrattuali tipo della Commissione europea e lo stato dell'arte in Italia 77

7.4. Gli incentivi per l'adozione dell'IA nel settore privato 82

8. ANALISI DELLE SFIDE DA AFFRONTARE 87

CONCLUSIONI E SPUNTI DI POLICY 93

EXECUTIVE SUMMARY

I progressi nel campo dell'intelligenza artificiale hanno condotto governi e organizzazioni sovranazionali a interessarsi sempre di più di questa materia. Pertanto, la prima parte del presente studio si focalizza dapprima sull'analisi delle principali linee guida e documenti programmatici sull'IA che si sono susseguiti in questi ultimi anni, per poi soffermarsi sul confronto tra Stati Uniti e Cina in merito alla regolazione dell'IA, sino a esaminare i recenti sviluppi dell'Unione Europea, con un focus sull'AI Act. Infine, si fornirà una panoramica delle iniziative più rilevanti in ambito nazionale, in vista della nuova strategia nazionale sull'intelligenza artificiale.

Lo scorso 21 dicembre è stato raggiunto il primo step dell'*AI Advisory Body* dell'ONU con la pubblicazione dell'*Interim Report: Governing AI for the Humanity,* il quale sarà in fase di consultazione fino al prossimo 31 marzo al fine di raccogliere spunti utili per la versione finale del rapporto – prevista entro il 31 agosto 2024 – che conterrà raccomandazioni specifiche sulle funzioni, la forma e i termini per l'istituzione di una nuova agenzia internazionale che si occupi della governance globale dell'intelligenza artificiale, conformemente ad alcune direttrici tracciate già nell'*Interim Report* (inclusività, interesse pubblico, *data governance*, partecipazione, diritto internazionale).

Un punto di riferimento è rappresentato dalla Raccomandazione sui Principi dell'IA emanata dall'OCSE nel 2019 ed emendata lo scorso 8 novembre per aggiornare la nozione di "*AI system*". Lo scopo dichiarato è quello di promuovere un utilizzo dell'IA che sia innovativo, sicuro e affidabile, rispettoso dei diritti umani e dei valori democratici, adottando un approccio pragmatico e flessibile affinché le regole e le raccomandazioni ivi contenute resistano al passare del tempo. Allo scopo di evitare che tali Principi rimanessero su un piano meramente astratto, a ottobre scorso è stato reso noto il report "*The state of implementation of the OECD AI principles four years on*", il quale offre una panoramica circa le strategie nazionali sull'IA, dei relativi organismi di supervisione, nonché dei quadri di monitoraggio e valutazione, evidenziando altresì esempi di politiche consolidate ed emergenti in materia, così da stimolare l'apprendimento incrociato tra i *policymakers*.

Dato il forte interesse dimostrato dall'OCSE sul tema, non sorprende che nell'ultimo anno si sia occupato anche di IA generativa, in particolare con il *working paper* "*Initial policy considerations for generative artificial intelligence*" del 18 settembre scorso, che accanto al riconoscimento delle molteplici opportunità, soprattutto per alcuni campi di applicazione (educazione, sanità, sviluppo di codice, industria creativa e motori di ricerca), evidenzia altresì alcune sfide critiche con riguardo al mercato del lavoro, alla tutela del diritto d'autore e alla possibilità di un uso improprio di tali sistemi per la creazione di contenuti manipolati che alimentino la disinformazione.

Dato il forte interesse dimostrato dall'OCSE sul tema, non sorprende che nell'ultimo anno si sia occupato anche di IA generativa, in particolare con il working paper "Initial policy considerations for generative artificial intelligence" del 18 settembre scorso

Con l'inizio del nuovo anno si è concluso il periodo di presidenza giapponese del G7, durante il quale si sono fatti importanti passi avanti verso la definizione di una governance dell'IA. Difatti, in tale contesto è stato istituito il Processo di Hiroshima sull'IA con lo scopo di promuovere un'intelligenza artificiale sicura e affidabile a livello globale, con una particolare attenzione verso i sistemi di IA più avanzati. In particolare, l'iniziativa si è concretizzata – in primo luogo – nella definizione di 11 Principi Guida rivolti a sviluppatori, distributori e utilizzatori di sistemi di IA avanzati, ivi compresi enti del mondo dell'accademia, società civile, settore privato e pubblico; in secondo luogo, sulla base dei predetti principi, è stato plasmato il Codice di condotta che – pur non essendo giuridicamente vincolante – è immediatamente applicabile e data l'autorevolezza del consesso da cui proviene, ispirandosi peraltro a quanto già stabilito in ambito OCSE, rappresenta un utile strumento per quelle organizzazioni che intendano dimostrare una maggiore attenzione dei sistemi di IA da loro sviluppati, distribuiti o comunque utilizzati.

Il Codice di condotta promosso dal G7 rappresenta un utile strumento per quelle organizzazioni che intendano dimostrare una maggiore attenzione dei sistemi di IA da loro sviluppati, distribuiti o comunque utilizzati

L'1 e il 2 novembre scorso il Regno Unito ha ospitato il primo *AI Safety Summit*, che ha riunito allo stesso tavolo rappresentanti governativi di 28 Paesi e dell'Unione Europea, alla presenza degli esponenti delle principali aziende del settore dell'intelligenza artificiale, al fine di discutere della direzione da intraprendere affinché l'umanità abbia a disposizione un'IA responsabile e sicura. Il prodotto di questa importante iniziativa politica si è rivisto nella cosiddetta "Dichiarazione di Bletchley Park", sottoscritta da tutti e 29 i partecipanti istituzionali (tra i quali Stati Uniti, Cina e Unione Europea). Un punto ricorrente nella Dichiarazione concerne il riconoscimento del fatto che la regolazione dell'IA può basarsi su approcci differenti dovuti alle circostanze nazionali e ai framework normativi applicabili. Per tale ragione, occorrerebbe richiedere agli attori che sviluppano IA una maggiore trasparenza di tali sistemi, nonché responsabilizzare gli stessi attraverso l'istituzione di elevati standard di sicurezza e di valutazione dei rischi volti alla tutela dei relativi utilizzatori, puntando prioritariamente sulle misure tecnologiche e poi sulla definizione di regole giuridiche, oltre che sulla collaborazione e la cooperazione multi-stakeholder.

Un punto ricorrente nella Dichiarazione di Bletchley Park concerne il riconoscimento del fatto che la regolazione dell'IA può basarsi su approcci differenti dovuti alle circostanze nazionali e ai framework normativi applicabili

Tra gli organismi internazionali che hanno fornito il loro contributo relativamente alla costruzione di una governance in materia di intelligenza artificiale vi è anche la Banca Mondiale, la quale nel 2020 ha reso pubblico un importante report ("*Artificial Intelligence in the Public Sector. Maximizing Opportunities, Managing Risks*"), il cui titolo esplicita sin da subito il focus del documento: massimizzare le opportunità (ad esempio, per contrastare la corruzione, sostenere il settore sanitario e giudiziario, nonché favorire il coinvolgimento dei cittadini, il *public procurement*

e le attività di audit) e, al contempo, gestire i rischi derivanti dall'uso dell'IA nel settore pubblico (performance, fiducia e bias; cybersicurezza; controllo; privacy). In particolare, lo scopo è quello di stimolare la modernizzazione del settore pubblico attraverso l'intelligenza artificiale e le sue applicazioni, suggerendo due priorità: da un lato, l'adozione di politiche e quadri di governance che promuovano, al contempo, un'IA umano-centrica e la massimizzazione delle relative opportunità; dall'altro, gli investimenti in capitale umano e infrastrutture digitali.
Altro contributo di rilievo sulla definizione di principi etici e di governance applicabili all'intelligenza artificiale è stato fornito dall'Organizzazione Mondiale di Sanità (OMS). Infatti, a giugno 2021 è stato pubblicato un report ("*Ethics and Governance of Artificial Intelligence for Health*") contenente linee guida e raccomandazioni in merito allo sviluppo, all'applicazione e all'uso delle tecnologie di IA nel settore sanitario, con particolare riferimento ai seguenti ambiti: ricerca sanitaria, sviluppo di farmaci, gestione e pianificazione dei sistemi sanitari e monitoraggio della salute pubblica. Inoltre, lo scorso 18 ottobre la stessa OMS ha pubblicato un nuovo report denominato "*Regulatory considerations on artificial intelligence for health*" in risposta alla crescente esigenza dei Paesi di gestire in modo responsabile le tecnologie di IA in sanità, focalizzandosi su 6 aree di interesse: documentazione e trasparenza; gestione del rischio; validazione dei dati; qualità dei dati; privacy e *data protection*; coinvolgimento e collaborazione.

In ultimo, ci si è soffermati sull'analisi dei punti di contatto e di contrasto delle iniziative in materia di IA a livello internazionale. Sul punto, è emerso un diffuso consenso sui principi etici e di sicurezza da applicare all'IA, oltre che in termini di cooperazione internazionale e collaborazione multi-stakeholder. Fa ben sperare, peraltro, che vi sia quantomeno unità di intenti nel riconoscere l'importante legame tra governance dell'IA e governance dei dati, nonché la centralità degli investimenti per accompagnare lo sviluppo delle tecnologie di IA con opportune competenze e infrastrutture necessarie al suo impiego a beneficio di tutti. D'altra parte, sono emersi alcuni punti di divergenza, con particolare riferimento ai modelli *open-source*, all'eventuale rafforzamento della regolamentazione in materia di IA, nonché a una maggiore attenzione sulla sostenibilità ambientale ed energetica dell'IA, oltre che in merito all'identificazione di criteri e meccanismi per l'individuazione della responsabilità in caso di danni causati da decisioni basate su algoritmi.
Inoltre, nelle linee guida e nei documenti programmatici analizzati non si registrano considerazioni di merito rispetto all'adesione a principi e valori democratici nello sviluppo, nella distribuzione e nell'utilizzo di soluzioni di IA. Questo aspetto è invece centrale nella bozza della prima convenzione internazionale sull'intelligenza artificiale, attualmente in discussione presso il Comitato sull'IA (CAI) del Consiglio d'Europa, il cui iter dovrebbe concludersi entro il mese di maggio.

Nelle linee guida e nei documenti programmatici analizzati non si registrano considerazioni di merito rispetto all'adesione a principi e valori democratici nello sviluppo, nella distribuzione e nell'utilizzo di soluzioni di IA. Questo aspetto è invece centrale nella bozza della prima convenzione internazionale sull'intelligenza artificiale, attualmente in discussione presso il Comitato sull'IA (CAI) del Consiglio d'Europa, il cui iter dovrebbe concludersi entro il mese di maggio

Sebbene sia globale l'interesse per l'IA e ampiamente diffusa la consapevolezza circa la necessità di comprendere e governare un fenomeno a così alto impatto e complessità, persistono importanti differenze negli approcci seguiti e gli intenti perseguiti dai governi ed in particolare da Cina e Stati Uniti. Se, infatti, la Cina, pur con la specifica attenzione rivolta a scongiurare rischi di conflitti tra sistemi di IA e l'ordine politico e sociale costituito dettate dal regime governante che la pone in una situazione peculiare rispetto ai principi affermati a livello internazionale in materia di IA, rappresenta uno dei primi paesi ad avere adottato una specifica legislazione in materia, gli USA al contrario, continuano, nonostante le evoluzioni in atto di cui si darà conto nel corso dell'analisi, a non disporre di una disciplina organica dell'IA e a conservare un approccio più incentrato sull'assunzione di impegni volontari da parte delle grandi aziende tecnologiche nel settore dell'IA attraverso il quale si punta a non frenare l'innovazione. La Cina, in particolare, è stata tra i primi paesi al mondo ad affrontare la sfida della regolamentazione con riferimento all'IA nel tentativo di indirizzarne e governarne lo sviluppo, sia pure attraverso una disciplina frammentata in diversi atti normativi, tra i quali in particolare l'Algorithm Recommendation Regulation, entrato in vigore il 1° marzo 2022, il Deep Synthesis Regulation entrato in vigore il 10 gennaio 2023 ed il Regolamento sull'IA generativa entrato in vigore lo scorso 15 agosto 2023.

La Cina è stata tra i primi paesi al mondo ad affrontare la sfida della regolamentazione con riferimento all'IA nel tentativo di indirizzarne e governarne lo sviluppo, sia pure attraverso una disciplina frammentata in diversi atti normativi, tra i quali in particolare l'Algorithm Recommendation Regulation, entrato in vigore il 1° marzo 2022, il Deep Synthesis Regulation entrato in vigore il 10 gennaio 2023 ed il Regolamento sull'IA generativa entrato in vigore lo scorso 15 agosto 2023

Il primo, in particolare, fissa principi ed obblighi da rispettare in relazione agli algoritmi di raccomandazione istituendo anche uno strumento di controllo particolarmente potente, ossia il registro degli algoritmi; il secondo, detta le regole sulla gestione dei dati e delle tecnologie di sintesi profonda, esortando i fornitori di servizi di sintesi profonda e i promotori di tale tecnologia a garantire il rispetto delle leggi e dei regolamenti prescrivendo una serie di obblighi e responsabilità in alcuni ambiti chiave. Le linee guida, infine, individuano una serie di principi e di procedure da seguire rispetto all'IA generativa nell'intento di promuovere lo sviluppo sano e l'uso regolamentato dell'IA generativa, preservare la sicurezza nazionale e l'interesse pubblico della società e proteggere i diritti e gli interessi legittimi di cittadini, persone giuridiche e altre organizzazioni.

Per quanto concerne invece gli USA, seppur con un approccio in chiara evoluzione, l'approccio tradizionalmente seguito è stato quello di limitare l'adozione di interventi prescrittivi favorendo, al contrario, l'adozione di modelli di sviluppo e condotta volontariamente accettati dalle aziende attive nel settore dell'IA nell'intenzione di non ostacolare lo sviluppo tecnologico e l'innovazione e di rendere l'America leader mondiale nella ricerca, lo sviluppo e l''utilizzo di sistemi di IA. Ed infatti, dopo l'adozione, nel 2020, del National AI Initiative Act (NAIIA), relativo al periodo 2020-

2025, in cui è evidenziato l'enorme potenziale innovativo dell'IA per ogni settore socio-economico ed affermata l'ambizione del Governo federale di giocare un ruolo centrale nella ricerca, nello sviluppo e nelle attività formative in materia di IA attraverso il coordinamento e la collaborazione tra governo, accademia e settore privato, nel 2023, prima a luglio e poi a settembre, è stata annunciata dall'amministrazione Biden l'assunzione di impegni volontari da parte di complessive 15 grandi aziende a contribuire allo sviluppo sicuro, protetto e trasparente dell'IA. Si tratta di una serie corposa di impegni tesi a garantire la sicurezza dei prodotti e la capacità degli stessi di catturare la fiducia del pubblico. Da ultimo, il 30 ottobre scorso, il Presidente Biden ha adottato un Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, nel quale, partendo dalla constatazione delle straordinarie opportunità offerte dai sistemi di IA ma anche dei potenziali rischi connessi ad "irresponsible use" degli stessi e della volontà di governare lo sviluppo e l'utilizzo dell'IA, sono declinate 8 priorità e principi guida da osservare nello sviluppo ed utilizzo dei sistemi di IA e previsti adempimenti di importanza crescente a seconda del bene che si intende tutelare (in primis la sicurezza nazionale) a carico delle imprese che, di fatto, sembrano segnare l'inizio di un cambiamento rispetto all'approccio, volutamente poco restrittivo ed affatto prescrittivo, fino ad oggi seguito dagli USA. Lo scorso 29 gennaio è stato reso pubblico dalla Casa Bianca il completamento delle attività richieste dall'order alle varie agenzie ed autorità mentre il 7 febbraio è stata annunciata la creazione dell'AI Safety Institute Consortium (AISIC), composto da oltre 200 stakeholders provenienti dal mondo della politica, della ricerca, dell'industria, dell'università e della società civile, che opererà sotto l'AI Safety Institute (USAISI) al fine di contribuire attivamente al raggiungimento degli obiettivi fissati dall'order con specifico riguardo alla fissazione di standard e all'individuazione di strumenti di mitigazione dei rischi dell'IA.

Accanto a Cina e Stati Uniti, anche l'UE sta affrontando le sfide regolamentari associate ai sistemi di IA, con l'ambizione, così come accaduto attraverso il GDPR per la tutela dei dati personali, di definire un set di regole in grado di assurgere a modello per lo sviluppo e le applicazioni di IA nel contesto globale. A tal fine, il 21 aprile 2021, la Commissione ha lanciato una proposta per un "regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate in materia di intelligenza artificiale e che modifica alcuni atti legislativi dell'Unione" (AI Act), teso ad istituire un quadro di riferimento legale volto a normare il mercato dell'UE dell'IA che sta giungendo al traguardo, dopo il raggiungimento di un'intesa da parte di Consiglio e Parlamento lo scorso 8 dicembre. Si tratta di un'iniziativa straordinariamente importante che, seppur con le difficoltà derivanti dal disciplinare un fenomeno in continua e rapida evoluzione e dunque con la necessità di predisporre una disciplina *future oriented,* adotta un approccio fondato sul rischio che distingue tra usi dell'IA che creano un rischio inaccettabile, un rischio elevato ed un rischio basso o minimo, da cui discendono obblighi diversificati e dal rigore crescente. Nello specifico, è previsto un divieto per le pratiche considerate inaccettabili in quanto contrarie ai valori dell'Unione, ad esempio perché violano i diritti fondamentali (ad es. pratiche manipolative dei minori o dei disabili o che prevedono l'uso di tecniche subliminali che sfruttano l'inconsapevolezza degli individui etc.), mentre per i sistemi di IA ad alto rischio, il regolamento distingue le principali tipologie di sistemi che rientrano in tale categoria, individua i criteri da seguire per valutare se un sistema di IA presenta alti rischi e fissa una serie di requisiti obbligatori oltre a subordinare l'accesso al mercato europeo di tali sistemi ad una valutazione della conformità ex ante secondo procedure dettagliatamente descritte.

Accanto a Cina e Stati Uniti, anche l'UE sta affrontando le sfide regolamentari associate ai sistemi di IA, con l'ambizione, così come accaduto attraverso il GDPR per la tutela dei dati personali, di definire un set di regole in grado di assurgere a modello per lo sviluppo e le applicazioni di IA nel contesto globale

Mentre l'UE si trova ad affrontare la sfida della regolamentazione con l'obiettivo di diventare leader nel settore, anche l'Italia cerca di giocare la propria partita per non restare indietro e mettere a sistema, nel rispetto dei diritti fondamentali degli individui, gli enormi benefici che l'IA assicura. Dopo che i precedenti tentativi si erano conclusi con un nulla di fatto, nel novembre 2021 è stato approvato dal Governo il Programma Strategico per l'Intelligenza Artificiale (IA) 2022-2024, che, seppur con un orizzonte temporale decisamente limitato e senza indicazione specifiche delle risorse destinate a ciascuna iniziativa, ha delineato 6 obiettivi, 11 settori prioritari, 3 aree di intervento e 24 politiche da implementare nei successivi tre anni al fine di rafforzare il sistema IA in Italia e renderla un esempio di eccellenza, attraverso specifiche iniziative mirate alla creazione ed il potenziamento delle competenze, della ricerca, dei programmi di sviluppo, delle applicazioni ed incentivando il trasferimento tecnologico.
Mentre scriviamo, il Governo sta lavorando a una nuova strategia che dovrebbe contenere, tra l'altro, due iniziative particolarmente importanti: da un lato, la costituzione di un fondo pubblico-privato di venture capital, sotto l'egida di CDP, per favorire la crescita delle startup innovative italiane; dall'altro, l'istituzione di una Fondazione, che dovrebbe chiamarsi AI4Industry, dotata di un budget finanziato dallo stato di venti milioni di euro l'anno.

Anche il Parlamento è attivo su questi temi, con diverse indagini in corso. Inoltre, lo scorso 19 ottobre è stato presentato dal PD un disegno di legge che persegue il fine di rendere chiaramente riconoscibili agli utenti i contenuti editoriali, compresi testi, video, immagini e voci, che sono creati, generati o sintetizzati, in tutto o in parte, da sistemi basati su intelligenza artificiale (ivi compresi algoritmi di apprendimento automatizzato, cosiddetto machine learning e reti neurali artificiali).

La seconda parte del presente studio si focalizza sugli impatti economici dell'intelligenza artificiale, andando in primis ad analizzare l'andamento del mercato. Successivamente, si concentra sull'adozione e sulla diffusione delle tecnologie IA nel settore pubblico e privato. In seguito, si esamina lo stato dell'arte a livello internazionale, eurounitario e nazionale sul *procurement* di soluzioni di IA nel settore pubblico, per poi proseguire con gli incentivi a sostegno delle imprese. Infine, come ultima disamina, l'ultimo capitolo analizza le sfide ed i potenziali rischi da affrontare.
Recenti stime prevedono che le dimensioni del mercato mondiale dell'intelligenza artificiale raggiungeranno i $305,9 miliardi nel 2024. Il mercato IA continuerà poi la sua ascesa, trainato principalmente dalle applicazioni di machine learning e toccherà i $738,8 miliardi nel 2030, mostrando un tasso di crescita medio annuo (CAGR) nel periodo 2024-2030 del 15,83%.
Dunque, si tratta di un mercato in continua espansione, le cui dinamiche positive sono destinate a durare anche nei prossimi anni grazie all'avvento dell'IA generativa, ovvero un nuovo campo di ricerca che utilizza tecniche di machine Learning e deep Learning per generare nuovi dati, tra cui immagini, musica e testo, che non esistevano in precedenza.

L'IA generativa rappresenta già oggi una fetta rilevante del mercato IA. Nel 2023, stando ad alcune stime, ha coperto il 19% del mercato IA totale ed entro il 2024 si prevede un aumento dell'incidenza di 3 punti percentuali. Nel 2030, il mercato dell'IA generativa dovrebbe pesare sull'intero mercato IA per il 28%.

L'IA generativa rappresenta già oggi una fetta rilevante del mercato IA. Nel 2023, stando ad alcune stime, ha coperto il 19% del mercato IA totale ed entro il 2024 si prevede un aumento dell'incidenza di 3 punti percentuali. Nel 2030, il mercato dell'IA generativa dovrebbe pesare sull'intero mercato IA per il 28%

Sono sempre di più le imprese di quasi tutti i settori economici che, rispetto al passato, utilizzano le tecnologie IA, compresa l'IA generativa, per svolgere diverse funzioni.
Secondo i dati riportati dalla Stanford University, l'IA è ampiamente utilizzata per le operazioni di servizio, la strategia e la finanza aziendale, e le attività di risk management e quasi tutti i settori economici presi in considerazione riportano circa il 20% di utilizzo dell'IA in queste tre funzioni. Il maggiore utilizzo dell'IA nello sviluppo di prodotti/servizi si riscontra, invece, nel settore dei servizi finanziari, in cui raggiunge il 30%. Infine, nel settore High Tech/Telecomunicazioni, l'IA è ampiamente utilizzata per le attività di risk management, dove la percentuale di adozione si attesta sul 38%.
Inoltre, un recentissimo sondaggio di McKinsey, condotto a livello globale evidenzia che il 33% degli intervistati del settore tecnologico, dei media e delle telecomunicazioni utilizza regolarmente l'IA generativa per lavoro o al di fuori del lavoro, mentre il 37% degli intervistati dello stesso settore ha dichiarato di aver utilizzato questa tecnologia almeno una volta. Altri due settori che usano questi nuovi strumenti sono i servizi finanziari e i servizi aziendali, legali e professionali in cui quasi un quarto degli intervistati utilizza regolarmente l'IA generativa.

Nonostante a livello globale si riscontri un incremento dell'adozione dell'IA da parte delle imprese, a livello europeo c'è ancora della strada da compiere. Infatti, nel 2023, solo l'8% delle imprese UE ha adottato almeno una tra le tecnologie IA più comuni. Il tasso di adozione più elevato si registra in Danimarca, dove il 15% circa delle imprese fa uso di almeno una tecnologia IA. In fondo alla classifica si colloca la Romania, che presenta il livello più basso di adozione IA con solo l'1,5% di imprese che utilizzano almeno una tecnologia. L'Italia si posiziona al di sotto della media UE, con il 5% delle imprese che ha adottato almeno una tra le tecnologie IA a disposizione.
Nello specifico, tra le diverse tecnologie IA, l'Italia presenta la performance migliore nella robotica di servizio, unico campo dove si colloca nettamente al di sopra della media UE, con oltre il 4% delle imprese che ha adottato robot di servizio. Relativamente alle altre tecnologie, tra cui (text mining, *speech recognition*, machine learning) l'adozione da parte delle imprese italiane è sempre al di sotto della media europea.

Tra le diverse tecnologie IA, l'Italia presenta la performance migliore nella robotica di servizio, unico campo dove si colloca nettamente al di sopra della media UE, con oltre il 4% delle imprese che ha adottato robot di servizio

Tra i settori economici che in Italia utilizzano maggiormente le tecnologie IA spicca quello dell'informatica dove il 23,6% delle imprese utilizza almeno uno tra i software o sistemi di IA più comuni: text mining; riconoscimento vocale, generazione del linguaggio naturale; riconoscimento, elaborazione delle immagini; machine learning, deep learning, reti neurali; software robot; robot o droni autonomi. Seguono poi il settore delle telecomunicazioni e della produzione cinematografica dove, rispettivamente, il 13,3% e l'11% circa delle imprese ha adottato l'IA. Tra le imprese italiane che utilizzano l'IA, le tecnologie più comuni riguardano l'automatizzazione di flussi di lavoro attraverso software robot (40,1%), le applicazioni di text mining (39,3%) e il riconoscimento vocale (31%).

Tra le imprese italiane che utilizzano l'IA, le tecnologie più comuni riguardano l'automatizzazione di flussi di lavoro attraverso software robot (40,1%), le applicazioni di text mining (39,3%) e il riconoscimento vocale (31%)

Infine, gli ambiti aziendali in cui vengono più spesso adottati sistemi di intelligenza artificiale sono quelli relativi a processi di produzione, ad esempio per la manutenzione predittiva o il controllo qualità della produzione, in cui il 39% delle imprese italiane fa uso di questi strumenti (nel caso del settore manifatturiero si raggiunge il 52,5% delle imprese), alla funzione di marketing o vendite, ad esempio per funzioni di assistenza ai clienti o campagne promozionali personalizzate (33,1% per il totale delle attività economiche mentre si raggiunge il 41,3% nel settore dei servizi), alla sicurezza informatica (23,7%, al 50,6% nel settore dell'energia) e alle attività di ricerca e sviluppo (R&S) o innovazione per analizzare dati, sviluppare un prodotto/servizio nuovo o significativamente migliorato (21,1%).

Come negli altri settori economici, anche nell'ambito del settore pubblico, l'adozione di soluzioni basate sull'IA potrebbe portare a vantaggi significativi. Numerose amministrazioni, oltre a sondare le opportunità tramite progetti pilota, hanno avviato l'implementazione di questa tecnologia in svariati contesti e utilizzano l'IA nello svolgimento di diverse operazioni quotidiane. A tal proposito, esistono studi che sottolineano come le soluzioni basate sull'IA nella gestione e nell'erogazione dei servizi pubblici, laddove implementate, stanno già offrendo notevoli vantaggi e valore pubblico ai cittadini.

In Europa, i casi di impiego dell'intelligenza artificiale nel settore pubblico sono in costante crescita dal 2010, con il numero più elevato di casi che si registra nei Paesi Bassi (124 casi registrati), seguito dall'Italia (75 casi registrati) e Portogallo (60 casi registrati). In Italia, notiamo che un numero considerevole di casi (il 29%), è costituito da iniziative lanciate a livello regionale, seguite

da iniziate nazionali (28%), iniziative che coinvolgono più paesi europei (il 24%) e iniziative lanciate a livello locale (19%).

In Europa, i casi di impiego dell'intelligenza artificiale nel settore pubblico sono in costante crescita dal 2010, con il numero più elevato di casi che si registra nei Paesi Bassi (124 casi registrati), seguito dall'Italia (75 casi registrati) e Portogallo (60 casi registrati). In Italia, notiamo che un numero considerevole di casi (il 29%), è costituito da iniziative lanciate a livello regionale, seguite da iniziate nazionali (28%), iniziative che coinvolgono più paesi europei (il 24%) e iniziative lanciate a livello locale (19%)

Dunque, mentre a livello europeo lo sviluppo dell'IA nel settore pubblico sembra essere guidato principalmente dai governi nazionali, in Italia notiamo un notevole coinvolgimento delle amministrazioni regionali e locali nel promuovere lo sviluppo e l'uso delle soluzioni di intelligenza artificiale. Per quanto concerne lo specifico settore di impiego, le applicazioni di IA sono usate principalmente nell'ambito dei servizi pubblici (37%), troviamo in seguito il settore sanitario (27%) e quello dell'ordine pubblico e della sicurezza (12%), mentre l'utilizzo dell'IA sembra essere meno diffuso nei settori della formazione scolastica, della protezione ambientale, della cultura e religione, e della protezione sociale. Guardando agli aspetti tecnologici, il 25% dei casi di impiego di IA in Italia si riferisce all'utilizzo del Machine Learning. Tra i casi di impiego di IA più interessanti in Italia e che fanno uso del Machine Learning, vi è sicuramente quello del progetto "Giurisprudenza Predittiva" del Tribunale di Pisa. A livello nazionale, inoltre, merita di essere menzionato il progetto dell'INPS per la classificazione e smistamento automatico delle PEC, premiato nella Top 10 mondiale dell'Intelligenza Artificiale di IRCAI UNESCO.

Al fine di incentivare il *public procurement* dei moderni sistemi di IA sono state promosse diverse iniziative normative. In virtù del fatto che tale pratica risulta spesso sfavorita a causa della particolare cautela che molte istituzioni pubbliche pongono nella scelta di impiegare questa nuova tecnologia, una prima forma di incentivo si rivede nelle linee guida internazionali per l'approvvigionamento dell'IA, sviluppate nel 2019 dal World Economic Forum. L'obiettivo è quello di esortare i governi all'elaborazione di considerazioni fondamentali prima dell'acquisizione e della distribuzione delle soluzioni e dei servizi di IA, sostenendo tutti i soggetti coinvolti in tale procedura.

Uno sviluppo importante in questo senso è arrivato dalla pubblicazione nel settembre scorso, nell'ambito della piattaforma "*Procurement of AI community*", presente sul sito web della Commissione Europea, di due proposte di clausole standard per l'acquisto da parte delle pubbliche amministrazioni di sistemi di intelligenza artificiale. Le clausole si ispirano a quelle già elaborate dalla città di Amsterdam nel 2018 e sono suddivise in due documenti che si distinguono in base all'acquisizione delle due diverse categorie di sistemi di intelligenza artificiale: quelli "ad alto rischio" e quelli che non presentano un rischio elevato.

Il 12 febbraio 2024 è stato pubblicato, ad opera dell'Agenzia per l'Italia Digitale, il Piano Triennale per l'Informatica nella Pubblica Amministrazione 2024-2026 che vede delineati al suo interno alcuni principi generali che le pubbliche amministrazioni dovranno adottare e applicare in risposta alle sfide derivanti dall'impiego dell'IA.

Il 12 febbraio 2024 è stato pubblicato, ad opera dell'Agenzia per l'Italia Digitale, il Piano Triennale per l'Informatica nella Pubblica Amministrazione 2024-2026 che vede delineati al suo interno alcuni principi generali che le pubbliche amministrazioni dovranno adottare e applicare in risposta alle sfide derivanti dall'impiego dell'IA

Anche nel settore privato è necessario delineare strumenti idonei a incentivare l'acquisizione di sistemi e servizi di intelligenza artificiale. La descrizione delle forme a sostegno della Transizione 4.0 affonda le radici nel 2016, anno in cui è stato lanciato il Piano Nazionale Industria 4.0. Questo, a seguito di numerose modifiche, è stato sostituito dal Piano Nazionale Impresa 4.0 (ad opera della legge di bilancio 2018) e poi dal Piano Transizione 4.0, che da ultimo è stato rimodulato ad opera del PNRR, mentre all'orizzonte vi è l'evoluzione verso il paradigma 5.0. Tuttavia, accanto alle misure previste o in via di definizione, occorre immaginarne delle altre, soprattutto sul fronte dell'orientamento tecnologico delle imprese e della formazione del personale ma anche della capacità del sistema di elevare l'asticella della ricerca e dell'innovazione, oggi decisamente più bassa rispetto ai Paesi più avanzati. In quest'ultimo senso, occorre sperimentare nuove forme di collaborazione pubblico-privato che consentano una condivisione di risorse e di esperienze. Che in ultima analisi lascino la tipica impresa italiana, per definizione piccola, meno sola (e più resiliente) di fronte alle raffiche più violente e veloci che in passato dell'innovazione.

Tante sono le opportunità legate allo sviluppo e all'adozione dell'intelligenza artificiale sia nel settore pubblico, sia privato. Nonostante il largo impiego delle tecnologie IA nelle imprese di tutto il mondo, permane la necessità di gestire i rischi legati a questa tecnologia. Secondo i dati rilevati da McKinsey[1], tra i principali rischi IA che le organizzazioni considerano rilevanti rientrano quelli legati alla sicurezza informatica. Oltre il 50% degli intervistati infatti ha timore di minacce informatiche correlate all'uso dell'IA. A questi poi si aggiungono i timori relativi alla privacy e alla sicurezza dei dati personali nonché rischi reputazionali. Accanto alla regolamentazione, appare importante il coinvolgimento attivo del settore privato per internalizzare nei processi di business un approccio appropriato, ispirato a principi di responsabilità e trasparenza e già testimoniato da una pluralità di iniziative. Relativamente al contesto italiano, nel 2023 una quota pari al 4,4% di imprese ha preso in considerazione l'utilizzo di tecnologie di IA ma non le ha ancora utilizzate per timore di alcuni rischi o per mancanza di competenze adeguate o costi troppo elevati.
In particolare, la mancanza di skill costituisce un ostacolo per oltre la metà delle imprese italiane che non utilizzano ancora l'IA ma vorrebbero farlo; i costi troppo elevati sono, invece, una preoccupazione del 49,6% di imprese. I timori, invece, relativi a violazione della privacy e alla protezione dei dati personali riguardano il 37,2% di imprese che vorrebbero utilizzare tali tecnologie se avessero più garanzie.

[1] *https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/*

PRIMA PARTE: REGOLAMENTAZIONE E DIALOGO TRA GOVERNI IN TEMA DI INTELLIGENZA ARTIFICIALE

1. LINEE GUIDA E DOCUMENTI PROGRAMMATICI IN MATERIA DI IA DEI PRINCIPALI ORGANISMI INTERNAZIONALI

1.1. Road to Global Digital Compact 2024: il primo rapporto dell'AI Advisory Body ONU su un quadro globale di governance dell'IA

L'essere umano si è mostrato interessato a creare forme di intelligenza non umana sin dall'antichità, a partire dal concetto di "automi" introdotto dalla civiltà egizia. Tuttavia, è solo dagli anni '40 del secolo scorso che si assiste agli atti fondativi di una nuova disciplina scientifica sino a giungere al celebre articolo "*Computing Machinery and Intelligence*" di Alan Turing che ha posto le basi per il dibattito sull'IA, introducendo anche quello che sarebbe diventato noto come il Test di Turing, ossia una metodologia per determinare se una macchina può essere considerata indistinguibile da un essere umano per capacità intellettive. Negli ultimi decenni si sono fatti importanti passi avanti, soprattutto grazie ai progressi nel campo dell'apprendimento automatico con lo sviluppo delle reti neurali profonde, il che consente all'IA di analizzare grandi quantità di dati. Naturalmente, ciò ha sollevato preoccupazioni relativamente all'etica, alla privacy, alla sicurezza e all'impatto sociale dell'IA, nonché con riguardo alle decisioni da essa intraprese, avviando un ampio dibattito, anche attraverso la costituzione di appositi forum di discussione a livello internazionale.

Ciò ha sollevato preoccupazioni relativamente all'etica, alla privacy, alla sicurezza e all'impatto sociale dell'IA, nonché con riguardo alle decisioni da essa intraprese, avviando un ampio dibattito, anche attraverso la costituzione di appositi forum di discussione a livello internazionale

Ciò premesso, lo scorso 26 ottobre è stato annunciato l'insediamento di un comitato consultivo di esperti sull'intelligenza artificiale in seno all'ONU (*AI Advisory Body*), composto da 39 membri provenienti da governi, imprese, società civile e accademia di ciascuno dei cinque continenti, avente come principale obiettivo quello di raccomandare regole condivise in merito a un modello di governance dell'IA a beneficio dell'intera umanità. Più nel dettaglio, per il primo anno di attività tale organismo è stato incaricato di preparare un primo report entro il 31 dicembre 2023 e un secondo, integrato con l'apporto degli stakeholders interessati, entro il 31 agosto 2024, che

confluirà nel *Global Digital Compact*[2], il quale verrà discusso durante il *Summit of the Future*[3] del 21-22 settembre 2024. Il risultato auspicato consiste nella predisposizione di raccomandazioni specifiche sulle funzioni, la forma e i termini per l'istituzione di una nuova agenzia internazionale che si occupi della governance dell'intelligenza artificiale.

Ebbene, lo scorso 21 dicembre è stato effettivamente raggiunto il primo step con la pubblicazione dell'*Interim Report: Governing AI for the Humanity*, il quale sarà in fase di consultazione sino al prossimo 31 marzo. Tale documento pone in rilievo le problematiche da affrontare e ne indica la metodologia, proponendo 5 principi guida e 7 funzioni istituzionali, che dovranno consentire – una volta giunti alla versione finale di agosto – di fornire soluzioni pratiche e condivise tra i membri dell'ONU su come delineare la futura governance dell'IA. Ciò poiché – precisa il report nella parte introduttiva – si tratta di un aspetto carente allo stato attuale, che è tuttavia cruciale affinché si possa trarre vantaggio dall'IA in maniera equamente distribuita a livello globale. Ad esempio, lo sviluppo dell'IA è basato sostanzialmente sulla disponibilità di tre elementi: dati, capacità computazionali e talenti, ai quali però hanno accesso solo alcuni Paesi (prevalentemente del nord del mondo) e grandi aziende tecnologiche, il che può condurre a una concentrazione preoccupante in mano a pochi soggetti.

Lo scorso 21 dicembre è stato effettivamente raggiunto il primo step con la pubblicazione dell'Interim Report: Governing AI for the Humanity, il quale sarà in fase di consultazione sino al prossimo 31 marzo

Pertanto, il documento in esame si focalizza anche sui fattori abilitanti a cogliere le numerose opportunità dell'IA, come la necessità di maggiori investimenti in infrastrutture di base (es: banda larga ed elettricità), talenti, dati e risorse computazionali, così come in capacità di *procurement* e framework regolatori, senza tralasciare un quadro di governance che consenta di gestire l'impatto dell'IA sull'intera società. Invece, viene sottolineato come sia ancora aperto il dibattito sul potenziale valore abilitante dell'*open-source* e di modelli aperti di condivisione di dati.

Proseguendo, l'Interim Report cerca di categorizzare i rischi esistenti e potenziali dei sistemi di IA, richiamando quelli relativi a: i) individui (manipolazione, inganno, pregiudizio, armi autonome, veicoli autonomi, diagnostica e neurotecnologia, polizia predittiva, riconoscimento biometrico, educazione, lavoro, ecc.); ii) gruppi (discriminazione e marginalizzazione di comunità e sotto-gruppi in base al genere, l'età e caratteristiche fisiche); iii) società nel suo complesso (disinformazione, elezioni, deepfake, usi militari e di polizia, ecc.); iv) economia (concentrazione di potere, dipendenza tecnologica, distribuzione di risorse, ecc.); v) ecosistemi (stabilità dei sistemi finanziari, rischi per le infrastrutture critiche, ecc.); vi) valori e norme di vario tipo. Sul punto, il documento riconosce come non vi sia ancora un consenso sufficiente su come valutare e contrastare questi rischi, richiamando pertanto la necessità di maggiore coordinamento e

[2] Si tratta di un accordo ufficiale proposto nell'ambito del report "Our common agenda" (2021) del Segretario Generale dell'ONU António Guterres che, partendo da una consultazione multi-stakeholder continua e quanto più ampia e inclusiva possibile, mira a stabilire dei principi per un futuro digitale aperto, libero e sicuro per tutti.

[3] L'obiettivo dichiarato del Summit è duplice: i) accelerare gli sforzi per rispettare gli impegni internazionali esistenti; ii) adottare misure concrete per rispondere alle sfide e alle opportunità emergenti. Tale obiettivo si intende raggiungere attraverso l'approvazione di un documento di output ("Pact for the Future") a seguito del Summit del Futuro di settembre.

cooperazione a livello globale, in particolar modo attraverso il ruolo delle organizzazioni internazionali.

Inoltre, il report ONU, sebbene riconosca e veda con favore le iniziative già esistenti sulla governance dell'IA[4], in particolar modo nel momento in cui convergono su scelte linguistiche e/o contenutistiche, evidenzia altresì un disallineamento globale circa l'implementazione delle stesse, sia in termini di interoperabilità tra le varie giurisdizioni, sia con riguardo a forme di incentivo all'interno delle medesime. A complicare tale quadro – prosegue il documento – contribuiscono una serie di elementi, tra cui la mancanza di standard e parametri di riferimento circa la gestione dei rischi, nonché la sussistenza di molteplici definizioni di IA, col risultato che, nella maggior parte delle giurisdizioni, la governance dell'IA si traduce in meccanismi di auto-regolazione da parte degli stessi sviluppatori, distributori e utilizzatori dei sistemi di intelligenza artificiale, non incoraggiando nel lungo termine l'inclusione di diverse parti interessate, soprattutto nel Sud del globo.

Sulla scorta di tali premesse, il comitato consultivo dell'ONU ha elaborato i seguenti principi guida riferiti alle caratteristiche che dovrebbe avere la futura governance dell'IA:

1. Inclusività: i meccanismi di governance dovrebbero promuovere sia un'equa partecipazione nello sviluppo, distribuzione e uso dell'IA, sia nella governance stessa;
2. Interesse pubblico: l'IA dovrebbe essere governata nell'interesse pubblico, per cui le politiche pubbliche dovrebbero tendere a raggiungere obiettivi relativi all'equità, l'inclusione, la sostenibilità, il benessere individuale e collettivo, la competitività dei mercati e un sano ambiente per l'innovazione;
3. *Data governance*: la governance dell'IA e quella dei dati dovrebbero andare di pari passo, in modo che i framework legali e regolatori siano funzionali a tutelare la privacy e la sicurezza dei dati personali e di quei dati pubblici rilevanti per affrontare le sfide della società;
4. Partecipazione: l'IA dovrebbe essere universale, interconnessa e radicata in una collaborazione multi-stakeholder nell'ambito di istituzioni esistenti o nuove, che siano dotate di strutture per coinvolgere concretamente i governi, le imprese e la società civile;
5. Diritto internazionale: la governance dell'IA dovrebbe essere ancorata alla Carta delle Nazioni Unite, al Diritto Internazionale dei Diritti Umani e ad altri impegni internazionali concordati, come gli Obiettivi di Sviluppo Sostenibile dell'ONU.

Dalla visione che emerge dall'Interim Report, la futura governance dell'IA dovrebbe occuparsi quantomeno di 7 funzioni istituzionali (e 15 sotto-funzioni) di rigidità crescente:

I. Valutare regolarmente le direzioni e le implicazioni future dell'IA, effettuando uno *scanning* dei rischi e opportunità, al fine di informare e guidare l'attività dei *policymaker*;
II. Rafforzare l'interoperabilità delle iniziative di governance emergenti, in conformità con le norme internazionali, tramite un Quadro di Governance Globale dell'IA emanato da un consesso internazionale;

[4] *Si v. infra, par. 1.2 ss.*

III. Sviluppare e armonizzare gli standard, la sicurezza e i framework di gestione dei rischi, con un particolare focus su nuovi indicatori per misurare e tracciare l'impatto ambientale dell'IA e delle risorse energetiche e naturali di cui necessita per funzionare;
IV. Agevolare lo sviluppo, la diffusione e l'uso dell'IA a vantaggio dell'economia e della società attraverso la cooperazione internazionale;
V. Promuovere la collaborazione internazionale per lo sviluppo di talenti, capacità e infrastrutture computazionali, dataset di alta qualità, condivisione di modelli *open-source* responsabili e strumenti pubblici abilitati dall'IA per raggiungere gli Obiettivi di Sviluppo Sostenibile;
VI. Monitorare i rischi, segnalare gli incidenti e coordinare la risposta alle emergenze derivanti da un uso improprio dell'IA (es: accesso ad armi di distruzione di massa, maggiori rischi di attacchi verso infrastrutture critiche, rapida diffusione di informazioni dannose per i mercati e le istituzioni pubbliche);
VII. Curare la *compliance* e l'*enforcement* rispetto a norme giuridicamente vincolanti, anche in combinazione con norme non vincolanti.

Nella parte conclusiva, il report in esame sottolinea come siano chiari agli occhi dei membri dell'*AI Advisory Body* sia i benefici che i rischi relativi all'intelligenza artificiale e che, proprio per tale ragione, una governance globale dell'IA basata su principi guida, che siano attuati attraverso funzioni chiare, appare necessaria. Pertanto, i prossimi passi del comitato – come anticipato – si focalizzeranno sul proseguimento delle interlocuzioni con diversi stakeholders nel mondo, al fine di elaborare il rapporto finale, il quale comprenderà metodologie per la valutazione dei rischi e per garantire l'interoperabilità dei differenti modelli di governance, fornendo casi di studio relativi a contesti specifici, in particolar modo con riguardo all'open source, all'IA nel settore finanziario, nonché con riferimento alla proprietà intellettuale, ai diritti umani e all'impatto sul mondo del lavoro.

I prossimi passi del comitato – come anticipato – si focalizzeranno sul proseguimento delle interlocuzioni con diversi stakeholder nel mondo, al fine di elaborare il rapporto finale, il quale comprenderà metodologie per la valutazione dei rischi e per garantire l'interoperabilità dei differenti modelli di governance, fornendo casi di studio relativi a contesti specifici

1.2. OCSE: dai principi del 2019 alle prime considerazioni sull'IA generativa

Facendo un passo indietro dal punto di vista cronologico, è opportuno fare riferimento ai Principi dell'OCSE sull'IA (OECD/LEGAL/0449), sottoscritti originariamente da 42 Paesi – oggi 46 – in occasione del rispettivo Consiglio ministeriale di maggio 2019, a cui si è data la forma della raccomandazione, per cui essi non sono giuridicamente vincolanti ma sono comunque caratterizzati da un'importante autorevolezza, anche perché costituiscono le prime linee guida approvate da un consesso intergovernativo in materia di intelligenza artificiale. Tali Principi furono

elaborati da un gruppo di oltre 50 esperti provenienti da governi, imprese, accademia, società civile, organismi internazionali e sindacati. Passando alla struttura, si delineano dapprima una serie di termini e concetti chiave, tra cui la definizione di "*AI system*" modificata lo scorso 8 novembre, a cui si è adeguato anche l'AI Act europeo[5], per proseguire con l'indicazione di 5 regole valoriali e 5 raccomandazioni per i *policymaker*. Lo scopo dichiarato è quello di promuovere un utilizzo dell'IA che sia innovativo, sicuro e affidabile, rispettoso dei diritti umani e dei valori democratici, adottando un approccio pragmatico e flessibile affinché tali regole e raccomandazioni resistano al passare del tempo.

Lo scopo dichiarato è quello di promuovere un utilizzo dell'IA che sia innovativo, sicuro e affidabile, rispettoso dei diritti umani e dei valori democratici, adottando un approccio pragmatico e flessibile affinché tali regole e raccomandazioni resistano al passare del tempo

Più nel dettaglio, la prima definizione è proprio quella di *AI system*, così come emendata lo scorso novembre: "*An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment*". La *ratio* alla base dell'aggiornamento della presente nozione si rivede innanzitutto nel chiarire quale può essere la natura degli obiettivi di un sistema di intelligenza artificiale, ossia espliciti o impliciti; sul punto, è stato eliminato il riferimento al fatto che tali obiettivi possono essere definiti esclusivamente dall'essere umano, in quanto è ormai pacifico che quelli inizialmente fissati in fase di sviluppo e progettazione vengano successivamente integrati da appositi *prompt* dell'utente una volta che l'IA entra in funzione, così come potrebbe verificarsi un disallineamento imprevisto tra l'obiettivo esplicitato in prima istanza da quello che diviene poi l'output effettivo. In secondo luogo, la nuova definizione evidenzia il ruolo dell'input, il quale può provenire sia da un essere umano, sia da un'altra macchina. In terzo luogo, tra gli output sono ora annoverati i "contenuti" (come testo, video o immagini), così da chiarire che la presente Raccomandazione si applica anche ai sistemi di IA generativa, divenuti piuttosto popolari negli ultimi tempi. Ulteriore modifica di rilievo concerne il riferimento all'adattabilità, riconoscendo in questo modo la possibilità che alcuni sistemi di IA possano evolversi dopo la loro progettazione e distribuzione, soprattutto se basati su tecniche di apprendimento automatico.

Le successive definizioni riguardano: a) "*AI system lifecycle*", precisando che le quattro fasi che lo caratterizzano (*design, data and models; verification and validation; deployment; operation and monitoring*) non seguono necessariamente una logica sequenziale, bensì iterativa; b) "*AI knowledge*", riferendosi a skill e risorse come dati, codici, algoritmi, modelli, ricerche, know-how, processi, best practices, programmi di formazione e governance, che risultino necessari per comprendere e intervenire nel ciclo di vita dell'IA; c) "*AI actors*", includendovi organizzazioni e individui che distribuiscono o utilizzano l'IA; d) stakeholders, ossia organizzazioni e individui che, direttamente o indirettamente, sono coinvolti (come sviluppatori, distributori e utilizzatori) oppure colpiti (tra cui coloro che ne subiscono in qualche modo gli effetti) da sistemi di IA.

[5] *Si v. infra, par. 4.*

Pertanto, gli stakeholders sono considerati un sotto-insieme degli attori di cui alla lettera precedente.

Ad ogni modo, tutte le precedenti definizioni vanno considerate insieme alle 5 regole valoriali di seguito esplicitate:

1. Crescita inclusiva, sviluppo sostenibile e benessere: l'IA deve arrecare benefici alle persone e al pianeta, promuovendo la crescita inclusiva, lo sviluppo sostenibile e il benessere;
2. Valori centrati sulla persona ed equità: i sistemi di IA devono essere progettati al fine di rispettare i diritti umani e i valori democratici, includendo meccanismi di salvaguardia per garantire una società giusta ed equa (es: ammettere l'intervento umano);
3. Trasparenza e spiegabilità: gli utenti hanno il diritto di conoscere e comprendere le modalità di funzionamento dei sistemi di IA, per cui gli vanno fornite informazioni complete e coerenti, compatibilmente con lo stato dell'arte, così da consentire un adeguato sfruttamento dei risultati;
4. Robustezza, sicurezza e protezione: i sistemi di IA devono essere monitorati e supervisionati dall'essere umano costantemente, affinché ci si assicuri un funzionamento sicuro e protetto durante l'intero ciclo di vita, il che si traduce nell'opera di prevenire e gestire i rischi in ciascuna delle fasi del ciclo di vita di tali sistemi (per cui è centrale il ruolo degli *AI actors*, ognuno rispetto al rispettivo ruolo di competenza);
5. Responsabilità: le organizzazioni e gli individui che sviluppano, implementano o utilizzano i sistemi di IA sono da ritenersi responsabili del loro corretto funzionamento, in conformità con le regole valoriali sin qui richiamate, tenuto conto dei rispettivi ruoli, del contesto e dello stato dell'arte.

In maniera complementare ai principi appena descritti, si collocano le 5 raccomandazioni che i Paesi aderenti dovrebbero tenere in considerazione per l'implementazione delle politiche nazionali e per la cooperazione internazionale:

I. Promuovere investimenti in R&S, sia pubblici che privati, per stimolare l'innovazione verso un'IA affidabile, nonché per realizzare dataset aperti che rispettino la privacy e la protezione dei dati;
II. Sostenere un ecosistema digitale per lo sviluppo di un'IA affidabile, che includa in particolare tecnologie digitali, infrastrutture e meccanismi per la condivisione della conoscenza associata all'IA (*AI knowledge* di cui sopra);
III. Favorire la diffusione di sistemi di IA affidabili attraverso la sperimentazione di ambienti controllati, nonché la revisione e l'adattamento del quadro regolatorio e dei meccanismi di *assessment* applicabili ai sistemi di IA;
IV. Costruire, attraverso la cooperazione con gli stakeholders, le competenze e la forza lavoro per un'equa e sicura trasformazione del mercato del lavoro e della società che possa cogliere i benefici dell'IA;
V. Collaborare a livello internazionale, ricorrendo a un approccio intersettoriale e multi-stakeholders, per la condivisione di conoscenza sull'IA che sia funzionale a sviluppare standard tecnici globali sull'IA, cosicché essa resa sia interoperabile e affidabile.

Allo scopo di evitare che tali Principi rimanessero su un piano meramente astratto, l'OCSE ha sancito, in chiusura del testo della presente Raccomandazione, di pubblicare entro cinque anni un report sullo stato di implementazione. Difatti, a ottobre scorso è stato reso noto il report "*The state of implementation of the OECD AI principles four years on*" (n.3/2023)[6], il quale offre una panoramica circa le strategie nazionali sull'IA, dei relativi organismi di supervisione, nonché dei quadri di monitoraggio e valutazione, evidenziando altresì esempi di politiche consolidate ed emergenti in materia, così da stimolare l'apprendimento incrociato tra i *policymakers*.

Pertanto, appare opportuno menzionare alcuni importanti risultati che emergono dal rapporto. Ad esempio, nel 2017 erano piuttosto ridotti i Paesi dotati di una strategia nazionale sull'IA, mentre oggi ve ne sono oltre 50 solo tra le economie aderenti all'OCSE o relativi partner. Inoltre, a maggio 2023, l'Osservatorio OCSE sulle politiche in materia di IA (*OECD.AI Policy Observatory*) ha registrato ben 930 iniziative a livello di policy provenienti da oltre 70 giurisdizioni; naturalmente, sul punto sono ancora molti e sostanziali i punti di divergenza – primo fra tutti la differenza tra l'approccio a una regolamentazione più rigida, che si scontra con una regolazione più snella e, talvolta, del tutto assente – ma da quest'ultimo monitoraggio emerge anche una convergenza su alcuni aspetti (es: framework sull'etica e principi per un'IA affidabile; regole giuridicamente vincolanti e specifiche per l'IA; standard tecnici; ambienti controllati per testare prodotti e servizi che utilizzano l'IA, ossia le c.d. *regulatory sandboxes*).

Dal quadro sin qui delineato, si può affermare come l'OCSE si stia impegnando notevolmente nel supportare i decisori per far fronte alle sfide connesse all'intelligenza artificiale. Pertanto, non sorprende che nell'ultimo anno si stia occupando anche di IA generativa, in particolare con il *working paper* "*Initial policy considerations for generative artificial intelligence*" del 18 settembre scorso, che accanto al riconoscimento delle molteplici opportunità, soprattutto per alcuni campi di applicazione (educazione, sanità, sviluppo di codice, industria creativa e motori di ricerca), evidenzia altresì alcune sfide critiche con riguardo al mercato del lavoro, alla tutela del diritto d'autore e alla possibilità di un uso improprio di tali sistemi per la creazione di contenuti manipolati che alimentino la disinformazione.

Non sorprende che nell'ultimo anno si stia occupando anche di IA generativa, in particolare con il working paper "Initial policy considerations for generative artificial intelligence" del 18 settembre scorso, che accanto al riconoscimento delle molteplici opportunità, soprattutto per alcuni campi di applicazione (educazione, sanità, sviluppo di codice, industria creativa e motori di ricerca), evidenzia altresì alcune sfide critiche

[6] *In verità, la versione dell'anno scorso è un parziale aggiornamento del primo report datato 2021 che, tuttavia, prendeva in considerazione esclusivamente l'implementazione delle 5 raccomandazioni dirette ai governi e quindi tralasciava la traduzione in iniziative concrete delle 5 regole valoriali.*

Di conseguenza, il documento evidenzia alcune questioni relative all'IA generativa che i policymaker dovrebbe affrontare, ossia:

1) Misinformazione e disinformazione: i contenuti generati con IA aumentano tali rischi, in quanto cresce la difficoltà di distinguerli da quelli generati dall'essere umano, con conseguenze dannose a livello individuale e sociale, in particolar modo su questioni scientifiche. Per ovviare – o comunque attenuare – tali problematiche, il paper suggerisce alcune misure, come lo sviluppo di modelli che forniscano vere e proprie evidenze e fonti di partenza, nonché sistemi di IA che agevolino la rilevazione dei contenuti sintetici e il *watermarking* (apposizione di una filigrana), oltre a stimolare la ricerca verso approcci innovativi al fine di superare i limiti di quelli attualmente esistenti (es: difficoltà a sviluppare sistemi di IA che rilevino un testo generato dall'IA e, più in generale, scarsa affidabilità degli algoritmi di rilevamento; elusione delle linee guida e dei codici di condotta da parte di attori commerciali scorretti o *state-sponsored*; manipolazione malevola, tramite attacchi di *spoofing*, dei modelli alla base del *watermarking);*
2) Amplificazione dei pregiudizi: l'IA può amplificare e alimentare pregiudizi sociali, stereotipi e discriminazioni sulla base dei dati di addestramento ricevuti, per cui è necessario promuovere una maggiore inclusività e cura di tali dati e ammettere la verifica da parte dell'utente umano;
3) Diritti di proprietà intellettuale: i modelli di IA generativa sono spesso addestrati con ingenti quantità di dati coperti da diritto d'autore, senza la previa autorizzazione dei rispettivi titolari. Inoltre, il documento sottolinea come sia ancora dibattuto se i contenuti generati dall'IA possano essere essi stessi coperti da diritto d'autore o brevettati e, in caso affermativo, da quale soggetto;
4) Impatti positivi e negativi sul mercato del lavoro: se da un lato l'IA generativa può far aumentare la disponibilità di lavoratori dotati di competenze specializzate in materia, contribuire al progresso tecnologico e alla riduzione dei costi, dall'altro può esporre al rischio di automazione alcuni lavori e specifiche mansioni, soprattutto se a bassa specializzazione, con ripercussioni sul livello occupazionale. Tuttavia – specifica il documento – la letteratura sugli effetti dell'IA generativa sul mercato del lavoro è piuttosto recente, per cui è stata sottoposta limitatamente al processo di revisione tra pari.

Ciò premesso, il paper in esame sottolinea che i rischi dell'IA – e in particolar modo quelli correlati all'IA generativa – richiedono soluzioni sistemiche su larga scala, ricorrendo, ad esempio, a regolamentare la materia, adottare principi etici e standard tecnici, nonché rilasciare modelli e strategie di accesso al mercato. In conclusione, vi è un richiamo esplicito al lavoro del G7 per migliorare la governance (anche) dell'IA generativa nell'ambito dell'ormai celebre Processo di Hiroshima sull'IA[7].

[7] *Si v. infra, par. 1.3.*

Il paper in esame sottolinea che i rischi dell'IA – e in particolar modo quelli correlati all'IA generativa – richiedono soluzioni sistemiche su larga scala, ricorrendo, ad esempio, a regolamentare la materia, adottare principi etici e standard tecnici, nonché rilasciare modelli e strategie di accesso al mercato

1.3. Il dialogo tra governi: i Principi guida e il Codice di condotta internazionale per i sistemi avanzati di IA del G7 nell'ambito dell'Hiroshima Process

Con l'inizio del nuovo anno si è concluso il periodo di presidenza giapponese del G7, durante il quale si sono fatti importanti passi avanti, come anticipato, verso la definizione di una governance dell'IA. Difatti, in tale contesto è stato istituito il già citato Processo di Hiroshima sull'IA con lo scopo di promuovere un'intelligenza artificiale sicura e affidabile a livello globale, con una particolare attenzione verso i sistemi di IA più avanzati. Più nel dettaglio, l'obiettivo prefissato constava nella definizione di Principi Guida e di un Codice di condotta volontario per gli sviluppatori di IA, al fine di favorire regole flessibili e armonizzate in materia. Entrambe le iniziative si sono effettivamente concretizzate in data 30 ottobre a seguito di un costante confronto con gli stakeholders.

In primo luogo, gli 11 Principi Guida delineati in tale sede – per i quali la Commissione europea aveva aperto un'apposita consultazione preparatoria dal 13 al 20 ottobre – si rivolgono a un'ampia gamma di soggetti come sviluppatori, distributori e utilizzatori di sistemi di IA avanzati, ivi compresi enti del mondo dell'accademia, società civile, settore privato e pubblico. Più nel dettaglio, tali Principi riguardano:

1. L'adozione di misure adeguate durante lo sviluppo di sistemi avanzati di IA, sia in fase di progettazione che di successiva distribuzione, al fine di identificare, valutare e ridurre i rischi che possano emergere durante l'intero ciclo di vita di tali sistemi;
2. L'identificazione e la conseguente riduzione delle relative vulnerabilità, nonché – se del caso – degli incidenti e degli abusi emersi dopo la fase di distribuzione;
3. La comunicazione al pubblico di capacità, limiti e ambiti di utilizzo appropriato e improprio dei sistemi avanzati di IA, con lo scopo di garantire la trasparenza e un uso responsabile degli stessi;
4. La condivisione responsabile delle informazioni e la segnalazione degli incidenti tra le principali organizzazioni di sviluppatori di sistemi avanzati di IA;
5. Lo sviluppo, l'implementazione e la divulgazione di politiche in materia di governance e gestione dei rischi (che siano fondate su un approccio basato sul rischio), ivi inclusi aspetti relativi alla protezione dei dati personali;
6. L'investimento in controlli di sicurezza solidi e sistematici, che comprendano il controllo sulle minacce fisiche e cibernetiche, sia interne che esterne all'organizzazione;

7. L'implementazione di meccanismi affidabili di autenticazione e provenienza dei contenuti, per consentire agli utilizzatori di identificare quando si trovano dinanzi a un contenuto generato dall'IA;
8. La prioritizzazione della ricerca per la mitigazione dei rischi sociali e di sicurezza, nonché gli investimenti in efficaci misure di mitigazione;
9. Lo sviluppo di sistemi avanzati di IA per affrontare adeguatamente le sfide globali, tra cui il cambiamento climatico, la salute e l'istruzione;
10. La promozione di standard tecnici a livello internazionale;
11. L'attuazione di misure tecniche adeguate per l'input, la protezione dei dati personali e della proprietà intellettuale.

Sulla base di tali principi è stato poi plasmato il Codice di condotta volontario, pubblicato – come anticipato – nella medesima occasione. In particolare, il Codice contiene indicazioni specifiche e pragmatiche per ciascuno dei Principi guida prima descritti. Più nel dettaglio:

I. Adottare misure adeguate per identificare, valutare e ridurre i rischi in tutto il ciclo di vita dell'IA, tramite test (come il *red-teaming)* da eseguirsi in ambienti sicuri nelle diverse fasi del ciclo di vita per identificare sin da subito le vulnerabilità di uno specifico sistema di IA, con una particolare attenzione per le fasi pre-commercializzazione, al fine di implementare i rimedi opportuni. Sul punto, il Codice evidenzia di valutare attentamente i rischi relativi alla sicurezza chimica, biologica, radiologica, nucleare, alle capacità cibernetiche offensive, nonché ai rischi per la salute, l'auto-riproduzione di sistemi di IA ("*self-replicating*"), oltre ai rischi che possano impattare la società nel suo complesso e la democrazia (es: discriminazione, disinformazione, ecc.);
II. Ricorrere a sistemi e procedure commisurate al livello di rischio, che siano funzionali al monitoraggio di vulnerabilità, incidenti, rischi emergenti e usi impropri, nonché misure adeguate per affrontarli, oltre a prendere in considerazione la possibilità di agevolare e promuovere la ricerca e la segnalazione di problemi e vulnerabilità anche a terze parti e utenti, ad esempio, tramite sistemi premiali e incentivi;
III. Pubblicare rapporti sulla trasparenza contenenti informazioni e istruzioni d'uso aggiornate, chiare e comprensibili per l'utente, congiuntamente alla documentazione tecnica. In tema, il Codice prescrive anche il contenuto minimo di tali documenti, tra cui rilevano i dettagli degli *assessment* sui potenziali rischi per la sicurezza e la società, le capacità del modello/sistema, le limitazioni significative che lo caratterizzano, nonché i risultati dei test condotti per valutare l'idoneità del sistema di IA a essere distribuito e commercializzato;
IV. Stabilire o aderire a meccanismi di cooperazione internazionale per sviluppare standard condivisi, strumenti, meccanismi e best practices per garantire la sicurezza e l'affidabilità dei sistemi di IA avanzati, impegnandosi altresì a condividere e segnalare al pubblico e, ove opportuno, alle autorità competenti le informazioni rilevanti in tal senso;
V. Puntare sullo sviluppo e l'implementazione di politiche organizzative di gestione e mitigazione del rischio, nonché sull'elaborazione di politiche sulla privacy, le quali coprano un'ampia gamma di rischi e siano costantemente aggiornate, garantendo al contempo la continua formazione del personale;

VI. Valutare la sicurezza di pesi, modelli e dati tramite l'implementazione di misure di sicurezza operative e di controllo degli accessi fisici e informatici adeguate, nonché programmi di rilevamento delle minacce interne (anche cyber) e di revisione delle misure di sicurezza adottate affinché rimangano efficaci e appropriate nel tempo;

VII. Prevedere meccanismi di autenticazione e provenienza dei contenuti creati con i sistemi di IA avanzati, sviluppando altresì strumenti o API per permettere agli utilizzatori di verificare se un contenuto è generato dall'IA o meno (es: *watermarking*);

VIII. Investire nella ricerca per migliorare la sicurezza, l'affidabilità, la trasparenza e l'interoperabilità dei modelli alla base dei sistemi di IA avanzati, con un focus sui rischi in tema di tenuta della democrazia, diritti umani, protezione dei soggetti vulnerabili, proprietà intellettuale, bias cogntivi e disinformazione;

IX. Impegnarsi nello sviluppo dell'IA a beneficio di tutti, in linea con gli Obiettivi di sviluppo sostenibile dell'ONU, dando priorità alla gestione responsabile dell'IA, all'alfabetizzazione digitale e alla collaborazione con la società civile;

X. Contribuire allo sviluppo e all'adozione di standard tecnici e di prassi internazionali, cooperando con le Organizzazioni di sviluppo degli standard, con particolare riferimento a metodologie di test sui dati, meccanismi di autenticazione e provenienza dei contenuti, politiche di cibersicurezza e di *reporting*, nonché a misure per distinguere i contenuti generati dall'IA;

XI. Adottare misure di gestione della qualità dei dati sin dalla fase di raccolta e creazione dei dataset, al fine di mitigare i bias *by design*. Tra le misure indicate rilevano la trasparenza, tecniche di addestramento *privacy-oriented* e il *fine-tuning* del sistema di IA affinché preservi la proprietà intellettuale e i contenuti coperti da diritto d'autore.

In conclusione, i principi sin qui descritti costituiscono certamente un importante passo avanti verso la definizione di una nuova governance globale dell'IA, coerentemente col lavoro che sta procedendo in altre sedi a livello internazionale (molte delle quali citate nel presente studio). Inoltre, nonostante non sia uno strumento giuridicamente vincolante, va considerato che il Codice di condotta è immediatamente applicabile e data l'autorevolezza del consesso da cui proviene, ispirandosi peraltro a quanto già stabilito in ambito OCSE (v. *supra*), rappresenta un utile strumento per quelle organizzazioni che intendano dimostrare una maggiore attenzione dei sistemi di IA da loro sviluppati, distribuiti o comunque utilizzati.

Il Codice di condotta è immediatamente applicabile e data l'autorevolezza del consesso da cui proviene, ispirandosi peraltro a quanto già stabilito in ambito OCSE, rappresenta un utile strumento per quelle organizzazioni che intendano dimostrare una maggiore attenzione dei sistemi di IA da loro sviluppati, distribuiti o comunque utilizzati

1.4. Una strategia comune per l'IA: la Dichiarazione di Bletchley Park

L'1 e il 2 novembre scorso il Regno Unito ha ospitato il primo *AI Safety Summit*, che ha riunito allo stesso tavolo rappresentanti governativi di 28 Paesi[8] e dell'Unione Europea, alla presenza degli esponenti delle principali aziende del settore dell'intelligenza artificiale, al fine di discutere della direzione da intraprendere affinché l'umanità abbia a disposizione un'IA responsabile e sicura. Il prodotto di questa importante iniziativa politica si è rivisto nella cosiddetta "Dichiarazione di Bletchley Park" – dal nome del celebre luogo in cui si è svolto l'incontro – sottoscritta da tutti e 29 i partecipanti istituzionali.

Il documento parte dalla constatazione che per cogliere appieno le numerose opportunità dell'IA, quest'ultima vada progettata, sviluppata, diffusa e utilizzata in maniera sicura, affidabile e responsabile secondo un approccio incentrato sull'essere umano, garantendo che essa possa favorire l'inclusione, lo sviluppo e una crescita sostenibile per tutti. Ciò si ritiene possibile solo attraverso un'efficace e quanto più ampia possibile cooperazione internazionale, che si focalizzi su un modello regolatorio e di governance proporzionato e pro-innovazione (es: classificazione e categorizzazione secondo un approccio basato sul rischio; emanazione di principi e codici di condotta in materia), il quale possa assicurare la massimizzazione dei benefici e la riduzione dei rischi associati all'IA. Altro aspetto di rilievo in tal senso riguarda la collaborazione tra Stati, fora internazionali competenti, imprese, società civile e mondo accademico.

Ciò si ritiene possibile solo attraverso un'efficace e quanto più ampia possibile cooperazione internazionale, che si focalizzi su un modello regolatorio e di governance proporzionato e pro-innovazione

La Dichiarazione prosegue ammettendo i molteplici rischi che porta ancora con sé questa tecnologia in termini di diritti umani, privacy e valori democratici, senza sottovalutare anche le sfide etiche e sociali. Inoltre, con particolare riferimento a quei sistemi di IA capaci di auto-apprendere, viene evidenziato come questi possano addirittura sfuggire al controllo e alla supervisione dell'essere umano, esponendo quest'ultimo a rischi sostanziali derivanti da un uso improprio dell'IA, sia di tipo intenzionale, sia correlato a problemi involontari di controllo rispetto all'intento umano. Sul punto, il documento ammette che tali problemi non sono pienamente conoscibili allo stato attuale e sono, peraltro, difficili da prevedere, soprattutto in alcuni settori come la cybersicurezza e le biotecnologie, oppure circa la possibilità di amplificare le ripercussioni dovute alla disinformazione potenziata con l'IA. Pertanto, la Dichiarazione pare voler insistere innanzitutto su un maggiore sforzo sul piano internazionale verso la comprensione di tali rischi e delle azioni concrete per affrontarli adeguatamente.

Un punto ricorrente nella Dichiarazione concerne il riconoscimento del fatto che la regolazione dell'IA può basarsi su approcci differenti dovuti alle circostanze nazionali e ai framework normativi applicabili. Per tale ragione, occorrerebbe richiedere agli attori che sviluppano IA una maggiore

[8] *Arabia Saudita, Australia, Brasile, Canada, Cile, Cina, Corea del Sud, Emirati Arabi Uniti, Filippine, Francia, Germania, Giappone, India, Indonesia, Irlanda, Israele, Italia, Kenya, Nigeria, Paesi Bassi, Regno Unito, Ruanda, Singapore, Spagna, Stati Uniti, Svizzera, Turchia, Ucraina.*

trasparenza di tali sistemi, nonché responsabilizzare gli stessi attraverso l'istituzione di elevati standard di sicurezza e di valutazione dei rischi volti alla tutela dei relativi utilizzatori.

Sulla base di queste premesse, la Dichiarazione impegna i suoi firmatari a porre al centro delle rispettive agende politiche due macro-temi:

1. Identificare i rischi per la sicurezza dell'IA di interesse comune, sostenendo una comprensione di tali rischi che sia basata su dati scientifici e vada di pari passo col progresso di questa tecnologia, accelerando altresì la condivisione della conoscenza a livello globale al fine di comprendere l'impatto dell'IA sull'intera società;
2. Puntare sulla collaborazione per dar vita a politiche basate sul rischio funzionali a garantire la sicurezza dell'IA, tenendo in considerazione le opportune differenze di approccio a livello nazionale.

Anche se la Dichiarazione non si traduce immediatamente in azioni specifiche, sancisce l'intenzione di sostenere una rete inclusiva di ricerca scientifica sulla sicurezza dell'IA a livello internazionale, che possa comprendere e integrare la collaborazione multilaterale, plurilaterale e bilaterale, anche ricorrendo a fora internazionali esistenti e a nuove iniziative rilevanti in tal senso (tra cui le prossime riunioni dell'*AI Safety Summit*). In conclusione, dal documento presentato a Bletchley Park emerge chiaramente quanto sia ritenuto fondamentale che l'IA sia sviluppata e implementata in maniera affidabile e trasparente, indipendentemente da chi sia l'utilizzatore e dal relativo campo di applicazione, ponendo un focus prioritario sulle misure tecnologiche e poi sulla definizione di regole giuridiche (diversamente dall'approccio UE[9]). Ad ogni modo, secondo quanto emerge dalla Dichiarazione, risulta evidente che sia assolutamente indispensabile puntare sulla collaborazione e la cooperazione multi-stakeholder affinché si possa garantire la sicurezza e l'affidabilità dell'IA.

Secondo quanto emerge dalla Dichiarazione, risulta evidente che sia assolutamente indispensabile puntare sulla collaborazione e la cooperazione multi-stakeholder affinché si possa garantire la sicurezza e l'affidabilità dell'IA

Peraltro, va evidenziato che contestualmente a tale iniziativa, il Regno Unito ha istituito l'*AI Safety Institute* avente lo scopo dichiarato di ridurre al minimo le sorprese derivanti da progressi rapidi e inaspettati dell'IA, sviluppando in tal senso l'infrastruttura sociale e tecnologica necessarie per comprendere i rischi dei sistemi avanzati di IA e consentirne la governance. Più nel dettaglio, esso porterà avanti, quantomeno in prima battuta, tre funzioni principali: i) sviluppo e valutazione dei sistemi di IA avanzati; ii) ricerca sicura sull'IA; iii) facilitazione dello scambio di informazioni.

[9] *Si v. infra, par. 4..*

1.5. World Bank: opportunità e rischi dell'IA nel settore pubblico

Nel corso degli anni, diversi organismi internazionali hanno fornito il loro contributo relativamente alla costruzione di una governance in materia di intelligenza artificiale e tra questi vi è anche la Banca Mondiale, la quale nel 2020 ha pubblicato un importante report ("*Artificial Intelligence in the Public Sector. Maximizing Opportunities, Managing Risks*"), il cui titolo esplicita sin da subito il focus del documento: massimizzare le opportunità e, al contempo, gestire i rischi derivanti dall'uso dell'IA nel settore pubblico. In particolare, lo scopo è quello stimolare la modernizzazione del settore pubblico attraverso l'intelligenza artificiale e le sue applicazioni.

In merito alle opportunità, vengono evidenziati specifici casi d'uso dell'IA provenienti da tutto il mondo, tra cui:

1) Corruzione: in Brasile è stato sviluppato un sistema di IA – uno dei più grandi *data lake* esistenti, che al tempo includeva già 27 dataset differenti e 250 milioni di dati – capace di identificare 225 indicatori (*red flags*) di possibili frodi nei processi di *public procurement*. Tale sistema ha consentito di ottimizzare il meccanismo di rilevamento delle frodi nel settore dei contratti pubblici, conseguendone un minor dispendio risorse e accrescendo l'efficacia degli audit e delle investigazioni. Inoltre, il progetto è stato valutato come altamente scalabile grazie a un meccanismo di "*Scalabale Data Unification*", che sostanzialmente consente di abbattere i costi di replicazione attraverso la costruzione di uno schema globale della spesa pubblica, identificando e poi convertendo i singoli dataset locali (es: riguardanti una singola Regione o Stato) in quello globale;
2) Coinvolgimento dei cittadini: in Nigeria è stata testata una soluzione con IA per raccogliere le opinioni dei cittadini su alcune iniziative pubbliche. Le caratteristiche più rilevanti di tale soluzione riguardano la possibilità di riassumere il testo, mostrare le parole chiave e la loro valenza nel contesto, escludere i feedback ricevuti dall'esterno di un'area geografica, classificare le immagini sulla base dei criteri e dei dati di addestramento, analizzare le opinioni fornite per coglierne sensazioni negative, neutrali o positive utili a confermare o modificare taluni aspetti delle iniziative pubbliche;
3) Controllo della dogana e delle frontiere: il documento riporta il caso degli Stati Uniti che sin dal 2016 ricorrono a un sistema di IA capace di monitorare le imbarcazioni a chilometri di distanza, riconoscere alcuni movimenti insoliti e segnalarli alle autorità competenti, cosicché possano intercettare le possibili fonti di contrabbando e altre attività illecite al confine;
4) Sanità: durante la pandemia da Covid-19, l'IA è stata utilizzata prevalentemente per predire i potenziali effetti di differenti tipi di politiche di quarantena, nonché per controllare possibili infezioni dei pazienti tramite modelli di riconoscimento termico e facciale. In entrambi i casi, il report evidenzia come si possa affermare una buona accuratezza dei risultati rispetto alla situazione poi realmente verificatasi. Inoltre, a Singapore è stato lanciato un chatbot dotato di IA per fornire al personale sanitario informazioni aggiornate, tra cui gli ultimi protocolli sanitari e il dosaggio di alcuni farmaci per contrastare i sintomi della malattia;

5) Settore giudiziario: nel Regno Unito è stato sviluppato un assistente legale potenziato con l'IA capace di effettuare, fra l'altro, l'analisi dei contratti e consulenza legale in maniera guidata con una precisione del 86,6%, ossia oltre il 20% in più di quella ottenuta da avvocati umani. Altro caso d'uso ha riguardato la soluzione "*DoNotPay*", un bot di IA che aiuta le persone a contestare le multe per divieto di sosta con una percentuale di successo del 64%;
6) *Procurement*: il governo statunitense ha sviluppato un progetto di IA per la scansione dei documenti di gara (*Solicitation Review Tool*), che ha lo scopo di verificare la conformità ad alcune normative federali e, in caso negativo, segnalare alle parti di effettuare le opportune azioni correttive;
7) Tassazione: negli Stati Uniti l'IA viene utilizzata congiuntamente a piattaforme avanzate di *analytics* per verificare il rispetto della disciplina fiscale da parte delle imprese, consentendo di tracciare piste investigative, identificare schemi ricorrenti, scoprire frodi fiscali e condurre più agevolmente attività investigative sul riciclaggio di denaro. Inoltre, una soluzione di IA – denominata "*AI Economist*" – viene impiegata per modellare le politiche fiscali secondo un approccio *data-driven* e già nel 2020 aveva dimostrato di poter trovare, partendo dall'osservazione di attori umani in casi reali, politiche che garantissero il 16% in più di uguaglianza e produttività rispetto al modello fiscale classico di Saez;
8) Audit: le società di revisione contabile utilizzano l'IA per massimizzare l'efficienza dei controlli e ridurre al minimo il costo dei processi. In questo senso, l'IA può analizzare i documenti contrattuali e presentare potenziali anomalie all'operatore umano tramite la produzione di un punteggio di rischio che si basa su alcuni parametri contabili, consentendo un risparmio di tempo di oltre il 90%. Uno degli esempi di questo tipo di soluzioni è *MindBridge AI Auditor*, adottato anche dal governo britannico e canadese.

Come anticipato, accanto alle opportunità, il report si focalizza anche sui rischi dell'uso dell'IA nel settore pubblico, con particolare riferimento a quattro categorie principali:

I. Performance, Fiducia e Bias: il documento indica alcuni casi da cui si può evincere che l'IA può condurre all'assunzione di decisioni pregiudizievoli e discriminatorie in favore o contro uno o più gruppi di persone nel momento in cui il modello è stato addestrato su dati incompleti, imprecisi o corrotti intenzionalmente, proponendo come possibile soluzione la diffusione di dataset e algoritmi aperti per mitigare questa tipologia di rischi – il che condurrebbe a garantire trasparenza, possibilità di contestare la decisione assunta dall'IA e supervisione umana – oltre a sviluppare e adottare quadri di gestione del rischio da parte dei governi;
II. Cybersicurezza: i cyberattacchi, soprattutto grazie a *phishing* e *spear-phishing,* possono danneggiare l'algoritmo di addestramento e modificarlo, inserendovi diverse tipologie di malware. In tema, le soluzioni proposte nel documento spaziano dalla previsione di buone pratiche solide in ambito cybersecurity, all'adottare un atteggiamento proattivo per scovare e mitigare le vulnerabilità di sicurezza prima degli attori malevoli, sino a implementare soluzioni di IA che possano svolgere un ruolo difensivo nella prevenzione, identificazione e gestione delle minacce;
III. Controllo: il controllo proattivo, il monitoraggio, i test e la convalida dei risultati dei sistemi di IA sono ritenuti aspetti essenziali per evitare casi limite, in cui il funzionamento e gli

output dell'IA sfuggano alla comprensione dell'essere umano. Difatti, il documento riporta la situazione – realmente verificatasi – di due sistemi di IA entrati in competizione tra di loro nel 2010, che hanno causato un'inflazione artificiale del mercato finanziario, oppure il caso in cui due chatbot che, interagendo tra di loro, creino un linguaggio non comprensibile dall'utilizzatore;

IV. Privacy: si suggerisce di implementare una solida base giuridica per mitigare i rischi legati alla privacy e alla protezione dei dati personali, oltre a un quadro di governance che promuova l'autovalutazione (*self-assessment*), la revisione tra pari e l'inclusione del settore pubblico per rafforzare il rispetto del framework normativo di riferimento;

Il report in esame prosegue fornendo una panoramica dei modelli di governance adottati a livello governativo, con un focus specifico su tre aspetti, quali i principi etici, il ruolo di un'agenzia governativa centrale e il quadro operativo. Con riguardo al primo aspetto, si evidenzia come siano ricorrenti alcuni principi: a) privacy e protezione dei dati personali; b) *accountability*; c) sicurezza, intesa anche come *safety*; d) trasparenza e spiegabilità; e) correttezza; f) controllo umano della tecnologia; g) collaborazione multistakeholder; h) promozione di una visione umano-centrica. Sul punto, si sottolinea che la vera sfida nell'adozione di questi principi si rivede nel bilanciare le attività di supervisione ed *enforcement* con la flessibilità in fase di attuazione.

Si sottolinea che la vera sfida nell'adozione di questi principi si rivede nel bilanciare le attività di supervisione ed enforcement *con la flessibilità in fase di attuazione*

Il secondo aspetto pone l'attenzione sull'implementazione dell'IA a livello centralizzato o decentralizzato, evidenziando come il primo caso sia quello che si presenta più frequentemente. Infatti, tendenzialmente i governi costituiscono un hub centrale che mette in comune i talenti disponibili, collabora con le diverse agenzie, fornisce un laboratorio per l'IA e sviluppa alleanze con il mondo accademico, il settore privato e le start-up. In ultimo, il terzo aspetto concerne un livello più operativo, in cui un soggetto pubblico – solitamente un'agenzia governativa centrale – definisce i problemi da affrontare e li concettualizza insieme agli esperti del settore, propone una possibile strada da percorrere per risolvere il problema e sviluppa le conseguenti azioni concrete. Inoltre, è ritenuto essenziale che il quadro operativo affronti le questioni inerenti la necessità di competenze e i rispettivi ruoli all'interno di un'organizzazione.

L'ultima parte del report si concentra sulla costruzione di un adeguato approccio governativo per l'utilizzo dell'IA, che non possa prescindere da caratteristiche volte a garantire l'interoperabilità e la sicurezza, oltre che la continuità dell'architettura propria dei diversi sistemi di IA progettati per essere utilizzati da una medesima struttura organizzativa. Più nel dettaglio, ci si riferisce a tre fattori chiave:

A. Architettura governativa: il documento parte dal presupposto che la maggior parte delle economie afferenti alla Banca Mondiale utilizzano sistemi *standalone* (denominati, spesso, "silos") che non sono interoperabili tra di loro, mentre i modelli di IA hanno tendenzialmente bisogno di grandi quantità di dati per restituire un output corretto e che

fornisca un valore aggiunto per i più differenti scopi. Pertanto, sarebbe auspicabile che i governi implementino a livello centrale una piattaforma comune per i dati, che consenta di raccogliere e gestire i dataset di interesse lungo l'intera struttura organizzativa (es: ministeri, dipartimenti, direzioni, ecc..). Una delle soluzioni ideali sarebbe quella di realizzare un "*data fabric*", ossia un'architettura di big data su larga scala che fornisce capacità di archiviazione, calcolo e sicurezza a organizzazioni che hanno bisogno di gestire grandi quantità di dati, come i governi e le multinazionali;

B. Interoperabilità: quanto sin qui esplicitato conduce al secondo fattore, considerato abilitante rispetto alla scalabilità, all'interscambio e all'interconnessione dei sistemi che contengono ed elaborano dati, anche attraverso l'ausilio dell'IA, ovviando in tal modo a effetti quali i "colli di bottiglia" che di fatto impediscono – o rendono più complicata e onerosa – la condivisione dei dati;
C. Standardizzazione dei dati: altro tassello cruciale riguarda la promozione della standardizzazione dei dati tramite la previsione di meccanismi di governance dei dati stessi. Ad esempio, in Estonia sono stati creati consessi che si occupano di *data governance* e nominano dei *"data steward"* in ciascuna agenzia governativa per coordinare la standardizzazione e l'interoperabilità dei dati. Si tratta di un meccanismo che fa parte di una più ampia strategia sulla governance dei dati che, fra le altre cose, definisce l'autorità competente in materia e si occupa di politiche, processi, standard, definizioni e accordi per lo scambio di dati a livello governativo. Parallelamente, viene raccomandato di impegnarsi verso lo sviluppo e la diffusione di buone pratiche a livello internazionale, come nel caso degli *Open Contracting Data Standard (OCDS).*

In sostanza, si raccomanda l'implementazione di un *data fabric*, così come appena definito, in quanto ritenuto intrinsecamente resiliente, adattivo e decentralizzato, il che consentirebbe di puntare sull'accesso standardizzato ai dati da parte delle singole soluzioni di IA lungo l'intera struttura governativa, con chiari vantaggi in termini competitivi sia per il settore pubblico, sia privato.

Si raccomanda l'implementazione di un data fabric, così come appena definito, in quanto ritenuto intrinsecamente resiliente, adattivo e decentralizzato, il che consentirebbe di puntare sull'accesso standardizzato ai dati da parte delle singole soluzioni di IA lungo l'intera struttura governativa

Nella parte conclusiva, il report suggerisce alcune priorità ai policymakers per affrontare al meglio i rischi e massimizzare le opportunità correlate all'uso dell'IA nel settore pubblico, con particolare riferimento verso:

1) L'adozione di politiche e quadri di governance che promuovano, al contempo, un'IA umano-centrica e la massimizzazione delle relative opportunità;
2) Gli investimenti in capitale umano e infrastrutture digitali.

In particolare, il primo aspetto richiederebbe: l'adozione di una strategia sull'intelligenza artificiale che affondi le radici in uso etico dell'IA; la promozione della trasparenza e della responsabilizzazione tramite un approccio multistakeholder inclusivo e che copra ciascuna fase del processo di creazione di policy, dalla progettazione all'implementazione; l'attenzione verso le competenze digitali, l'educazione e il *reskilling*; un quadro regolatorio per contrastare la propaganda e la disinformazione online, nonché il cybercrime, che sfruttino o colpiscano sistemi di IA; rafforzare la privacy, la protezione dei dati personali e le libertà civili, nonché monitorare la relativa compliance.

In maniera complementare, il secondo aspetto pone al centro investimenti diretti a garantire fondi per la ricerca, l'educazione e lo sviluppo di competenze digitali, sia più generiche, sia focalizzate sull'IA. Al contempo, andrebbe incentivata l'imprenditorialità innovativa tramite un apposito fondo per l'innovazione, programmi di prestito attraverso la collaborazione con banche di sviluppo nazionali e altri programmi di prestito per le PMI, come si sta già verificando in alcuni Paesi e in Unione Europea. Allo stesso modo, gli investimenti dovrebbero puntare sulla realizzazione del già citato *data fabric*. Infine, andrebbero incentivate formule quali i progetti pilota e i *proof-of-concept*, che dovrebbero includere un ampio coinvolgimento dei cittadini, al fine di esplorare le opportunità dell'IA a risolvere problemi specifici del settore pubblico.

1.6. OMS: dai principi etici e di governance del 2021 agli aspetti regolatori dell'IA per la sanità

Altro contributo di rilievo sulla definizione di principi etici e di governance applicabili all'intelligenza artificiale è stato fornito dall'Organizzazione Mondiale di Sanità (OMS) che a giugno 2021 ha pubblicato un report ("*Ethics and Governance of Artificial Intelligence for Health*") contenente linee guida e raccomandazioni in merito allo sviluppo, all'applicazione e all'uso delle tecnologie di IA nel settore sanitario.

A giugno 2021 è stato pubblicato un report ("Ethics and Governance of Artificial Intelligence for Health") contenente linee guida e raccomandazioni in merito allo sviluppo, all'applicazione e all'uso delle tecnologie di IA nel settore sanitario

Il documento è piuttosto corposo in quanto tocca numerosi aspetti, che vanno dall'analisi delle principali opportunità e sfide poste dall'IA in ambito medico-sanitario, sino all'illustrazione di 6 principi etici elaborati da un gruppo di 20 esperti dell'OMS per guidare lo sviluppo e l'adozione dell'IA in sanità, occupandosi altresì di fornire indicazioni pratiche per l'implementazione di tali principi da parte di tre gruppi di stakeholders: sviluppatori e programmatori di tecnologie di IA, ministri della salute e operatori sanitari.

Più nel dettaglio, le principali applicazioni dell'IA in sanità riportate nel documento riguardano:

A. La ricerca sanitaria: in questo campo la qualità dei risultati dipende sostanzialmente dai dati generati dalle cartelle cliniche elettroniche, il che naturalmente presuppone un database e

un sistema informatico idonei. Ad ogni modo, l'IA può essere applicata per identificare le migliori pratiche cliniche e sviluppare linee guida e strumenti di supporto clinico;

B. Lo sviluppo di farmaci: l'IA può essere utilizzata per semplificare e accelerare la scoperta e lo sviluppo di farmaci, un processo complesso che necessita di risorse umane, capitale da investire e grandi quantità di dati. In questo campo, l'IA è stata già utilizzata per identificare potenziali trattamenti per la malattia del virus Ebola e si prevede che in futuro possa consentire l'evoluzione del settore verso l'esecuzione di test in modalità virtuale, ricorrendo a modelli computerizzati del corpo umano;
C. La gestione e la pianificazione dei sistemi sanitari: l'IA può essere utilizzata per assistere il personale nelle attività logistiche, tra cui l'ottimizzazione delle forniture mediche, per supportare processi decisionali complessi, oppure per eseguire compiti banali o ripetitivi. In uno studio condotto in Brasile l'IA è stata utile per ottimizzare l'allocazione delle risorse del sistema sanitario nazionale, in base a dove risultasse più urgente intervenire. Tuttavia, sul punto sono emerse delle criticità relative al rischio che le risorse possano non essere allocate in maniera equa in caso di dati di addestramento poco inclusivi, errati o caratterizzati da pregiudizi;
D. Il monitoraggio della salute pubblica: gli strumenti dotati di IA posso essere impiegati per identificare popolazioni o località specifiche che sono caratterizzate da un comportamento ad alto rischio o che potrebbe beneficiare in particolar modo di comunicazioni a scopo sanitario, anche attraverso il *micro-targeting*. In particolare, quest'ultima modalità può presentare criticità, amplificate dall'opacità degli algoritmi, rispetto a pubblicità di natura commerciale e/o politica. L'IA può anche essere di supporto per la prevenzione delle malattie identificando, ad esempio, possibili contaminazioni nell'acqua utilizzata per l'agricoltura, oppure analizzando il livello di inquinamento dell'aria, o ancora fornendo modelli predittivi funzionali a prepararsi alle emergenze sanitarie.

Con riguardo ai già citati principi etici per l'uso dell'IA in sanità, il documento sottolinea che, mentre essi sono da considerarsi universali, la loro implementazione può differire in base al contesto culturale, religioso e sociale di riferimento. Ad ogni modo, i principi sono:

1. Protezione dell'autonomia umana: l'essere umano deve mantenere il controllo dei sistemi sanitari e delle decisioni mediche, il che implica anche che le persone comprendano il ruolo dell'IA nei trattamenti sanitari che li riguardano;
2. Promozione del benessere delle persone, della loro sicurezza e del pubblico interesse: le tecnologie di IA non devono danneggiare le persone, per cui chi le progetta deve farlo in maniera conforme ai requisiti normativi di sicurezza e accuratezza, oltre a rispettare i casi d'uso consentiti;
3. Predisposizione di garanzie adeguate in termini di trasparenza, spiegabilità e comprensibilità dei sistemi di IA: le soluzioni di IA devono essere intelligibili o comprensibili per gli sviluppatori, i professionisti sanitari, i pazienti, gli utenti e le autorità di regolamentazione, per cui è necessario puntare sulla trasparenza e sulla spiegabilità alla base delle applicazioni di IA, avendo come parametro le capacità di coloro ai quali vengono spiegate;

4. Promozione della responsabilità e dell'affidabilità delle tecnologie impiegate: gli stakeholders devono assicurarsi che l'IA sia utilizzata in condizioni adeguate e da persone opportunamente formate. Inoltre, deve essere stabilita la supervisione umana sulle tecnologie di IA, anche per individuare un punto di responsabilità nel caso in cui individui o gruppo siano danneggiati a seguito di decisioni basate su algoritmi;
5. Garanzia che siano attuati i criteri di inclusione ed equità nello sviluppo e nell'uso dell'IA: l'inclusività richiede che l'IA in ambito sanitario sia progettata per un uso equo ed appropriato, indipendentemente da età, sesso, genere, reddito, razza, etnia, orientamento sessuale, abilità o altre caratteristiche protette dalla disciplina sui diritti umani, così da evitare di sostenere – o comunque amplificare – i pregiudizi e le discriminazioni già esistenti;
6. Sostenibilità ed efficacia dell'IA rispetto alle esigenze delle persone: Chi progetta, sviluppa e utilizza tecnologie di IA dovrebbe valutarle in maniera continua, sistematica e trasparente, al fine di stabilire se l'IA stia funzionando in modo adeguato e in base alle aspettative e ai requisiti comunicati, i quali devono risultare comunque legittimi. Inoltre, l'uso dell'IA dovrebbe essere coerente con gli sforzi globali per ridurre l'impatto degli esseri umani sull'ambiente, gli ecosistemi e il clima. Tale principio dovrebbe concretizzarsi anche nell'impegno da parte di governi e imprese ad affrontare l'impatto dell'IA sul mondo del lavoro, il che comprende di curare la formazione degli operatori sanitari affinché sappiano adattarsi all'uso delle tecnologie di IA nel settore.

Proseguendo, il report in esame fornisce elementi utili per garantire la governance dell'IA in ambito sanitario, a partire dai modelli in fase di sviluppo o già elaborati in materia, focalizzandosi su aspetti quali:

A. *Data governance*: partendo dal presupposto che le tipologie, la quantità e gli ambiti di utilizzo dei dati sanitari sono notevolmente aumentati, anche per scopi commerciali, è necessario salvaguardare la privacy e la protezione dei dati personali degli individui, garantendo al contempo i benefici connessi alla raccolta e all'uso dei loro dati (es: istituendo e curando meccanismi sicuri di *data sharing*), il che conduce alla previsione e al rispetto della legislazione di riferimento (es: GDPR), sia da parte degli attori pubblici che privati;
B. Controllo e condivisione dei benefici: l'OMS dovrebbe assicurare la comprensione circa quali diritti si applicano all'uso dei dati sanitari e dell'IA contenuti in tecnologie sanitarie, mentre i governi, gli istituti di ricerca e le università coinvolte nello sviluppo di tecnologie di IA dovrebbero mantenere l'interesse nella proprietà dei risultati raggiunti, cosicché i benefici siano condivisi e resi ampiamenti accessibili, soprattutto verso quelle popolazioni che hanno contribuito con i loro dati. Inoltre, i governi dovrebbero considerare forme di incentivi alternativi ai diritti di proprietà intellettuale, come premi e finanziamenti, al fine di stimolare attività di R&S appropriate;
C. Governance del settore privato: i governi dovrebbero prendere in considerazione l'adozione di modelli di co-regolamentazione col settore privato per comprendere le tecnologie di IA e instaurare forme di partenariato pubblico-privato che sviluppino o impieghino soluzioni di IA per la sanità. Le imprese, d'altro canto, dovrebbero investire in misure per migliorare la

progettazione, la supervisione e l'affidabilità dei loro prodotti, nonché propendere a concedere licenze o certificazioni per gli sviluppatori di sistemi di IA ad alto rischio, come quelli utilizzati in ambito sanitario;

D. Governance del settore pubblico: i governi dovrebbero sempre svolgere valutazioni di impatto (sull'etica, sui diritti umani, sulla sicurezza e la protezione dei dati) trasparenti e inclusive prima di selezionare una qualsiasi tecnologia di IA per il settore sanitario e successivamente regolamentare i casi e le modalità di utilizzo. Inoltre, essi dovrebbero definire standard etici e legali per l'acquisto di tecnologie di IA e promuovere la loro integrazione nei confronti dei fornitori di servizi sanitari pubblici e privati. Peraltro, sarebbe auspicabile che i governi e le autorità sanitarie nazionali decidano di avvalersi di un sistema di IA per scopi sanitari (e non solo) solo a seguito di un confronto democratico e multistakeholder, che includa i pazienti e i rappresentanti dei gruppi minoritari, anche al fine di evitare che l'uso dell'IA possa creare o amplificare le diseguaglianze sociali o sanitarie;
E. Considerazioni regolatorie: i governi dovrebbero introdurre e rafforzare standard legali da applicarsi alle nuove tecnologie di IA, con lo scopo di evitare usi insicuri o pericolosi in ambito sanitario. Inoltre, andrebbe richiesta la trasparenza su alcuni aspetti dell'IA, come il codice sorgente e i dati di input, oltre a incentivare gli sviluppatori a identificare, monitorare e mitigare i rischi rilevanti per la sicurezza e i diritti umani sin dalla fase di progettazione della tecnologia;
F. Osservatorio sulle politiche e sulla legislazione in materia di IA: l'OMS intende supportare iniziative, anche con altri fora internazionali, che siano funzionali alla formulazione di leggi, policy e best practices per lo sviluppo, la distribuzione e l'uso sicuro dell'IA in sanità;
G. Governance globale dell'IA: i governi dovrebbero supportare una nuova governance globale dell'IA per il settore sanitario nel rispetto delle norme etiche, dei diritti umani e degli obblighi di legge, garantendo la partecipazione della società civile e, in particolare, di organizzazioni non governative e quelle che rappresentano gruppi minoritari.

In ultimo, come anticipato, il documento fornisce indicazioni pratiche per gli sviluppatori di IA, i ministri della salute e gli operatori sanitari, oltre a specificare che simili considerazioni costituiscono un punto di partenza per ulteriori occasioni di discussione e per orientare i policymakers e gli stakeholders nello sviluppare e impiegare l'IA per scopi medico-sanitari. Difatti, lo scorso 18 ottobre è stato pubblicato un nuovo report dall'OMS denominato "*Regulatory considerations on artificial intelligence for health*" in risposta alla crescente esigenza dei Paesi di gestire in modo responsabile le tecnologie di IA in sanità.

Lo scorso 18 ottobre è stato pubblicato un nuovo report dall'OMS denominato "Regulatory considerations on artificial intelligence for health" in risposta alla crescente esigenza dei Paesi di gestire in modo responsabile le tecnologie di IA in sanità

In particolare, quest'ultimo report si focalizza essenzialmente sulle seguenti aree di interesse:

1) Documentazione e trasparenza: lo scopo medico previsto deve essere specificato preventivamente e documentato, così come il processo di sviluppo (es: selezione e uso di uno o più dataset, standard di riferimento, parametri, metriche, ecc.), considerando di adottare un approccio basato sul rischio in merito al livello di dettaglio della documentazione e alla conservazione dei registri utilizzati per lo sviluppo e la valutazione dei sistemi di IA;
2) Gestione del rischio: un approccio volto alla gestione dei rischi deve riguardare tutte le fasi del ciclo di vita di un sistema di IA – dallo sviluppo pre-commercializzazione alla sorveglianza post-vendita – e interessare, ad esempio, possibili bias cognitivi, problematiche legate alla complessità del modello (come l'*overfitting* e l'*underfitting*), nonché vulnerabilità e minacce legate alla cybersicurezza;
3) Validazione dei dati: la documentazione resa disponibile insieme al sistema di IA deve riguardare, fra l'altro, le caratteristiche dei dataset utilizzati e la dimostrazione delle performance tramite un processo di validazione dei dati, che abbia come riferimento un dataset esterno, come nell'ipotesi di dati su studi clinici randomizzati;
4) Qualità dei dati: gli sviluppatori dovrebbero considerare attentamente che i dati siano di una qualità sufficiente a sviluppare un sistema di IA con lo scopo prefissato e, in ogni caso, ricorrere a *pre-release* per garantire che tali sistemi non amplifichino pregiudizi ed errori già esistenti e noti;
5) Privacy e protezione dei dati personali: gli sviluppatori dovrebbero includere sin dalle prime fasi (*by design)* misure atte a garantire la privacy e la protezione dei dati personali, il che presuppone di avere una buona conoscenza del quadro normativo applicabile e le opportune competenze per implementare quanto stabilito all'interno dei sistemi di IA. In tal senso, sarebbe opportuno prevedere un *compliance programme* che attesti e assicuri che le pratiche sulla privacy e la cybersicurezza siano state prese in considerazione, sia per affrontare i potenziali danni, sia per le attività di *enforcement*;
6) Coinvolgimento e collaborazione: durante le prime fasi di sviluppo di una tecnologia di IA è importante considerare l'implementazione di piattaforme informative che facilitino il coinvolgimento e la collaborazione dei principali stakeholder, ricorrendo, ove opportuno, a un approccio di co-regolamentazione a beneficio di tutte le parti interessate.

1.7. Analisi dei punti di contatto e di contrasto delle iniziative in materia di IA a livello internazionale

Avendo delineato le principali iniziative dei fora e di altri organismi internazionali, appare ora opportuno soffermarsi su quei punti che si sono rintracciati diffusamente nei documenti programmatici e nelle linee guida analizzate e, al contempo, quelli che potrebbe essere i punti di divergenza o di contrasto tra le stesse. Innanzitutto, va evidenziato che sino ad oggi si è teso maggiormente a emanare principi guida da tradursi in misure di implementazione più pragmatiche, lasciando un certo margine di flessibilità ai destinatari, anche se in qualche occasione sono state previste azioni concrete a supporto (che hanno assunto spesso la forma del codice di condotta o di raccomandazioni più o meno granulari).

Va evidenziato che sino ad oggi si è teso maggiormente a emanare principi guida da tradursi in misure di implementazione più pragmatiche, lasciando un certo margine ai destinatari, anche se in qualche occasione sono state previste azioni concrete a supporto

Altro elemento da precisare in via preliminare è che tra le iniziative a livello di policy sin qui tracciate, non si focalizzano specificamente – se non in un caso – tutte quelle precedenti il 2023 non si focalizzano sulle opportunità e i rischi dell'IA generativa. Inoltre, in virtù della forma giuridica dei consessi internazionali che hanno emanato tali principi e linee guida, essi non possono che puntare sull'autorevolezza di quanto stabiliscono, configurandosi in atti di *soft law* e quindi non giuridicamente vincolanti.

Ad ogni modo, possono essere rintracciati – seppur con alcune differenze – i seguenti punti di contatto:

- Focus sulla comprensione dei rischi esistenti e potenziali connessi ai sistemi di IA (tra cui: privacy, protezione dei dati personali, diritti umani, bias, cybersecurity, proprietà intellettuale e disinformazione), nonché sull'adozione di misure di mitigazione adeguate sin dalle prime fase di sviluppo dei sistemi di IA e successivamente durante l'intero ciclo di vita degli stessi;
- Cooperazione internazionale, in particolare enfatizzando il ruolo delle organizzazioni internazionali e, come nel caso dell'Interim Report dell'*AI Advisory Body* dell'ONU e – con le opportune differenze – della Dichiarazione di Bletchley Park, anche di nuove iniziative e consessi che si occupino di fornire indicazioni sulla governance dell'IA;
- Inclusività ed equità, in alcuni casi (ONU e World Bank) esplicitamente rivolte a popolazioni a basso reddito e a gruppi meno rappresentati, nonché ad aree del globo tendenzialmente meno sviluppate;
- Sviluppo di standard etici e tecnici, nonché di framework per la gestione dei rischi, che rendano l'IA sicura e affidabile durante l'intero ciclo di vita;
- Centralità dell'IA funzionale per garantire l'interesse pubblico e il benessere individuale e collettivo;
- Governance dell'IA legata fortemente alla *data governance* e alla qualità dei dati (anche se non vi è sempre un riferimento esplicito, se non nel caso di ONU, OCSE e, con un'attenzione particolare, della World Bank);
- Coinvolgimento dei cittadini e della società civile e attenzione sulla collaborazione inclusiva e democratica multi-stakeholder;
- Sviluppo di un'IA umano-centrica;
- Trasparenza, spiegabilità e interoperabilità dei sistemi di IA;
- Accompagnamento dello sviluppo tecnologico con la costruzione delle competenze e delle infrastrutture necessarie al suo impiego e funzionamento;
- Importanza degli investimenti nella ricerca per migliorare sicurezza, affidabilità, trasparenza e interoperabilità dei modelli alla base dei sistemi di IA;

- Garanzia che sia sempre possibile l'intervento e la supervisione dell'essere umano.

Se questi sono i principali aspetti su cui si registra un diffuso consenso, quelli di seguito esplicitati costituiscono punti rispetto ai quali si registra una divergenza di vedute:

- Modelli *open-source*: sul tema, come evidenziato chiaramente dall'Interim Report dell'*AI Advisory Body* dell'ONU, risulta ancora acceso il dibattito sulla potenziale funzione abilitante o, diversamente, di rischio circa il ricorso a un'IA libera e aperta. Di fatto, solo il comitato consultivo dell'ONU sembra schierarsi a favore di un simile approccio, che naturalmente potrebbe favorire un più rapido processo di *catching-up* da parte dei Paesi meno avanzati, mentre in tutti gli altri documenti analizzati non vi è alcun riferimento;
- Rafforzamento della regolamentazione in materia di IA: il tema è assolutamente centrale e ciò si evince anche dalla differenza di approccio che emerge dalle iniziative esaminate. In particolare, l'ONU promuove un framework normativo vincolante in materia di IA, senza però escludere che si possano affiancare anche norme non vincolanti, e risulta dello stesso parere anche la World Bank. L'OCSE aggiunge di sviluppare parallelamente ambienti controllati in cui testare le soluzioni di IA. Il G7, nell'ambito del Processo di Hiroshima sull'IA, sembra puntare maggiormente sulla collaborazione internazionale per l'adozione di standard tecnici e buone pratiche per garantire la sicurezza e l'affidabilità dell'IA, mentre la Dichiarazione di Bletchley Park si riferisce a un modello regolatorio proporzionato e pro-innovazione che sia focalizzato essenzialmente su principi e codici di condotta flessibili;
- Maggiore attenzione sull'aspetto di sostenibilità ambientale ed energetica dell'IA: sul punto, è possibile rilevare approcci differenti in quanto in sede OCSE si sostiene in via generale di sviluppare l'IA in modo che arrechi benefici alle persone, al pianeta e che promuova lo sviluppo sostenibile, il G7 spinge a sviluppare sistemi avanzati di IA per affrontare il cambiamento climatico e un'indicazione analoga è fornita dall'OMS, mentre la World Bank mantiene un focus incentrato sostanzialmente sulla persona. L'impegno maggiore in tal senso sembra essere mostrato dall'Interim Report ONU che ritiene opportuno promuovere la misurazione e il tracciamento dell'impatto dell'IA in termini di sostenibilità e di necessità di risorse energetiche;
- Il riferimento esplicito verso l'importanza di identificare criteri e meccanismi per l'individuazione della responsabilità in caso di danni causati da decisioni basate su algoritmi di IA sembra emergere chiaramente nei documenti sin qui analizzati con riguardo a ONU, World Bank, OMS e OCSE. Diversamente, il G7 e la Dichiarazione di Bletchley Park non si soffermano esplicitamente su questo aspetto.
- Un punto rilevante, anche se finora piuttosto trascurato in tutti i documenti analizzati, compresi quelli che scaturiscono da consessi di Paesi like-minded come il G7, riguarda l'adesione a principi e valori democratici, ad esempio prevedendo precisi limiti alla censura dei governi sia rispetto ai dati di input che di output dei modelli. Un tema che è emerso prepotentemente con la rapida diffusione unita a una larga accessibilità dei *Large Language Model* e i conseguenti provvedimenti restrittivi di diverse nazioni, a cominciare dalla Cina.

2. CONSIGLIO D'EUROPA: VERSO LA PRIMA CONVENZIONE INTERNAZIONALE IN MATERIA DI IA

Come visto in chiusura del paragrafo precedente, nelle linee guida e nei documenti programmatici analizzati non si registrano considerazioni di merito rispetto all'adesione a principi e valori democratici nello sviluppo, nella distribuzione e nell'utilizzo di soluzioni di IA. Questo aspetto è invece centrale nella bozza della prima convenzione internazionale sull'intelligenza artificiale ("*Draft Framework Convention On Artificial Intelligence, Human Rights, Democracy And The Rule Of Law*"), attualmente in discussione presso il Comitato sull'IA (CAI) del Consiglio d'Europa[10], il cui iter dovrebbe concludersi entro il mese di maggio.

Il CAI, riunitosi per la prima volta a Roma nell'aprile 2022, è stato istituito dal Comitato dei Ministri del Consiglio d'Europa e ha una serie di prerogative connesse all'IA, fra cui quella di stabilire un framework normativo che si applichi allo sviluppo, alla progettazione e all'applicazione dell'IA, potendo comprendere strumenti legali sia di natura vincolante, sia non vincolante. Nello svolgimento di queste attività, tale organo deve tenere in debita considerazione questioni connesse al genere, alla minore età, ai diritti dei bambini e delle persone con disabilità, nonché tendere a rafforzare il ruolo della società civile, oltre a contribuire al raggiungimento (o alla revisione) dell'Agenda ONU 2030 per lo Sviluppo Sostenibile, con particolare riguardo ai Goal 5 (*Gender Equality*) e 16 (*Peace, Justice and Strong Institutions*).

La composizione del CAI, anche in virtù della significatività connessa all'emanazione della prima convenzione internazionale in materia di IA, è piuttosto ampia e include – oltre i 46 Stati Membri del Consiglio d'Europa e il Presidente del CAI – alcuni partecipanti a cui non spetta diritto di voto, ma che possono partecipare agli incontri di discussione tramite propri rappresentanti. Tra questi ultimi, rilevano l'Unione Europa, i Paesi Osservatori del Consiglio d'Europa (Canada, Santa Sede, Giappone, Messico, USA), altri organismi internazionali impegnati sul tema (OSCE, OCSE, UNESCO e altre agenzie ONU), oltre a una serie di soggetti – qualificati come "osservatori" – tra cui Israele, GPAI (*Global Partnership on Artificial Intelligence*) e varie organizzazioni della società civile.

La composizione del CAI, anche in virtù della significatività connessa all'emanazione della prima convenzione internazionale in materia di IA, è piuttosto ampia e include – oltre i 46 Stati Membri del Consiglio d'Europa e il Presidente del CAI – alcuni partecipanti a cui non spetta diritto di voto, ma che possono partecipare agli incontri di discussione tramite propri rappresentanti

Passando al contenuto della bozza di convenzione, così come da ultimo modificata il 18 dicembre scorso[11], essa è strutturata in 36 articoli, di cui una parte sostanziosa dedicata ai principi relativi alle attività che si posizionano durante l'intero ciclo di vita dei sistemi di IA (e ai rimedi applicabili in caso di violazione degli stessi), oltre che con riguardo all'implementazione della convenzione e

[10] https://www.coe.int/en/web/artificial-intelligence/cai.
[11] https://rm.coe.int/cai-2023-28-draft-framework-convention/1680ade043.

ai meccanismi di follow-up e cooperazione tra le Parti. Più nel dettaglio, all'art. 1 viene specificato l'obiettivo della convenzione: assicurare che durante l'intero ciclo di vita dei sistemi di IA siano garantiti i diritti umani, la democrazia e lo stato di diritto, attraverso un *risk-based approach* da attuarsi mediante l'adozione (o il mantenimento) di misure legislative, amministrative o di altra natura.

Proseguendo, se per un verso è pacifico – quantomeno ai fini della convenzione in esame – quale sia la definizione di riferimento di "sistema di IA"[12], non può dirsi lo stesso per il rispettivo ambito di applicazione. Difatti, all'art. 3 sono presenti diverse formulazioni alternative circa l'applicabilità delle norme della convenzione al settore pubblico e/o privato, alle attività di ricerca e a quelle connesse alla sicurezza nazionale e alla difesa. Infatti, secondo le notizie pubblicate e i comunicati diffusi nelle ultime settimane, sembra esservi un dibattito piuttosto incalzante attorno a questa disposizione della convenzione e, in particolare, sull'esclusione del settore privato e delle attività riconducibili alla sicurezza nazionale. Naturalmente, una simile discussione sta interessando anche l'UE, con la Commissione che sembrerebbe difendere l'inclusione del settore privato, insieme a quello pubblico, mentre diversi Stati Membri starebbero richiedendo maggiore flessibilità sull'ambito di applicazione, al fine di salvaguardare un traguardo storico così importante. Ad ogni modo, se così fosse, la prima convenzione al mondo sull'IA si applicherebbe unicamente al settore pubblico. In tal senso, la prossima riunione plenaria del CAI, prevista per metà marzo, potrebbe risultare decisiva.

Secondo le notizie pubblicate e i comunicati diffusi nelle ultime settimane, sembra esservi un dibattito piuttosto incalzante attorno a questa disposizione della convenzione e, in particolare, sull'esclusione del settore privato e delle attività riconducibili alla sicurezza nazionale

In chiusura della parte più generale del testo, vi sono due clausole relative alla protezione dei diritti umani e all'integrità dei processi democratici e il rispetto dello stato di diritto. In particolare, su quest'ultimo aspetto viene sancita l'importanza di misure volte a tutelare, da un lato, l'integrità, l'indipendenza e l'efficacia delle istituzioni e dei processi democratici, dall'altro, la partecipazione ai processi democratici e l'accesso equo al dibattito pubblico. Invece, è ancora in discussione il riferimento esplicito al tema delle influenze o manipolazioni esterne dannose e malevoli, che richiama naturalmente i concetti di deepfake e disinformazione.

Il terzo capitolo si focalizza sui principi applicabili ai sistemi di IA ai fini della presente convenzione, ossia:

- Dignità umana e autonomia individuale;
- Trasparenza e supervisione;
- Accountability e responsabilità;
- Uguaglianza e non discriminazione;
- Privacy e protezione dei dati personali[13];

[12] *Si v. supra, par. 1.2.*
[13] *La formulazione di tale principio è ancora in fase di discussione.*

- Salvaguardia della salute e dell'ambiente;
- Affidabilità e fiducia[14];
- Innovazione sicura.

Quest'ultimo principio, seppur non ancora consolidato con riguardo alla terminologia da utilizzare, impegnerebbe ciascuna Parte a incoraggiare l'innovazione e, contemporaneamente, evitare impatti negativi sui diritti umani, la democrazia e lo stato di diritto. Le norme che si occupano specificamente dei rimedi in caso di violazione dei principi (artt. 14-15) e di valutazione e mitigazione dei rischi (art. 16) sono fra i più in discussione. Più nel dettaglio, con riguardo al primo aspetto si dovrà optare per una formulazione più generica ("*violations of human rights*"), oppure per una più puntuale ("*unlawful harm or damage to the rights of individuals and legal persons*"), in quanto al requisito della violazione. Sul punto, un'ulteriore proposta interessante concerne la possibilità effettiva per la persona interessata dalla violazione di presentare un reclamo a un'autorità pubblica, in conformità col diritto interno della Parte.

Con riguardo al secondo aspetto, si propone di adottare una serie corposa di misure – secondo il *risk-based approach* sopra richiamato – per identificare, valutare, prevenire e mitigare i rischi e gli impatti sui diritti umani, la democrazia e lo stato di diritto relativamente alla progettazione, sviluppo, utilizzo e dismissione dei sistemi di IA. A titolo esemplificativo, rilevano: i) l'integrazione del punto di vista di tutti gli stakeholders interessati, comprese le persone i cui diritti possono essere potenzialmente colpiti dalla progettazione, dallo sviluppo, dall'uso e dalla dismissione dei sistemi di IA; ii) pubblicare le informazioni sugli sforzi intrapresi per identificare, valutare, mitigare e prevenire i rischi e gli impatti negativi.

La quarta parte (art. 17 ss.) – sostanzialmente consolidata – copre i principi applicabili all'implementazione della convenzione, tra cui: la non discriminazione; il rispetto dei diritti dei minori e delle persone con disabilità; la promozione di un'adeguata alfabetizzazione digitale e di un maggiore livello di competenze digitali per tutti i segmenti della popolazione; l'armonizzazione con altri strumenti giuridici a cui la singola Parte è obbligata; la possibilità di prevedere misure di protezione più ampie rispetto a quanto stabilito nella convenzione. Interessante, anche se ancora in discussione, il riferimento specifico alla tutela dei *whistleblowers* e al ruolo delle consultazioni pubbliche come strumenti per prevenire e mitigare i rischi e gli impatti sin qui richiamati.

Allo stesso modo, risulta particolarmente rilevante la parte dedicata al meccanismo di follow-up e alla cooperazione (art. 24 ss.), poiché si prevede che le Parti si consultino periodicamente – attraverso una Conferenza delle Parti – per facilitare l'implementazione e il rispetto effettivo della convenzione, anche proponendo integrazioni o emendamenti alla stessa, come pure segnalando problematiche e fornendo raccomandazioni inerenti l'interpretazione e l'applicazione delle norme ivi contenute. Tra le recenti aggiunte non ancora consolidate va evidenziato il riferimento all'UE che, nei settori di sua competenza, potrebbe avvalersi del diritto di voto in tale consesso, con un numero di voti pari al numero degli Stati Membri che siano anche contraenti della convenzione, a meno che gli stessi non siano interessati a loro volta ad avvalersene.

[14] La formulazione di tale principio è ancora in fase di discussione e richiamerebbe misure atte a garantire, ad esempio, la qualità, l'integrità e la sicurezza dei dati, la cybersicurezza, la governance, l'accuratezza e le performance dei sistemi di IA durante il rispettivo ciclo di vita.

Si prevede che le Parti si consultino periodicamente – attraverso una Conferenza delle Parti – per facilitare l'implementazione e il rispetto effettivo della convenzione, anche proponendo integrazioni o emendamenti alla stessa, come pure segnalando problematiche e fornendo raccomandazioni inerenti l'interpretazione e l'applicazione delle norme ivi contenute

Nel testo, viene stressata particolarmente la cooperazione tra le Parti, le quali dovrebbero scambiarsi informazioni rilevanti sugli aspetti relativi all'IA che possano avere un significativo impatto positivo o negativo sul godimento dei diritti umani, sul funzionamento della democrazia e sull'osservanza dello stato di diritto, ivi compresi rischi ed effetti che possano emergere in contesti di ricerca. Per gli stessi fini, si incoraggia anche il supporto delle Parti verso Stati non contraenti della convenzione, così come di attori non statali. In ultima istanza, si prescrive che ciascuna Parte debba stabilire o designare uno o più meccanismi, caratterizzati da indipendenza e imparzialità, nonché dotati dei necessari poteri, competenze e risorse, per sorvegliare l'osservanza degli obblighi derivanti dalla convenzione.

3. STATI UNITI E CINA: IL DIVERSO APPROCCIO ALL'IA SULLE DUE SPONDE DEL PACIFICO

L'IA si caratterizza per essere una delle tecnologie più potenti dei nostri tempi, una promessa straordinaria in termini di innovazione ed efficienza, uno strumento rivoluzionario in grado di avere impatti profondi sulla società, sull'economia, finanche sulla vita degli individui e che pertanto impone l'individuazione non solo di tutte le opportunità da cogliere ma anche di tutte le potenziali criticità che ad essa si accompagnano al fine di comprenderle e governarle in maniera efficace.

I paragrafi precedenti hanno evidenziato la straordinaria attenzione prestata al tema dell'IA e l'importante serie di linee guida e documenti programmatici in materia di IA adottate dai principali organismi internazionali con l'obiettivo di coniugare l'esigenza di favorire lo sviluppo dei sistemi di IA con la necessità, altrettanto urgente, di assicurare la tutela dei diritti fondamentali degli individui.

Seppur in un contesto generale in cui si mira a comprendere e governare il fenomeno dell'IA, appaiono comunque diversi gli approcci seguiti e gli intenti perseguiti dai governi ed in particolare da Cina e Stati Uniti. Se, infatti, la Cina, pur con la specifica attenzione rivolta a scongiurare rischi di potenziali conflitti tra sistemi di IA e l'ordine costituito politico e sociale, rappresenta uno dei primi paesi ad avere adottato una specifica legislazione in materia, gli USA al contrario, continuano, nonostante le importanti evoluzioni in atto di cui si darà conto nel corso dell'analisi, a non disporre di un framework regolamentare dell'IA e a conservare un approccio più incentrato sull'assunzione di impegni volontari da parte delle grandi aziende tecnologiche nel settore dell'IA attraverso il quale si punta a non frenare l'innovazione.

La disciplina cinese, ancorché non organica, si basa su diversi atti normativi, in particolare l'Algorithm Recommendation Regulation, entrato in vigore il 1° marzo 2022, il Deep Synthesis Regulation entrato in vigore il 10 gennaio 2023 ed il Regolamento sull'IA generativa entrato in vigore lo scorso 15 agosto 2023.

La disciplina cinese, ancorché non organica, si basa su diversi atti normativi, in particolare l'Algorithm Recommendation Regulation, entrato in vigore il 1° marzo 2022, il Deep Synthesis Regulation entrato in vigore il 10 gennaio 2023 ed il Regolamento sull'IA generativa entrato in vigore lo scorso 15 agosto 2023

Ciò che emerge, a livello generale, è che, nonostante la comunanza di specifiche preoccupazioni (ad es. quelle relative al fatto di rendere visibile il contenuto prodotto dall'IA), esistono importanti differenze rispetto ai Paesi occidentali nell'approccio governativo particolarmente attento a chiarire che i sistemi di IA non debbano confliggere con l'ordine politico e sociale e a legittimare l'utilizzo di sistemi di sorveglianza di massa e di social scoring, secondo un approccio del tutto inammissibile nel sistema europeo ed americano.

L'Algorithm Recommendation Regulation, in particolare, persegue il fine espresso di promuovere i valori fondamentali del socialismo, salvaguardare la sicurezza nazionale e gli interessi pubblici sociali, proteggere i diritti e gli interessi legittimi dei cittadini e delle imprese e promuovere il sano e ordinato sviluppo di servizi di informazione su Internet. In tale logica, il regolamento prescrive che gli algoritmi di raccomandazione non perturbino l'ordine economico e sociale e non diffondano informazioni vietati dalle leggi e dai regolamenti amministrativi vigenti ed impone ai fornitori di servizi di raccomandazione di esaminare, valutare e verificare regolarmente i meccanismi, i modelli, i dati, i risultati e la sicurezza delle applicazioni degli algoritmi col divieto espresso di impostare modelli di algoritmi che inducano dipendenza dell'utente, consumo eccessivo o altre violazioni di leggi, regolamenti o dell'etica. A tali divieti si aggiunge quello di manipolare account utente, commentare o inoltrare falsamente notizie, bloccare informazioni, raccomandare eccessivamente, manipolare elenchi o ordinare o controllare i risultati delle ricerche, interferire nella presentazione di informazioni e mettere in atto comportamenti che influenzano l'opinione pubblica online o eludono i controlli. Specifica attenzione è riservata ai diritti degli utenti e in particolare, di consumatori, lavoratori, minori ed anziani; rispetto ai minori, nello specifico, è chiaramente vietato ai fornitori di servizi di raccomandazione fornire informazioni che potrebbero indurli a imitare comportamenti non sicuri o a violare l'etica sociale, ad avere cattive abitudini o informazioni che potrebbero influire sulla salute fisica e mentale degli stessi, mentre con riguardo agli anziani il focus è sulle frodi e sulla necessità di scongiurarne i relativi rischi.
Lo stesso regolamento ha poi introdotto uno strumento di controllo importante, il registro degli algoritmi che si concentra prevalentemente sugli algoritmi di raccomandazione e sul ruolo che essi giocano nella diffusione delle informazioni, prescrivendo che gli stessi non minino la sicurezza nazionale e l'interesse pubblico sociale e che forniscano una motivazione, una giustificazione, nel caso di violazione dei legittimi interessi degli utenti. Nello specifico, è prevista la registrazione, da parte di fornitori di servizi di raccomandazione di algoritmi di "public opinion characteristics" o che possiedono capacità di mobilitazione sociale, entro 10 gg. dalla data di fornitura, mediante comunicazione di una serie di informazioni, nonché la messa in campo di procedure di valutazione della sicurezza e di condotte collaborative nei confronti delle autorità, anche in termini di fornitura di dati tecnici. Specifici obblighi di riservatezza sono previsti in relazione alla tutela della privacy e dei segreti aziendali.
Specifica attenzione è rivolta ai potenziali comportamenti monopolistici delle piattaforme e alle questioni sociali più scottanti, tra cui il ruolo che gli algoritmi di dispacciamento svolgono nel creare condizioni di lavoro pericolose per i fattorini cinesi.

La Cyberspace Administration of China (CAC) ad Agosto 2022 ha fornito la prima pubblicazione di circa 30 registrazioni compiute dalle maggiori società cinesi di piattaforme internet, tra cui Tencent, Alibaba e Bytedance, che hanno fornito una serie di informazioni - non tutte disponibili, come, ad esempio, l'Algorithm Security Self-Assessment di cui si richiede il caricamento alle aziende ma che non è accessibile al pubblico, oppure l'"Algorithm Strategy," e l'"Algorithm Risk and Prevention Mechanism - tra cui i "Fondamenti dell'algoritmo" ed il "Meccanismo operativo dell'algoritmo" che tuttavia appaiono di così elevato tecnicismo da risultare difficilmente intellegibili.

Nel luglio 2023 è entrato in vigore il Deep Synthesis Regulation, emanato congiuntamente il 25 novembre 2022 da tre autorità di regolamentazione del governo centrale - la Cyberspace Administration of China (CAC), il Ministero dell'Industria e dell'Information Technology (MIIT) e il

Ministero della Pubblica Sicurezza (MPS). Si tratta di un regolamento di 25 articoli suddivisi in cinque capitoli, che fissa le regole sulla gestione dei dati e delle tecnologie di sintesi profonda, esortando i fornitori di servizi di sintesi profonda e i promotori di tale tecnologia a garantire il rispetto delle leggi e dei regolamenti prescrivendo una serie di obblighi e responsabilità in alcuni ambiti chiave. La CAC ha dichiarato che le disposizioni in questione perseguono il duplice scopo di incoraggiare la crescita sana delle imprese e di ridurre i rischi associati alle attività offerte dalle piattaforme che utilizzano il deep learning o la realtà virtuale per modificare qualsiasi materiale online. Specifica attenzione è rivolta al fenomeno noto come "deepfake", una combinazione di "deep learning" e "fake", una tecnologia che utilizza potenti tecniche di apprendimento automatico e di intelligenza artificiale per modificare o sintetizzare informazioni visive e audio che possono così produrre contenuti del tutto verosimili ma falsi.

Destinatari di tale disciplina sono da un lato i "fornitori di servizi di sintesi profonda", ossia le aziende che offrono servizi di sintesi profonda e quelle che forniscono loro assistenza tecnica e, dall'altro, gli "utenti di servizi di sintesi profonda", ossia le organizzazioni e gli individui che utilizzano i servizi di sintesi profonda per creare, duplicare, pubblicare o trasferire informazioni. Ebbene, ai fornitori di tali servizi sono dettate una serie di prescrizioni e riconosciute specifiche responsabilità in diversi ambiti tra cui: a) la sicurezza dei dati e la protezione delle informazioni personali, rispetto ai quali i fornitori sono chiamati ad adottare le misure necessarie per la protezione dei dati personali in base alla normativa vigente, a predisporre e migliorare i sistemi di gestione per la formazione del personale, la revisione degli algoritmi, la registrazione degli utenti, la sicurezza dei dati, la protezione dei minori e la protezione delle informazioni personali; b) trasparenza, in relazione alla quale i fornitori di servizi di sintesi profonda sono tenuti a stabilire linee guida, criteri e processi per riconoscere le informazioni false o dannose e per gestire i casi in cui gli utenti producano materiale falso o dannoso utilizzando la tecnologia di sintesi profonda. Gli stessi fornitori sono inoltre tenuti a formulare e divulgare regole di gestione ed accordi di piattaforma, a rafforzare la gestione dei contenuti sintetici approfonditi, adottare meccanismi di revisione dei dati sintetici, predisporre un database per l'identificazione di informazioni illegali e false e la registrazione dei log di rete pertinenti; c) gestione dei contenuti ed etichettatura. Le Deep Synthesis Provisions prevedono la creazione di un meccanismo per la dissipazione delle fake news, in modo che quando i servizi di sintesi profonda vengono utilizzati per produrre, copiare, pubblicare e diffondere informazioni false, i fornitori di servizi di sintesi profonda siano tenuti ad adottare misure per dissipare tali notizie, a tenere un registro e a segnalarle alle autorità competenti (come l'Internet Information Department). Inoltre, le previsioni in questione rendono obbligatoria l'aggiunta di etichette o tag sulle informazioni generate dall'uso di tecnologie di sintesi profonda (inclusa la simulazione della voce, la conversazione o la scrittura intelligente che simula lo stile di una persona reale, la sintesi dell'immagine del volto o la manipolazione del volto); d) sicurezza tecnica. Al fine di accrescere la sicurezza degli utenti su base tecnica, i fornitori di servizi di sintesi profonda sono chiamati a rivedere periodicamente gli algoritmi e a condurre valutazioni di sicurezza quando forniscono modelli, sagome e altri strumenti con la funzione di modifica del volto, della voce e di altre informazioni biometriche, oppure oggetti, scene e altre informazioni non biometriche che possono coinvolgere la sicurezza nazionale, l'immagine nazionale, gli interessi nazionali e gli interessi pubblici.

Si prevede infine che i fornitori di servizi di sintesi profonda con capacità di mobilitare l'opinione pubblica o la società debbano seguire le procedure di deposito previste dalle disposizioni sulla

gestione delle raccomandazioni algoritmiche per i servizi di informazione su Internet sopra descritte.

Il 13 luglio 2023 sono state pubblicate una serie di linee guida per regolamentare il settore dell'AI generativa, le cd. "Misure provvisorie per la gestione dei servizi di intelligenza artificiale generativa", entrate in vigore il 15 agosto al fine di promuovere lo sviluppo sano e l'uso regolamentato dell'IA generativa, preservare la sicurezza nazionale e l'interesse pubblico della società e proteggere i diritti e gli interessi legittimi di cittadini, persone giuridiche e altre organizzazioni.

Il fine espresso è quello di "incoraggiare l'uso innovativo dell'AI generativa in diversi settori e campi, che generino contenuti positivi, sani ed edificanti di alta qualità", nonché sostenere imprese e istituti di formazione e ricerca, istituzioni culturali pubbliche e professionali che contribuiscono all'innovazione tecnologica dell'AI generativa.

Entrando nel merito delle prescrizioni fissate, le regole varate prevedono innanzitutto che per la fornitura di servizi di intelligenza artificiale generativa al pubblico sia richiesta ed ottenuta una licenza per operare, che tali servizi aderiscano ai “valori fondamentali del socialismo“, che i fornitori di servizi di AI generativa conducano verifiche di sicurezza e registrino i propri algoritmi presso il governo in conformità con il "Regolamento sulla gestione delle raccomandazioni sugli algoritmi dei servizi di informazione su Internet", qualora i propri servizi siano in grado di influenzare l'opinione pubblica o siano in grado di "mobilitare" il pubblico". Ciò posto, ai fornitori di servizi che individuino contenuti illegali, è posto l’obbligo di adottare provvedimenti per interromperne la generazione, migliorare il proprio algoritmo e segnalare il materiale all’autorità competente.
Specifica attenzione è rivolta alla tutela della proprietà intellettuale. Le regole in esame, infatti, prevedono l’obbligo espresso per i fornitori e gli utenti dei servizi di AI generativa di "rispettare i diritti di proprietà intellettuale e l'etica commerciale, proteggere i segreti commerciali e non praticare il monopolio o la concorrenza sleale sfruttando algoritmi, dati, piattaforme e altri vantaggi".

Se con le iniziative appena descritte la Cina ha scelto un approccio molto invasivo, gli Usa, al contrario, seppur con un approccio in chiara evoluzione, come già anticipato sopra, hanno tradizionalmente preferito limitare l’adozione di interventi prescrittivi favorendo, al contrario, l’adozione di modelli di sviluppo e condotta volontariamente accettati dalle aziende attive nel settore dell’IA nell’intenzione di non ostacolare lo sviluppo tecnologico e l’innovazione e di rendere l’America leader mondiale nella ricerca, lo sviluppo e l’’utilizzo di sistemi di IA.

Gli Usa, al contrario, seppur con un approccio in chiara evoluzione, hanno tradizionalmente preferito limitare l’adozione di interventi prescrittivi favorendo, al contrario, l’adozione di modelli di sviluppo e condotta volontariamente accettati dalle aziende attive nel settore dell’IA nell’intenzione di non ostacolare lo sviluppo tecnologico e l’innovazione e di rendere l’America leader mondiale nella ricerca, lo sviluppo e l’utilizzo di sistemi di IA

Con questo spirito, nel 2020 è stato varato il National AI Initiative Act (NAIIA), relativo al periodo 2020-2025, in cui è evidenziato l'enorme potenziale innovativo dell'IA per ogni settore socio-economico ed affermata l'ambizione del Governo federale di giocare un ruolo centrale nella ricerca, nello sviluppo e nelle attività formative in materia di IA attraverso il coordinamento e la collaborazione tra governo, accademia e settore privato. I macro-obiettivi perseguiti, in particolare, consistono nell'assicurare la leadership statunitense nella ricerca e nello sviluppo dell'IA, guidare il mondo nello sviluppo e nell'utilizzo di sistemi di intelligenza artificiale affidabili ("trustworthy") nel settore pubblico e privato, massimizzare i benefici dell'IA per tutti gli americani e preparare la forza lavoro presente e futura degli Stati Uniti all'integrazione dei sistemi di intelligenza artificiale in tutti i settori dell'economia e della società. Al fine di raggiungere tali obiettivi, vengono individuate una serie corposa di iniziative - con al centro il National Artificial Intelligence Initiative Office - tese ad incrementare le risorse finanziarie, supportare la ricerca (di cui vengono individuati specifici ambiti rilevanti ,anche in maniera indiretta, per lo sviluppo dell'IA), la formazione, la consapevolezza, lo sviluppo di standard volontari e la creazione di alleanze strategiche. Dal punto di vista della Governance, a supporto del National Artificial Intelligence Initiative Office è prevista l'azione del National Artificial Intelligence Advisory Committee.

Tra le diverse attività messe in campo dall'amministrazione Biden sul fronte della regolazione dell'IA, ne spiccano in particolare due, per solidità, ampiezza di approccio e interlocuzione con gli stakeholder: il Blueprint for an AI Bill of Rights e l'AI Risk Management Framework. La prima è stata promossa direttamente dalla Casa Bianca e dal suo Ufficio per le politiche della scienza e della tecnologia, la seconda dal National Institute of Standards and Technology (NIST), agenzia tecnica del Dipartimento del Commercio. Entrambe condividono un lungo percorso di confronto con gli stakeholder, iniziato nel 2021, poco dopo l'arrivo della nuova amministrazione. Il Blueprint è stato varato nell'ottobre del 2022, il Framework nel gennaio del 2023. Seppure basandosi sullo stesso approccio per un IA responsabile, hanno finalità piuttosto diverse. Il primo definisce delle linee guida etiche, basate su 5 principi: sicurezza ed efficacia; protezioni da discriminazioni algoritmiche; privacy; notifica e spiegazione; alternative umane, attenzione e ripiego. Il secondo punta a dare delle direttive tecniche a chi sviluppa e usa prodotti AI per gestire al meglio i potenziali rischi emergenti durante il loro intero ciclo di vita, articolandole in quattro funzioni (GOVERN, MAP, MEASURE e MANAGE). Sono tuttavia entrambi documenti non vincolanti e dunque rimessi alla libera volontà degli attori ai quali sono destinati. Dunque, si tratta al più di *soft regulation*, più o meno paragonabile agli orientamenti etici per un'IA affidabile, definiti a livello europeo già nel 2019. Uno step certamente significativo ma anche un semplice antipasto dell'AI Act, la portata principale dell'impianto normativo comunitario.

Al fine di favorire l'adozione di standard e metodi condivisi senza tuttavia aderire ad una logica di tipo prescrittivo, nel luglio 2023 l'amministrazione Biden ha annunciato di aver ottenuto da 7 importanti aziende attive nel campo dell'IA, ossia Amazon, Anthropic, Google, Inflection, Meta, Microsoft ed OpenAI, l'impegno volontario a contribuire allo sviluppo sicuro, protetto e trasparente dell'IA. Gli impegni assunti da tali compagnie, in particolare, ruotano intorno a tre parole chiave, safety, security e trust e si sostanziano nel: a) garantire la sicurezza dei prodotti prima di renderli disponibili al pubblico attraverso la realizzazione di test dei sistemi realizzati da soggetti indipendenti e la condivisione tra industria, governi, società civile ed accademia sulla gestione dei rischi dell'IA; b) costruire sistemi di IA che pongano al sicurezza al primo posto attraverso investimenti nella cybersecurity e la predisposizione di meccanismi di rapida

individuazione, segnalazione e reportistica delle eventuali vulnerabilità riscontrate; c) guadagnare la fiducia del pubblico mediante lo sviluppo di meccanismi che assicurino la consapevolezza degli utenti circa la produzione di contenuti da parte di sistemi AI, la comunicazione delle capacità, dei limiti e delle aree di utilizzo appropriate e inappropriate dei propri sistemi di IA, la prioritizzazione della ricerca sui rischi sociali connessi all'IA e l'impegno a sviluppare e distribuire sistemi avanzati di IA per contribuire ad affrontare le maggiori sfide della società (es. dalla prevenzione del cancro alla mitigazione del cambiamento climatico). Nel settembre scorso è stata annunciata l'assunzione di tali impegni volontari da parte di altre 8 grandi aziende - Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, and Stability - attive nel settore dell'IA.

Da ultimo, il 30 ottobre scorso, il Presidente Biden ha adottato l'"Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence", nel quale, partendo dalla constatazione delle straordinarie opportunità offerte dai sistemi di IA ma anche dei potenziali rischi connessi ad "irresponsible use" degli stessi e della volontà di governare lo sviluppo e l'utilizzo dell'IA, sono declinate 8 priorità e principi guida da osservare nello sviluppo ed utilizzo dei sistemi di IA e previsti adempimenti di importanza crescente a seconda del bene che si intende tutelare (in primis la sicurezza nazionale) a carico delle imprese che, di fatto, sembrano segnare l'inizio di un cambiamento rispetto all'approccio, volutamente poco restrittivo ed affatto prescrittivo, fino ad oggi seguito dagli USA. Alle imprese che sviluppano tecnologie avanzate di intelligenza artificiale, infatti, sono imposti obblighi di rendicontazione e, sulla base del "Defense production act", i produttori di grandi modelli di AI sono chiamati a fornire al governo informazioni importanti, tra cui indicazioni relative alle fasi di addestramento di nuovi modelli ed alle misure di sicurezza informatica utilizzate, ivi inclusi i test di red teaming, attraverso i quali si mira ad individuare i punti deboli nei modelli di intelligenza artificiale. Il tutto, al fine di individuare e monitorare le potenziali minacce che l'IA può rappresentare per la sicurezza nazionale, la salute pubblica e l'economia.

Entrando nello specifico di alcune delle misure previste, l'executive order evidenzia innanzitutto come i sistemi di IA debbano essere "safe and secure" e, in linea con gli impegni volontariamente assunti dalle 15 aziende sopra citate, precisa la necessità ineludibile che i sistemi di IA siano sottoposti a solide, affidabili, ripetibili e standardizzate valutazioni e a meccanismi per testare, comprendere e ridurre i rischi di questi sistemi prima che vengano utilizzati. Specifica attenzione è rivolta ai rischi più urgenti da affrontare per la sicurezza dei sistemi di IA anche per quanto riguarda la biotecnologia, la sicurezza informatica, le infrastrutture critiche e altri pericoli per la sicurezza nazionale, alla necessità di scongiurare il rischio di usi impropri o modifiche pericolose ed all'opportunità di definire meccanismi di etichettatura efficaci che diano consapevolezza circa la provenienza di contenuti da sistemi di IA. La seconda priorità consiste nel promuovere l'innovazione, la concorrenza e la collaborazione responsabili attraverso investimenti in istruzione e formazione, lo sviluppo di conoscenze e competenze indispensabili per utilizzare i sistemi di IA, l'attrazione di talenti intenzionati a restare negli USA grazie ai quali favorire lo sviluppo di sistemi di IA in America, la promozione (anche attraverso l'attenta analisi delle condotte messe in campo dalle grandi aziende attive nel settore) di un ecosistema ed un mercato equi, aperti e competitivi per l'IA e le tecnologie correlate che consenta anche ai piccoli sviluppatori e imprenditori di guidare l'innovazione.

Non meno importanti e, dunque, oggetto di specifica attenzione, gli impatti dell'IA sul mondo del lavoro. Rispetto a questo tema l'executive order da un lato esorta a scongiurare il rischio che

l'impiego dell'IA possa tradursi in violazioni dei diritti, peggioramento della qualità del lavoro, indebita sorveglianza dei lavoratori, riduzione della concorrenza sul mercato e nuovi rischi per la salute e la sicurezza e, dall'altro, evidenzia la necessità di apprestare interventi specifici tesi a fornire l'istruzione e la formazione necessarie a consentire a tutti i lavoratori di beneficiare delle opportunità offerte dall'IA. Al quarto punto viene inoltre assunto l'impegno – in linea con l'approccio seguito con l'emanazione del Blueprint for an AI Bill of Rights, l'AI Risk Management Framework e l'Ordine Esecutivo 14091 del 16 febbraio 2023 ("Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government") - ad assicurare che l'IA sia sviluppata ed utilizzata in maniera da non accrescere le discriminazioni e le diseguaglianze, nel pieno rispetto delle leggi e previa attenta valutazione tecnica dei singoli sistemi.

Partendo dalla constatazione del crescente impiego di sistemi di IA o di prodotti abilitati all'IA da parte degli individui, è ribadita la piena vigenza degli obblighi gravanti sulle imprese e chiaramente espressa la volontà di garantire l'osservanza delle leggi a tutela dei consumatori e, soprattutto con riguardo a settori critici come l'assistenza sanitaria, i servizi finanziari, l'istruzione, l'edilizia, la legge e i trasporti, dove gli errori o l'uso improprio dell'IA potrebbero danneggiare i pazienti, costare ai consumatori o alle piccole imprese, o mettere a repentaglio la sicurezza o i diritti, l'intenzione di predisporre opportune misure di salvaguardia contro le frodi, i pregiudizi involontari, le discriminazioni, le violazioni della privacy ed altri eventuali pregiudizi derivanti dall'IA. La sesta priorità dell'executive order si concentra sulla tutela della privacy, assolutamente cruciale e oggetto di specifico impegno da parte dell'amministrazione, ove si consideri la maggior facilità, garantita dall'IA, di estrarre, reidentificare, collegare, dedurre e agire su informazioni sensibili relative all'identità, alla posizione, alle abitudini e ai desideri delle persone.
In linea con il macro-obiettivo di rendere gli USA leader nel campo dell'IA, viene ribadita la volontà di mettere in campo iniziative per attrarre talenti attivi nel settore, di formare la forza lavoro esistente per comprendere benefici, opportunità e rischi legati all'impiego dell'IA nell'ambito della propria attività lavorativa e di agire al fine di favorire ed accelerare l'impiego dell'IA da parte del Governo federale rispetto al quale viene altresì evidenziata la necessità di accrescerne la capacità di regolamentare, governare e sostenere un uso responsabile dell'IA al fine di ottenere risultati migliori per gli americani. Tutte le priorità fin qui descritte trovano piena sintesi nell'ultima che si concentra sulla volontà di rendere gli USA leader globali nello sviluppo dei sistemi di IA e nella predisposizione e promozione a livello mondiale, attraverso intese e collaborazioni, aperte anche ai competitor, di misure di salvaguardia necessarie per un impiego responsabile degli stessi.

Se queste sono le priorità perseguite, l'executive order, come anticipato, contiene poi specifiche previsioni e prescrizioni che vanno a rafforzare gli obblighi in capo alle imprese operanti nel settore dell'IA. Ed infatti, si prescrive l'adozione da parte del Segretario al Commercio, di concerto con una serie di soggetti specificamente individuati, entro 270 giorni dalla data di adozione dell'order, di linee guida e best practice, con l'obiettivo di promuovere standard industriali condivisi, per lo sviluppo e l'implementazione di sistemi di IA sicuri, protetti e affidabili, nonché linee guida relative alle attività di test da condurre per consentire l'implementazione di sistemi sicuri, protetti e degni di fiducia. Nello stesso termine, al fine di mitigare i rischi di sicurezza, il Segretario all'Energia è invece chiamato a sviluppare un piano per lo sviluppo degli strumenti di valutazione dei modelli di IA del Dipartimento dell'Energia, anche utilizzando soluzioni già esistenti se del caso, per analizzare le capacità dell'IA di generare output che possano rappresentare minacce o pericoli nucleari, di non proliferazione, biologici, chimici, rischi per le infrastrutture critiche e di sicurezza energetica. Nella logica di garantire la tutela della difesa nazionale e delle

infrastrutture critiche, si prescrive al Segretario al Commercio, di individuare, entro 90 gg. dalla data di adozione dell'order, i modelli e i cluster di calcolo per i quali richiedere, alle aziende che sviluppano o dimostrano l'intenzione di sviluppare dual-use foundation models, una serie di informazioni da fornire su base continuativa tra cui quelle relative alle misure di protezione (fisiche e non) adottate per assicurare la sicurezza ed i test compiuti. Specifici obblighi di comunicazione in merito al possesso, l'ubicazione e la potenza di calcolo, gravano sulle aziende, gli individui o altre organizzazioni o entità che acquisiscono, sviluppano o possiedono un potenziale cluster di calcolo su larga scala tra quelli che saranno individuati. Lo stesso Segretario è chiamato ad individuare - e successivamente rivedere ove necessario - obblighi informativi puntuali nel caso di accordi con soggetti stranieri per l'addestramento di modelli su larga scala con potenziali capacità utilizzabili in attività malevole di tipo informatico nonché a fissare gli standard minimi che i fornitori statunitensi devono esigere ai rivenditori stranieri dei propri prodotti per verificare l'identità di un soggetto straniero che apre o mantiene un account. Specifici termini e task sono posti a carico del dipartimento dell'energia americano al fine di valutare come le AI possano contribuire ad attacchi biologici o chimici o a violazioni informatiche di infrastrutture critiche.

Molto rilevanti le previsioni tese ad assicurare trasparenza circa la circostanza che determinati contenuti siano stati generati dall'Intelligenza Artificiale, in particolare attraverso la predisposizione di un'indicazione chiara dell'origine dei contenuti prodotti attraverso l'impiego di modelli e software basati sull'Intelligenza Artificiale.

Al centro dell'attenzione anche la tutela della privacy, rispetto alla quale l'ordine esecutivo prevede l'adozione di una legge bipartisan che si concentri sulla prevenzione, attraverso lo sviluppo di metodologie che, ab origine, proteggano i dati personali senza tuttavia impattare sul processo di addestramento dei modelli AI. Rispetto al tema della commercializzazione dei dati, di cui viene riconosciuta ed esaltata la crescente rilevanza economica, l'order sollecita un'attenta analisi delle modalità di immissione di tali dati sul mercato nel tentativo, niente affatto semplice, di garantire che tale commercializzazione sia rispettosa degli standard etici e legali e della privacy dei cittadini.

Anche la promozione dell'eguaglianza e la tutela dei diritti civili trova ampio spazio nell'ordine esecutivo che punta a scongiurare il rischio di discriminazioni in settori essenziali quali giustizia, sanità ed edilizia residenziale. Nello specifico, nell'ambito della tutela del diritto alla casa, si prevede l'adozione di chiare linee guida anti-discriminazione ai proprietari e ai contractor federali, mentre nel contesto giuridico-processuale, si promuove invece lo sviluppo di best practices per l'utilizzo dell'IA in sentenze, procedimenti cautelari, fase preliminare e detenzione con l'obiettivo esplicito di scongiurare il rischio che l'origine geografica di un individuo o l'appartenenza ad una specifica etnia costituiscano causa di iniquità applicative.
Sempre nella logica di porre rimedi ad eventuali criticità applicative, si prevede l'invio di report al Dipartimento della Salute e dei Servizi Umani nella logica di segnalare problematiche e proporre rimedi in relazione all'uso non sicuro o dannoso di pratiche mediche che coinvolgono l'IA.

Rispetto al mondo del lavoro, pur partendo da un approccio fondamentalmente ottimistico che vede nell'IA un motore di sviluppo economico e sociale e non un elemento di concorrenza teso alla contrazione dei salari, l'ordine esecutivo da un lato considera l'importanza dei possibili impatti dell'Intelligenza Artificiale Generativa sulle mansioni facilmente sostituibili con modelli di IA, prevedendo, conseguentemente, un'azione di supporto per tutti quei settori che subiranno una

significativa contrazione delle opportunità lavorative; dall'altro, rispetto al tema recruitment, lo stesso ordine prescrive che le valutazioni professionali non utilizzino modelli corrotti da bias.
Al fine di promuovere innovazione e competizione, se da un lato viene annunciato l'allentamento delle restrizioni sull'immigrazione negli USA (anche mediante revisione dei criteri di rilascio dei visti) nella logica di attrarre talenti stranieri, dall'altro si individua nel National AI Research Resource lo strumento che consentirà ai ricercatori e agli studenti di accedere direttamente alle risorse e ai dati, anche attraverso finanziamenti diretti per sviluppare metodologie di ricerca avanzate nei campi della salute e del cambiamento climatico. Sono inoltre previste specifiche iniziative di assistenza tecnica e risorse destinate ai piccoli sviluppatori e imprenditori ed è affidato alla Federal Trade Commission il supporto nella commercializzazione delle ultime scoperte da parte delle piccole imprese.

L'Ordine sinteticamente descritto ha dunque delineato i confini di un'azione di ampio respiro attraverso la quale rafforzare la sicurezza e la protezione dell'IA, proteggere la privacy degli americani, promuovere l'equità e i diritti civili, difendere i consumatori e i lavoratori, promuovere l'innovazione e la concorrenza, far progredire la leadership americana nel mondo. Per raggiungere tali obiettivi sono diverse le autorità coinvolte che sono state chiamate a mettere in campo una serie piuttosto copiosa di attività entro termini piuttosto sfidanti che andavano dai 30 ai 180 gg. dall'adozione dell'order.

Ebbene, lo scorso 29 gennaio, tre mesi dopo la pubblicazione dell'order, è stato reso pubblico il completamento delle attività prescritte dall'order, alcune delle quali anche in anticipo rispetto ai termini fissati come si evince dalla tabella di seguito riportata.

Tab. 3.1: Attività concluse in attuazione dell'executive order

Fonte: Casa Bianca, https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/

Action	Agency	Required Timeline	
Evaluated ways to prioritize agencies' adoption of AI through the Technology Modernization Fund	Technology Modernization Board	30 days	COMPLETE
Directed the Nontraditional and Emerging Transportation Technology Council to evaluate the transportation sector's need for AI guidance and technical assistance	Department of Transportation	30 days	COMPLETE
Reported federal agency resources available to incorporate into the National AI Research Resource (NAIRR) pilot	Agencies identified by the National Science Foundation	45 days	COMPLETE
Identified priority areas for increasing federal agency AI talent and accelerated hiring pathways	Office of Science and Technology Policy & Office of Management and Budget	45 days	COMPLETE
Convened AI and Tech Talent Task Force	White House Chief of Staff's Office	45 days	COMPLETE
Launched an AI Talent Surge to accelerate hiring AI professionals across the federal government, including through a large-scale hiring action for data scientists	Agencies coordinating with the AI and Tech Talent Task Force	45 days	COMPLETE
Published a Request for Information (RFI) on whether to revise the list of Schedule A job classifications that do not require permanent labor certifications	Department of Labor	45 days	COMPLETE
Convened an interagency council to coordinate federal agencies' use of AI	Office of Management and Budget	60 days	COMPLETE
Reviewed the need for -- and granted -- flexible hiring authorities including direct hire and excepted service authorities for federal agencies to hire AI professionals	Office of Personnel Management	60 days	COMPLETE
Used Defense Production Act authorities to compel developers of powerful AI systems to report vital information, especially AI safety test results	Department of Commerce	90 days	COMPLETE
Proposed a draft rule that compels U.S. cloud companies that provide computing power for foreign AI training to report that they are doing so	Department of Commerce	90 days	COMPLETE
Completed risk assessments covering AI's use in every critical infrastructure sector	Sector Risk Management Agencies	90 days	COMPLETE
Launched a pilot of the NAIRR	National Science Foundation	90 days	COMPLETE
Streamlined visa processing, including by renewing and expanding interview-waiver authorities	Department of State	90 days	COMPLETE
Established an AI Task Force to develop policies to provide regulatory clarity and catalyze AI innovation in healthcare	Department of Health and Human Services	90 days	COMPLETE
Convened federal agencies' civil rights offices to discuss the intersection of AI and civil rights	Department of Justice	90 days	COMPLETE
Directed key Federal Advisory Committees to advise on AI and transportation	Department of Transportation	90 days	COMPLETE
Launched a pooled hiring action, to accelerate federal AI hiring, by letting certain applicants apply for roles in multiple agencies with just one application	Office of Personnel Management	90 days	COMPLETE
Released a draft framework for prioritizing generative AI technologies in security authorizations for federally procured products and services	General Services Administration	90 days	COMPLETE
Announced the funding of new Regional Innovation Engines (NSF Engines), including with a focus on advancing AI	National Science Foundation	150 days	COMPLETE
Released an RFI on how federal agencies' privacy impact assessments may be more effective at mitigating privacy risks, including those that are further exacerbated by AI and other advances in technology and data capabilities.	Office of Management and Budget	180 days	COMPLETE
Established an office to coordinate development of AI and other critical and emerging technologies across the agency	Department of Energy	180 days	COMPLETE
Released for comment a draft policy on Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence	Office of Management and Budget	(none)	COMPLETE
Launched the EducateAI initiative, in order to prioritize AI-related workforce development	National Science Foundation	(none)	COMPLETE
Defined AI as a focus area for prize funds through the 2024 Growth Accelerator Fund Competition	Small Business Administration	(none)	COMPLETE
Confirmed the eligibility of AI-related expenditures for support via key programs that benefit small businesses	Small Business Administration	(none)	COMPLETE
Published an RFI on AI's implications for global development	U.S. Agency for International Development & Department of State	(none)	COMPLETE
Proposed changes to a privacy rule that would further limit companies' ability to monetize children's data, including by limiting targeted advertising	Federal Trade Commission	(none)	COMPLETE
Issued an advisory opinion to highlight that false, incomplete, and old information must not appear in background check reports, including for tenant screening	Consumer Financial Protection Bureau	(none)	COMPLETE

Con specifico riferimento alla fissazione di standard ed allo sviluppo di strumenti in grado di mitigare i rischi e consentire di beneficiare a pieno delle immense potenzialità dell'IA, il 7 febbraio scorso il Segretario del Commercio statunitense ha annunciato la creazione dell'AI Safety Institute Consortium (AISIC), composto da oltre 200 stakeholders tra creatori ed utilizzatori di IA, accademici, ricercatori, esponenti governativi e locali, rappresentanti di industrie, startup ed organizzazioni della società civile. Tale consorzio opererà sotto l'AI Safety Institute (USAISI) e contribuirà attivamente al raggiungimento degli obiettivi fissati nell'order.

4. UNIONE EUROPEA: L'ANALISI DELLA PRIMA REGOLAMENTAZIONE ORGANICA AL MONDO SULL'INTELLIGENZA ARTIFICIALE

Accanto a Cina e Stati Uniti, anche l'UE sta affrontando le sfide regolamentari associate ai sistemi di IA, con l'ambizione, così come accaduto attraverso il GDPR per la tutela dei dati personali, di definire un set di regole in grado di assurgere a modello per lo sviluppo e le applicazioni di IA nel contesto globale.

Anche l'UE sta affrontando le sfide regolamentari associate ai sistemi di IA, con l'ambizione, così come accaduto attraverso il GDPR per la tutela dei dati personali, di definire un set di regole in grado di assurgere a modello per lo sviluppo e le applicazioni di IA nel contesto globale

Alle enormi opportunità che l'IA assicura si accompagnano infatti un'importante serie di questioni e temi nuovi da comprendere e governare che hanno spinto la Commissione europea, sin dal 2018, con la comunicazione "AI per l'Europa", ad avviare una serie di iniziative nel campo dell'intelligenza artificiale tra cui la pubblicazione, nel febbraio 2020, del Libro Bianco sull'IA, fino a giungere al lancio, il 21 aprile 2021, di una proposta per un "regolamento del Parlamento europeo e del Consiglio che stabilisce norme armonizzate in materia di intelligenza artificiale e che modifica alcuni atti legislativi dell'Unione" (AI Act), con il quale si istituisce un quadro di riferimento legale volto a normare il mercato dell'UE dell'IA.

Dopo un lungo e complesso *iter* legislativo, nel corso del quale sono state tra l'altro inserite disposizioni tese a disciplinare i sistemi di IA generativa che non erano stati considerati dalla proposta della Commissione ma che negli ultimi due anni sono diventati centrali nel dibattito per le straordinarie applicazioni e le relative potenziali criticità, lo scorso 8 dicembre Parlamento e Consiglio hanno raggiunto un'intesa sul regolamento, di cui attende ora la pubblicazione del testo definitivo. Il testo circolato a fine gennaio, in attesa di essere approvato nei successi passaggi previsti, si compone di 85 articoli e nove allegati che saranno direttamente applicati in tutti gli Stati membri a partire dal 24° mese successivo all'entrata in vigore dello stesso.
Il regolamento, in particolare, detta norme armonizzate per l'immissione sul mercato, la messa in servizio e l'uso di sistemi di intelligenza artificiale nell'UE, compresi quelli di uso generale (*general purpose AI models*), vietando alcuni sistemi di IA ritenuti inaccettabili, fissa requisiti specifici per i sistemi di intelligenza artificiale ad alto rischio e declina obblighi per gli operatori di tali sistemi, prevede regole armonizzate sulla trasparenza per alcuni sistemi di IA, prevede specifiche norme sul monitoraggio del mercato, sulla governance della sorveglianza del mercato e sull'applicazione delle norme ed individua misure a sostegno dell'innovazione, con particolare attenzione alle PMI, comprese le start-up. Per quanto concerne l'ambito applicativo, il regolamento si rivolge ai fornitori che immettono sul mercato o mettono in servizio sistemi di IA nell'UE, indipendentemente dal luogo di stabilimento, ai distributori ed importatori di sistemi di IA, agli utenti dei sistemi di IA situati nell'Unione ed ai fornitori ed utenti di sistemi di IA situati in un paese

terzo, laddove l'output prodotto dal sistema sia utilizzato nell'UE, e persegue la finalità di accrescere la fiducia dei cittadini europei nell'IA. Sono tuttavia escluse dall'ambito di applicazione del regolamento, le tecnologie adoperate per scopi militari e per quelli di ricerca.

Entrando ora nel merito dell'analisi del regolamento, va innanzitutto evidenziato come dal punto di vista metodologico, siano previsti obblighi diversificati che seguono un approccio basato sul rischio, che distingue tra usi dell'IA che creano un rischio inaccettabile, un rischio elevato ed un rischio basso o minimo, da cui discendono evidentemente conseguenze diverse. Nello specifico, particolarmente rilevante, in quanto indicativo del modello che l'UE punta ad esportare nel contesto internazionale, l'elenco delle pratiche vietate in quanto considerate inaccettabili per violazione dei valori dell'Unione (art. 5 del regolamento). Ci si riferisce, nello specifico, ai sistemi che sfruttano tecnologie subliminali per manipolare i comportamenti degli individui e per quelli che abusano di persone vulnerabili e fragili. Parimenti vietata la categorizzazione biometrica che fa riferimento a dati personali sensibili, come il credo religioso, l'orientamento politico o sessuale così come la pratica di scraping da internet di volti, il riconoscimento delle emozioni sul posto di lavoro o a scuola ed i sistemi di punteggio o social scoring. Il regolamento vieta anche la polizia predittiva, ossia l'impiego di informazioni come ad es. tratti della personalità, nazionalità, situazione familiare o economica, per stabilire la probabilità che compia un reato.

Accanto a tali divieti assoluti vengono tuttavia elencate alcune eccezioni tassativamente individuate. Ed infatti, il divieto di categorizzazione biometrica non impedisce l'etichettatura o il filtro di dataset biometrici, legalmente acquisiti, per scopi di polizia, così come è possibile ricorrere a sistemi di analisi del rischio che non facciano profilazione di individui, come quelli per individuare transazioni sospette o per tracciare le rotte del narcotraffico, sulla base delle informazioni salvate nei database. Una delle questioni che più ha impegnato le istituzioni europee è quella relativa all'impiego di sistemi di riconoscimento facciale e biometrico in tempo reale.

Una delle questioni che più ha impegnato le istituzioni europee è quella relativa all'impiego di sistemi di riconoscimento facciale e biometrico in tempo reale

Si tratta, in particolare, di un'applicazione proibita in considerazione del potenziale discriminatorio che essa porta con sé, ma che può essere tuttavia impiegata in tre specifiche situazioni e con le cautele previste: 1) la ricerca di vittime di reati e di persone scomparse; 2) minacce certe alla vita o alla sicurezza fisica delle persone o di attacco terroristico; 3) localizzazione e identificazione dei presunti autori di una lista di 16 reati contenuti in un nell'allegato IIa (terrorismo, traffico di esseri umani, abusi sessuali su minori e pedopornografia, traffico di droghe e sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, omicidio o gravi feriti, traffico di organi, traffico di materiale radioattivo e nucleare, sequestro di persona e ostaggi, crimini sotto la giurisdizione della Corte penale internazionale, dirottamento di aerei e navi, stupri, crimini ambientali, rapine organizzate e armate, sabotaggio e partecipazione a una organizzazione criminale coinvolta in uno o più crimini tra quelli elencati). Quanto alla finalità perseguita col riconoscimento biometrico da remoto in tempo reale, essa consiste nella conferma dell'identità della persona che è stata individuata come target, mentre dal punto di vista procedurale, è prevista una valutazione degli impatti sui diritti fondamentali dei cittadini e un provvedimento di un giudice o di un ente

indipendente, con la possibilità, inoltre, di ricorrere ad una procedura d'urgenza che consente di attivare la sorveglianza biometrica e di richiedere l'autorizzazione entro le 24 ore successive (in caso di mancato rilascio del provvedimento autorizzatorio si prevede l'immediata interruzione e la cancellazione dei dati raccolti). Ai garanti nazionali dei dati personali e del mercato devono spedire ogni anno alla Commissione un rapporto sull'uso dei sistemi di riconoscimento biometrico in tempo reale, così come di eventuali usi proibiti. Ad ogni modo, gli Stati dell'Unione possono adottare leggi nazionali per ampliare il raggio d'azione della sorveglianza biometrica, nel rispetto dei paletti fissati dall'AI Act. Le stesse regole si applicano anche per il riconoscimento facciale usato ex post. In questo caso la finestra per ottenere l'ok in casi di urgenza è di 48 ore.

È riconosciuta agli Stati membri la facoltà di prevedere la possibilità di autorizzare, in tutto o in parte, l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico per finalità di enforcement della legge (prevedendo le procedure da seguire per la richiesta di autorizzazione e l'impiego di tali sistemi nonché per il controllo) così come di adottare una normativa più restrittiva sull'impiego di tali sistemi. Nel primo caso, si prescrive l'invio di un report annuale in relazione all'impiego di tali sistemi che andrà a confluire all'interno del report annuale curato dalla Commissione europea.

Se questi sono i sistemi vietati con le relative eccezioni di impiego, il regolamento detta una disciplina molto dettagliata con riguardo ai sistemi ad alto rischio. Si tratta, in particolare, si sistemi che pongono un significativo rischio per la salute, la sicurezza o i diritti fondamentali dei cittadini e, dunque, i sistemi di identificazione e categorizzazione biometrica o per il riconoscimento delle emozioni, applicativi di sicurezza di infrastrutture critiche, software educativi o di formazione, per valutare i risultati di studio, per assegnare corsi o per controllare gli studenti durante gli esami, algoritmi per valutare curriculum o distribuire compiti e impieghi, algoritmi impiegati dalla pubblica amministrazione o da enti privati per assegnare sussidi, per classificare richieste di emergenza, per smascherare frodi finanziarie o per stabilire il grado di rischio quando si sottoscrive un'assicurazione. A ciò si aggiungono gli algoritmi usati dalle forze dell'ordine, dal potere giudiziario e dalle autorità di frontiera per valutare rischi, scoprire flussi di immigrazione illegale o individuare pericoli sanitari con la precisazione però che se l'algoritmo serve solo per svolgere una procedura limitata, per ottimizzare il risultato di un lavoro realizzato da un individuo, per identificare deviazioni dagli usuali processi decisionali o per svolgere lavori preparatori di controllo, non viene considerato ad alto rischio. Dal punto di vista operativo sarà la Commissione, entro 18 mesi dall'entrata in vigore del regolamento, a fornire le linee guida per l'applicazione della disciplina concernente i sistemi ad alto rischio, così come a modificare, eventualmente, la lista degli algoritmi che ricadono sotto questa categoria.

Rispetto ai sistemi ad alto rischio, il regolamento individua i criteri da seguire per valutare se un sistema di IA presenta alti rischi e fissa una serie di requisiti obbligatori oltre a subordinare l'accesso al mercato europeo di tali sistemi ad una valutazione della conformità ex ante secondo procedure dettagliatamente descritte. Il regolamento a tale riguardo prescrive l'istituzione, la conservazione e la dimostrazione di un sistema di gestione dei rischi che sia frutto di un processo di aggiornamento costante e sistematico nel corso dell'intero ciclo di vita del sistema, l'adozione di adeguate misure di gestione dei rischi da adottare secondo una serie di criteri e principi dettagliatamente enucleati e a seguito di specifiche prove dirette a misurarne l'appropriatezza, la predisposizione e conservazione della documentazione tecnica a supporto, la registrazione automatica dei log (secondo standard minimi fissati) che riguardano il sistema per tutta la durata

di vita dello stesso per risalire a eventuali situazioni di rischio e indagarne le origini, la gestione trasparente dei dati trattati, una progettazione tesa ad assicurare un adeguato livello di accuratezza, robustezza e cibersicurezza, obblighi di monitoraggio successivo all'immissione in commercio e di segnalazione di incidenti gravi e garanzie di collaborazione con le autorità competenti. Lo stesso regolamento prevede inoltre la necessità di prevedere, in caso di pericolo imminente, la possibilità di bloccare l'intelligenza artificiale attraverso un "bottone di stop o una procedura simile, che consente al sistema di bloccarsi in modo sicuro", prescrive agli sviluppatori di istituire un sistema di verifica della qualità, di sottoporsi alle analisi di conformità, di applicare il marchio CE, nonché di comunicare eventuali incidenti alle autorità secondo le tempistiche e le procedure previste. Specifici obblighi sono posti a carico di importatori e distributori di sistemi di IA ad alto rischio.

Per quanto riguarda, invece, i sistemi di AI per uso generale, ossia in grado di svolgere compiti diversi come la produzione di testi o immagini e allenati attraverso un'enorme mole di dati non categorizzati (si pensi a GPT-4 o LaMDA), il regolamento prescrive agli sviluppatori di assicurarsi che i contenuti siano marcati in un sistema leggibile da una macchina e siano chiaramente riconoscibili come generati da un'AI al fine di garantire adeguata consapevolezza da parte degli utenti. Nel tentativo di arginare il dilagare delle fake news, si prevede, inoltre, che i contenuti deepfake siano etichettati come tali (attraverso sistemi come il watermarking, la filigrana digitale applicata a foto o video).

Particolare attenzione è dedicata ai sistemi di alto impatto, ossia quelli che avendo maggiori effetti sulla popolazione, sono soggetti ad obblighi più stringenti su sicurezza informatica, trasparenza dei processi di addestramento e condivisione della documentazione tecnica prima di sbarcare sul mercato. Sul punto il regolamento fissa una soglia - modificabile se opportuno per rispondere alle evoluzioni tecnologiche future - per identificare tali sistemi, identificata in un potere di calcolo pari a 10^25 FLOPs (floating point operations per second, un'unità di misura della capacità computazionale).

Dal punto di vista dei controlli, l'AI Act delega molte attività alle autorità locali che sono chiamate ad istituire almeno una sandbox regolatoria, ossia uno schema che consenta di compiere test in un ambiente sicuro, in deroga alla normativa, nella logica di non rallentare l'innovazione - a livello nazionale, entro 2 anni dall'entrata in vigore del regolamento.

Per quanto riguarda gli aspetti di governance, il regolamento proposto istituisce a livello dell'Unione un board europeo per l'intelligenza artificiale composto dalle autorità nazionali di controllo, rappresentate dal capo di tale autorità o da un alto funzionario di livello equivalente e dal Garante europeo della protezione dei dati e presieduto dalla Commissione, con il compito di raccogliere e condividere conoscenze e migliori pratiche tra gli Stati membri e contribuire all'uniformità delle pratiche amministrative negli Stati membri, formulare pareri, raccomandazioni o contributi scritti su questioni relative all'attuazione del regolamento.

Al fine di fornire supporto tecnico al Comitato e alla Commissione è istituito un forum consultivo (advisory forum) rappresentativo in maniera bilanciata dei vari stakeholders inclusa industria, PMI, start-up, società civile e università e di cui sono membri permanenti una serie di soggetti individuati tra cui l'Agenzia europea per la Cybersecurity (ENISA).

Sempre a supporto dell'attività della Commissione, nell'ambito della quale opera l'AI Office collocato sotto la DG Connect titolare di specifici compiti di monitoraggio e controllo (specie con riguardo ai sistemi per uso generale), è prevista la possibilità di istituire un panel scientifico di esperti indipendenti selezionati dalla stessa Commissione sulla base di una serie di criteri dettati dal regolamento.

A ciascun Stato Membro è rimessa invece la designazione di almeno un'autorità competente al fine di garantire l'applicazione e l'attuazione del regolamento (con il compito, anche, di fornire orientamenti e consulenza sull'attuazione dello stesso regolamento) e di una autorità di notifica e la formulazione di una relazione annuale da trasmettere alla Commissione.

Il regolamento incoraggia, infine, l'adozione di Codici di condotta elaborati da singoli fornitori di sistemi di IA o da organizzazioni che li rappresentano o da entrambi, anche con la partecipazione degli utenti e di tutti gli altri portatori di interessi e delle loro organizzazioni rappresentative tesi a promuovere l'applicazione volontaria ai sistemi di IA dei requisiti relativi, ad esempio, alla sostenibilità ambientale, all'accessibilità per le persone con disabilità.

A presidio dell'osservanza del regolamento, è previsto un aspro set di sanzioni che nel caso di violazione delle norme sugli usi proibiti possono arrivare fino a 35 milioni di euro o al 7% del fatturato globale mentre per le ipotesi di violazioni relative alla disciplina dettata per i sistemi ad alto rischio o quelli di suo generale, possono arrivare fino a un massimo di 15 milioni o del 3% del fatturato globale. Per le ipotesi di invio di informazioni errate, incomplete o fuorvianti alle autorità richiedenti, la sanzione raggiunge un tetto di 7,5 milioni di euro o dell'1% del fatturato globale.

5. ITALIA: VERSO LA NUOVA STRATEGIA NAZIONALE PER L'INTELLIGENZA ARTIFICIALE

Mentre l'UE si trova ad affrontare la sfida della regolamentazione con l'obiettivo di diventare leader nel settore, anche l'Italia cerca di giocare la propria partita per non restare indietro e mettere a sistema, nel rispetto dei diritti fondamentali degli individui, gli enormi benefici che l'IA assicura.

Dopo alcuni tentavi non finalizzati a causa del cambio dei governi e della scarsa attenzione politica verso il tema, nel novembre 2021 è stato approvato dal Governo il Programma Strategico per l'Intelligenza Artificiale (IA) 2022-2024[15], predisposto di concerto dal Ministero dell'Università e della Ricerca, dal Ministero dello Sviluppo Economico e dal Ministro per l'Innovazione tecnologica e la Transizione Digitale. Tale programma ha delineato, seppur con un orizzonte temporale decisamente limitato e senza indicazione specifiche delle risorse destinate a ciascuna iniziativa, le aree di intervento e ventiquattro politiche da implementare nei successivi tre anni al fine di rafforzare il sistema IA in Italia e renderla un esempio di eccellenza, attraverso specifiche iniziative mirate alla creazione ed il potenziamento delle competenze, della ricerca, dei programmi di sviluppo, delle applicazioni ed incentivando il trasferimento tecnologico.
Entrando nel concreto delle previsioni, la strategia ha individuato 6 obiettivi, 11 settori prioritari e 3 aree di intervento.

Gli obiettivi, in particolare, rappresentano le ambizioni italiane in materia di IA e consistono nel rafforzare la ricerca di frontiera nell'IA, ridurre la frammentazione della ricerca sull'IA, sviluppare e adottare un'IA antropocentrica e affidabile nel settore pubblico e privato, aumentare l'innovazione basata sull'IA e lo sviluppo della tecnologia di IA, sviluppare politiche e servizi basati sull'IA nel settore pubblico, creare, trattenere ed attrarre ricercatori di IA in Italia. Per centrare tali obiettivi, il piano individua specifiche aree di intervento, ossia undici settori prioritari: industria e manifatturiero, sistema educativo, agroalimentare, cultura e turismo, salute e benessere, ambiente, infrastrutture e reti, banche, finanza e assicurazioni, PA, città, aree e comunità intelligenti, sicurezza nazionale, tecnologie dell'informazione.

Le aree di intervento individuate, invece, si declinano nel rafforzamento delle competenze e nell'attrazione dei talenti per sviluppare un ecosistema dell'intelligenza artificiale in Italia, nell'incremento dei finanziamenti per la ricerca avanzata nell'IA e nell'incentivo all'adozione dell'IA e delle sue applicazioni, sia nella PA che nei settore produttivi in generale e prevedono l'adozione, in tre anni, di 24 policies. Si tratta di un'ampia gamma di iniziative che mirano a rafforzare le competenze anche attraverso la promozione di corsi e carriere in materie STEM, ad espandere l'IA negli ITS, a promuovere e sostenere finanziariamente ricerche in IA, anche attraverso il lancio della piattaforma italiana di dati e software per la ricerca sull'IA nonché a creare cattedre italiane di ricerca sull'IA. Quanto all'adozione dell'IA da parte di imprese e PA, da un lato, si punta ad accrescere la conoscenza e la consapevolezza attraverso campagne di informazione sull'IA, a promuovere il go-to-market delle tecnologie IA, a sostenere la crescita di spin-off innovativi e start-up e a rendere l'IA un pilastro a supporto della Transizione 4.0 delle imprese; dall'altro, a creare interoperabilità e dati aperti per favorire la creazione di modelli di IA, a rafforzare le

[15] *https://assets.innovazione.gov.it/1637777289-programma-strategico-iaweb.pdf*

soluzioni IA nella PA e nell'ecosistema GovTech in Italia, ad introdurre tecnologie per condivisione e risoluzione casi trasversali a varie autorità e a creare una banca dati IA/Computer Vision per il miglioramento dei servizi nella PA.

Partendo dai limiti di tale piano, che possiamo sintetizzare in un orizzonte temporale e spaziale troppo ridotti (tre anni e un focus preponderante sulla ricerca e sviluppo) e nella mancanza di un budget e di una governance dedicati, il Governo sta rimettendo mano alla strategia, nella quale in base agli annunci già fatti pubblicamente dovrebbero trovare posto, tra l'altro, due iniziative particolarmente importanti: da un lato, la costituzione di un fondo pubblico-privato di venture capital, sotto l'egida di CDP, per favorire la crescita delle startup innovative italiane; dall'altro, l'istituzione di una Fondazione, che dovrebbe chiamarsi AI4Industry, dotata di un budget finanziato dallo stato di venti milioni di euro l'anno.

Se il Governo sta rimettendo mano alla strategia, anche il Parlamento è impegnato nella sfida della regolamentazione. Ed infatti, lo scorso 19 ottobre è stato presentato dal PD un disegno di legge[16] che ha ad oggetto i contenuti editoriali, compresi testi, video, immagini e voci, che sono creati, generati o sintetizzati, in tutto o in parte, da sistemi basati su intelligenza artificiale (ivi compresi algoritmi di apprendimento automatizzato, cosiddetto machine learning e reti neurali artificiali) i quali devono essere "*chiaramente identificati come tali e resi riconoscibili agli utenti attraverso sistemi di etichettatura, cosiddetta label, e filigrana, cosiddetta watermark".*

Lo scorso 19 ottobre è stato presentato dal PD un disegno di legge[17] che ha ad oggetto i contenuti editoriali, compresi testi, video, immagini e voci, che sono creati, generati o sintetizzati, in tutto o in parte, da sistemi basati su intelligenza artificiale

I soggetti responsabili della pubblicazione e della diffusione dei contenuti generati da IA, in ogni mezzo trasmissivo, sono chiamati a fornire un'etichettatura e un avviso visibile, all'inizio e alla fine del contenuto, facilmente comprensibili agli utenti, che indichino che il contenuto è stato creato, in tutto o in parte, da un sistema di IA, secondo le modalità che saranno individuate dall'AGCom con proprio regolamento (entro 60 gg. dall'entrata in vigore della legge). Alla stessa Autorità è affidato il compito di monitorare il rispetto di tali prescrizioni e di definire gli strumenti di segnalazione e rimozione dei contenuti pubblicati e diffusi in violazione della disciplina descritta 2, nonché il regime sanzionatorio da applicare.

Accanto alla consapevolezza degli utenti circa l'origine dei contenuti prodotti da IA, l'altra irrinunciabile esigenza è quella di assicurare che l'IA sia sviluppata in maniera sicura.

A tale riguardo si segnala, per importanza, l'adesione dell'Agenzia per la Cybersicurezza Nazionale (ACN), nel novembre scorso, alle "*Linee guida per uno sviluppo sicuro dell'Intelligenza Artificiale",* promosse dal National Cyber Security Centre del Regno Unito cui hanno aderito 18 paesi e 23 agenzie. Tali linee guida, in particolare, sono state elaborate nel solco della prima conferenza

[16] *https://www.senato.it/service/PDF/PDFServer/DF/428771.pdf*

[17] *https://www.senato.it/service/PDF/PDFServer/DF/428771.pdf*

internazionale sull'Intelligenza Artificiale, il cosiddetto "AI Safety Summit" al quale si è fatto riferimento nel primo capitolo, e perseguono il fine di supportare gli sviluppatori di qualsiasi sistema basato sull'IA e di innalzare i livelli di cybersecurity dell'IA per assicurare che sia progettata, sviluppata e impiegata in maniera sicura.

Si tratta di un documento dall'elevato tecnicismo suddiviso in quattro aree chiave all'interno del ciclo di vita di sviluppo del sistema di IA: progettazione sicura, sviluppo sicuro, implementazione sicura e gestione sicura e manutenzione.

Si tratta di un documento dall'elevato tecnicismo suddiviso in quattro aree chiave all'interno del ciclo di vita di sviluppo del sistema di IA: progettazione sicura, sviluppo sicuro, implementazione sicura e gestione sicura e manutenzione

Nello specifico, il documento individua i rischi e le minacce ed offre consigli di progettazione del sistema e del modello affrontando anche il tema, cruciale, della sicurezza della catena di approvvigionamento, nonché la protezione dell'infrastruttura e dei modelli da compromissioni, minacce o perdite, lo sviluppo di processi di gestione degli incidenti e il rilascio responsabile, oltre a declinare indicazioni specifiche su monitoraggio e condivisione delle informazioni.

SECONDA PARTE: L'ADOZIONE E LA DIFFUSIONE DELL'IA NEL SETTORE PUBBLICO E PRIVATO

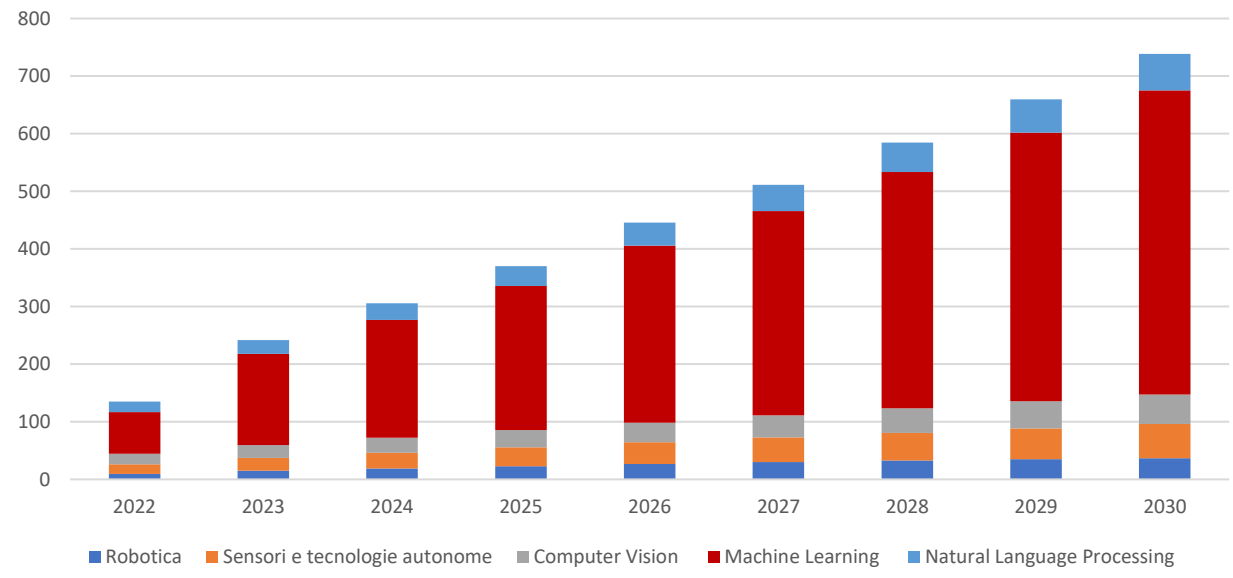
6. IL MERCATO DELL'INTELLIGENZA ARTIFICIALE: STATUS QUO E PROSPETTIVE FUTURE

L'intelligenza artificiale è sicuramente considerata una delle più strabilianti frontiere tecnologiche dei nostri tempi, per le grandi opportunità che offre sia al settore privato sia a quello pubblico e il crescente interesse nei confronti delle numerose applicazioni IA è confermato anche dalle dinamiche positive del mercato, destinate a durare nei prossimi anni.
Recenti stime prevedono, infatti, che le dimensioni del mercato mondiale dell'intelligenza artificiale[18] raggiungeranno in termini di spesa i $305,9 miliardi nel 2024. Il mercato IA continuerà poi la sua ascesa per toccare i $738,8 miliardi nel 2030, mostrando un tasso di crescita medio annuo (CAGR) nel periodo 2024-2030 del 15,83% (Fig.1). Come si evince dalla figura riportata di seguito, a trainare principalmente il mercato delle tecnologie intelligenti saranno le applicazioni basate sulle tecniche di machine learning.

Le dimensioni del mercato mondiale dell'intelligenza artificiale raggiungeranno in termini di spesa i $305,9 miliardi nel 2024

Fig.1: Il mercato mondiale dell'intelligenza artificiale ($ miliardi)

Fonte: Statista
Note: stime

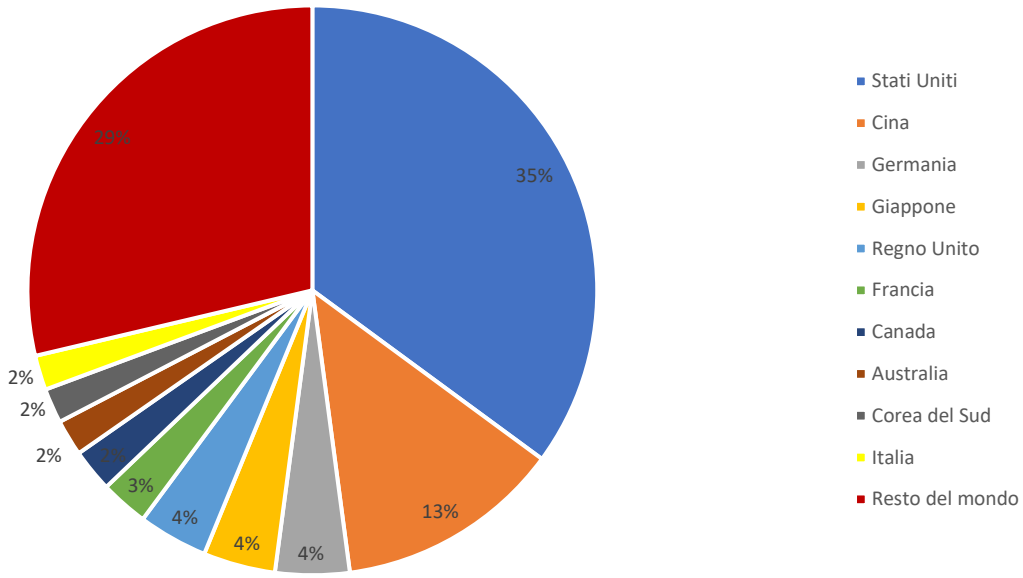


[18] *Le dimensioni del mercato sono generate dall'importo dei finanziamenti delle società di intelligenza artificiale (IA) in progetti ed iniziative IA.*

Nel confronto globale, gli Stati Uniti assumono il ruolo di leadership, coprendo il 35% del mercato IA, seguiti da Cina (13%), Germania (4%), Giappone (4%) e Regno Unito (4%). L'Italia si colloca al decimo posto dopo la Corea del Sud (Fig.2).

Fig.2: Il mercato mondiale dell'intelligenza artificiale, per Paese (in % del valore totale, 2024)

Fonte: Statista
Note: stime

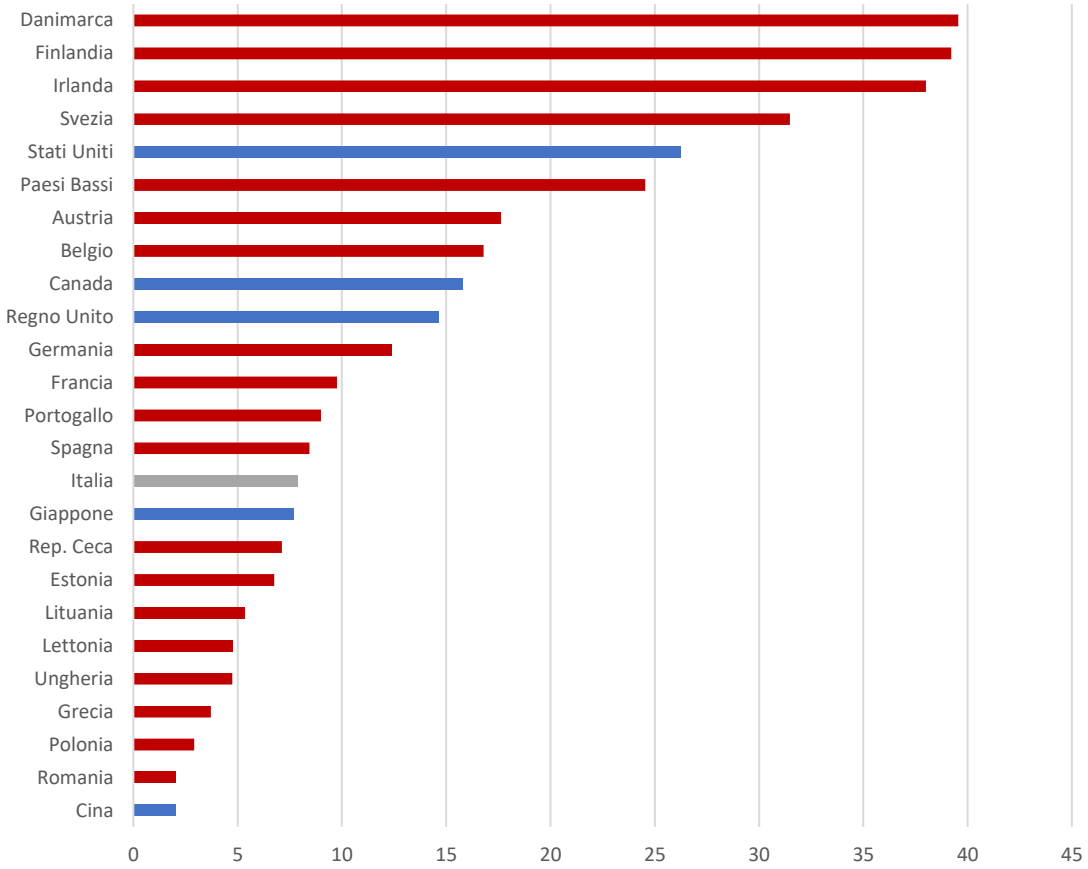


Tenendo conto, però, delle dimensioni della popolazione di ciascun paese, la Danimarca risulta il più grande mercato dell'IA a livello globale, con un valore di mercato per 100.000 abitanti di circa $39 milioni, seguita da Finlandia e Irlanda. Gli Stati Uniti si collocano, invece, in quinta posizione anche se in termini assoluti mostrano il valore di mercato più alto ($87,18 miliardi nel 2023). La Germania è il più grande mercato dell'IA in Europa in termini assoluti ($10,30 miliardi nel 2023). Tuttavia, in termini relativi si colloca in undicesima posizione nella classifica globale (e ottava tra i 20 Stati membri UE considerati, con $12 milioni per 100.000 abitanti); l'Italia quindicesima (e dodicesima UE) con circa $8 milioni per 100.000 abitanti, un valore pari a 1/5 di quello danese e meno di 1/3 di quello statunitense (Fig.3).

Tenendo conto delle dimensioni della popolazione di ciascun paese, la Danimarca risulta il più grande mercato dell'IA a livello globale, con un valore di mercato per 100.000 abitanti di circa $39 milioni, seguita da Finlandia e Irlanda

Fig.3: Valore di mercato IA/100.000 abitanti (in milioni di $; 2023) – Confronto tra gli Stati Membri e Stati Uniti, Canada, Regno Unito, Cina e Giappone

Fonte: elaborazioni I-Com su dati Statista, Eurostat e OCSE

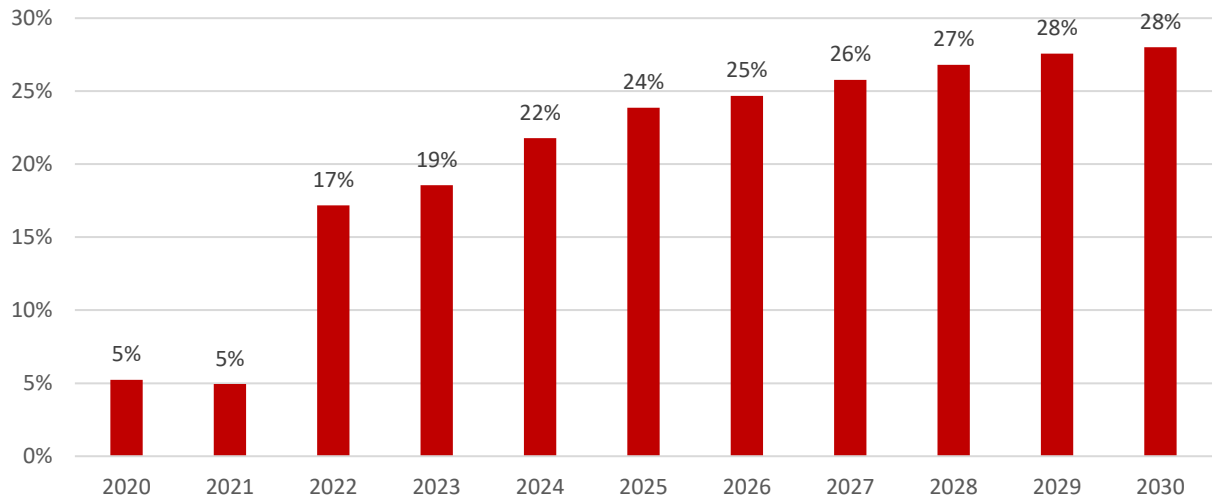


Negli ultimi anni, con un'evidente accelerazione nell'ultimo, il mercato dell'IA è sempre più trainato dall'IA generativa, che utilizza tecniche di machine learning e deep learning per generare nuovi dati, tra cui immagini, musica e testo, che non esistevano in precedenza.
L'IA generativa rappresenta già oggi una fetta rilevante del mercato IA, destinata ad aumentare nei prossimi anni. Nel 2023, stando ad alcune stime, ha coperto il 19% del mercato IA totale ed entro il 2024 si prevede un aumento dell'incidenza di 3 punti percentuali. Nel 2030, il mercato dell'IA generativa dovrebbe pesare sull'intero mercato IA per il 28% (Fig.4).

L'IA generativa rappresenta già oggi una fetta rilevante del mercato IA, destinata ad aumentare nei prossimi anni. Nel 2023, stando ad alcune stime, ha coperto il 19% del mercato IA totale ed entro il 2024 si prevede un aumento dell'incidenza di 3 punti percentuali

Fig.4: Il mercato mondiale dell'IA generativa (in % del totale del mercato IA)

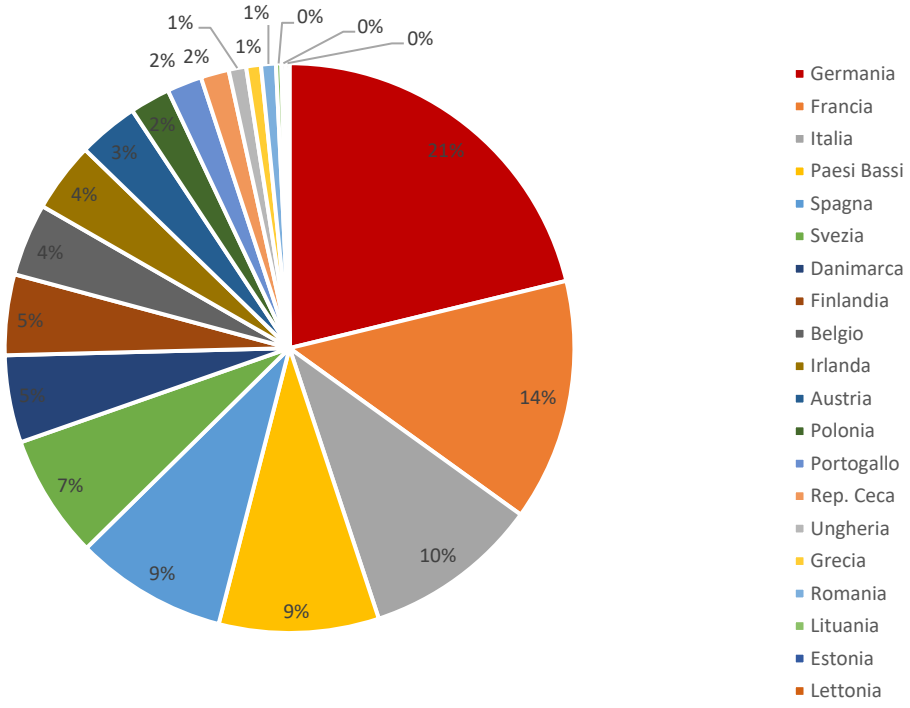
Fonte: elaborazioni I-Com su stime Statista



Il segmento dell'IA generativa è pronto, dunque, ad esplodere e tra i principali Stati Membri, la Germania si candida ad essere il più grande mercato dell'IA generativa coprendo il 21% del mercato europeo, seguita da Francia (14%) e Italia (10%) (Fig.5).

Fig.5: Il mercato dell'IA generativa negli Stati Membri (in %; 2024)

Fonte: elaborazioni I-Com su stime Statista



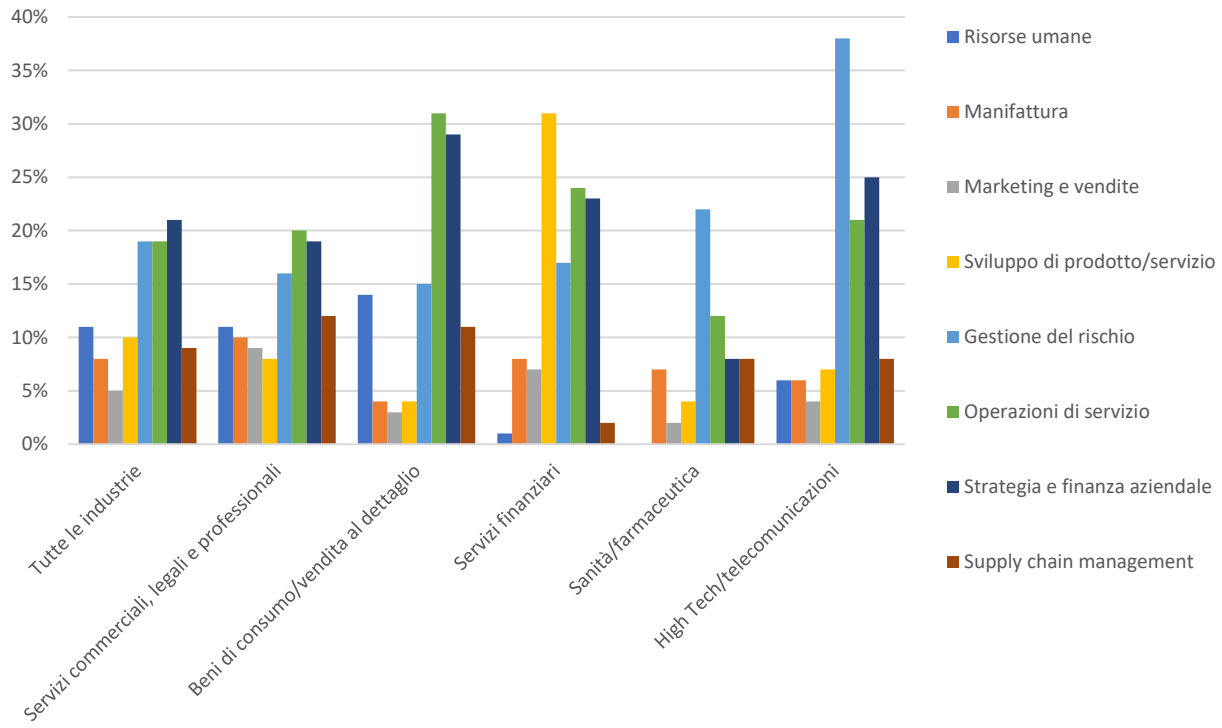
7. LE APPLICAZIONI IA A SUPPORTO DELLE IMPRESE E DELLA PA

7.1. Il livello di adozione dell'IA nei vari settori economici e la diffusione delle tecnologie intelligenti nelle imprese italiane ed europee

Gli ambiti applicativi dell'intelligenza artificiale sono davvero innumerevoli e spaziano dal settore ICT al settore manifatturiero, dal campo della sanità, al campo del fintech e dell'insurtech, fino ad interessare il settore dei servizi legali e professionali, con impatti importanti sulle attività di imprese e pubbliche amministrazioni, oltre che sulla vita delle persone.
Secondo i dati riportati dalla Stanford University[19], l'intelligenza artificiale (AI) è ampiamente utilizzata per le operazioni di servizio, la strategia e la finanza aziendale, e le attività di risk management e quasi tutti i settori economici presi in considerazione riportano circa il 20% di utilizzo dell'IA in queste tre funzioni. Il maggiore utilizzo dell'IA nello sviluppo di prodotti/servizi si riscontra, invece, nel settore dei servizi finanziari, in cui raggiunge il 30%.
Infine, nel settore High Tech/Telecomunicazioni, l'IA è ampiamente utilizzata per le attività di risk management, dove la percentuale di adozione si attesta sul 38% (Fig.6).

Fig.6: L'adozione dell'IA nei settori industriali per funzione aziendale (2022)

Fonte: Stanford University (2023)

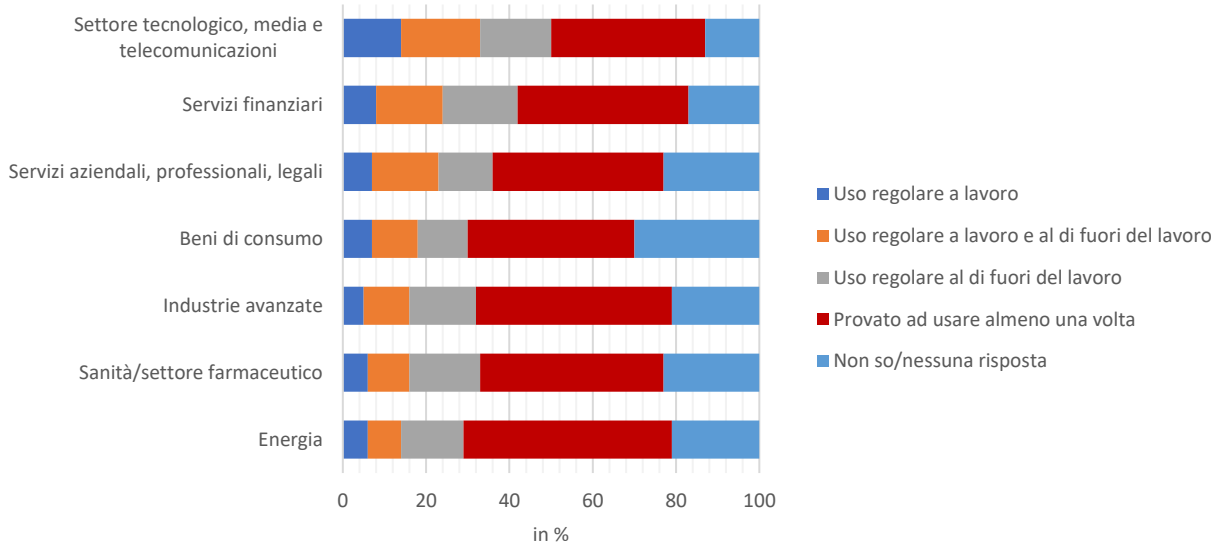


[19] https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf

Focalizzando l'attenzione sull'IA generativa, si evidenzia una crescita esponenziale dell'uso di questa nuova tecnologia in tutti i settori. Un recentissimo sondaggio di McKinsey a livello globale[20] evidenzia che il 33% degli intervistati del settore tecnologico, dei media e delle telecomunicazioni utilizza regolarmente l'IA generativa per lavoro o al di fuori del lavoro, mentre il 37% degli intervistati dello stesso settore ha dichiarato di aver utilizzato questa tecnologia almeno una volta. Altri due settori che usano questi nuovi strumenti sono i servizi finanziari e i servizi aziendali, legali e professionali in cui quasi un quarto degli intervistati utilizza regolarmente l'IA generativa (Fig.7). Questi risultati, per quanto riguarda i servizi finanziari e i settori tecnologici, sono piuttosto coerenti con quelli della survey realizzata per l'Italia da The European House – Ambrosetti in un recente studio[21] realizzato in collaborazione con Microsoft e pubblicato a settembre 2023 sui possibili impatti dell'AI generativa per l'economia italiana. L'impiego nei segmenti manifattura e salute & scienze della vita appaiono invece più intensi in Italia relativamente ai restanti settori. Fanalino di coda nella survey realizzata sull'Italia il settore pubblico, con livelli di diffusione "bassi" (quinto e ultimo cluster individuato dalla ricerca), ma in generale i servizi tradizionali accusano un ritardo significativo rispetto alla manifattura e ai servizi avanzati, con ogni probabilità un riflesso della minore digitalizzazione.

Fig.7: L'uso delle applicazioni di IA generativa nei principali settori economici

Fonte: McKinsey & Company (2023)



Anche se a livello globale si risconta un incremento dell'adozione dell'IA da parte delle imprese[22], a livello europeo si può osservare che, nel 2023, solo l'8% delle imprese UE ha adottato almeno una tra le tecnologie IA più comuni.

[20] *https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-AIs-breakout-year*

[21] The European House – Ambrosetti e Microsoft, *AI 4 Italy: impatti e prospettive dell'intelligenza artificiale generativa per l'Italia e il Made in Italy*, 2023.

[22] *https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/*

Il tasso di adozione più elevato si registra in Danimarca, dove il 15% circa delle imprese fa uso di almeno una tecnologia IA. In fondo alla classifica si colloca la Romania, che presenta il livello più basso di adozione IA con solo l'1,5% di imprese utilizza almeno una tecnologia. L'Italia si colloca al di sotto della media UE, con il 5% delle imprese che ha adottato almeno una tra le tecnologie IA a disposizione (Fig.8).
Tra le diverse tecnologie IA, l'Italia presenta la performance migliore nella robotica di servizio, unico campo dove si colloca nettamente al di sopra della media UE, con oltre il 4% delle imprese che ha adottato robot di servizio (Fig.9). Relativamente alle altre tecnologie, tra cui text mining, speech recognition, machine learning, l'adozione da parte delle imprese italiane è invece sempre al di sotto della media europea.

Anche se a livello globale si risconta un incremento dell'adozione dell'IA da parte delle imprese, a livello europeo si può osservare che, nel 2023, solo l'8% delle imprese UE ha adottato almeno una tra le tecnologie IA più comuni

Fig.8: Imprese UE che hanno adottato almeno una tecnologia IA (in % delle imprese totali, 2023)

Fonte: Eurostat

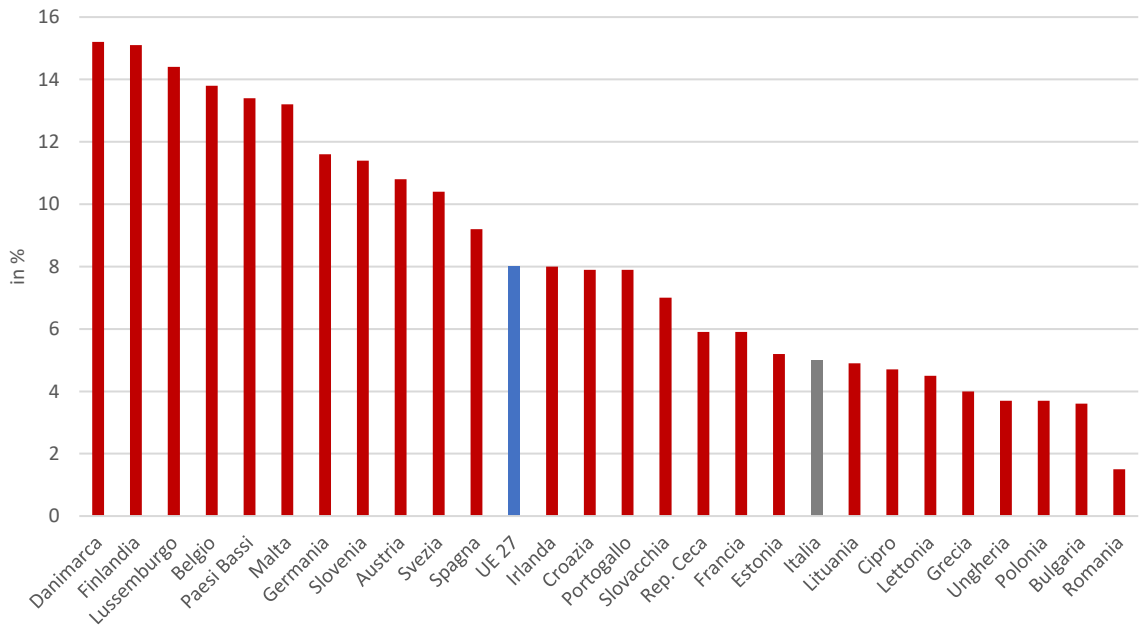
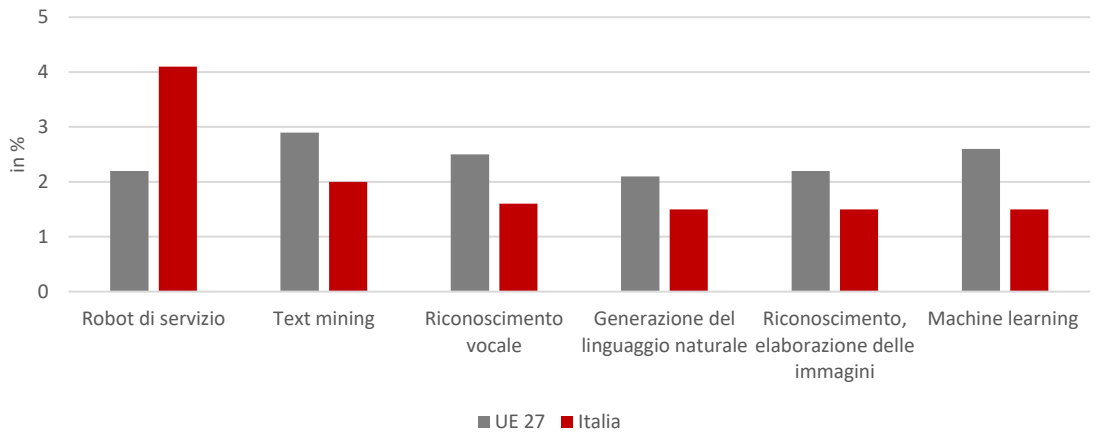


Fig.9: L'adozione delle tecnologie IA più comuni: confronto Italia – UE (in % delle imprese totali, 2023)

Fonte: Eurostat
Nota: Il dato sull'adozione di robot di servizio fa riferimento al 2022



Tra i settori economici che in Italia utilizzano maggiormente le tecnologie IA spicca non a caso quello informatico, dove il 23,6% delle imprese utilizza software o sistemi di IA per almeno una delle 7 finalità più comuni, ossia:

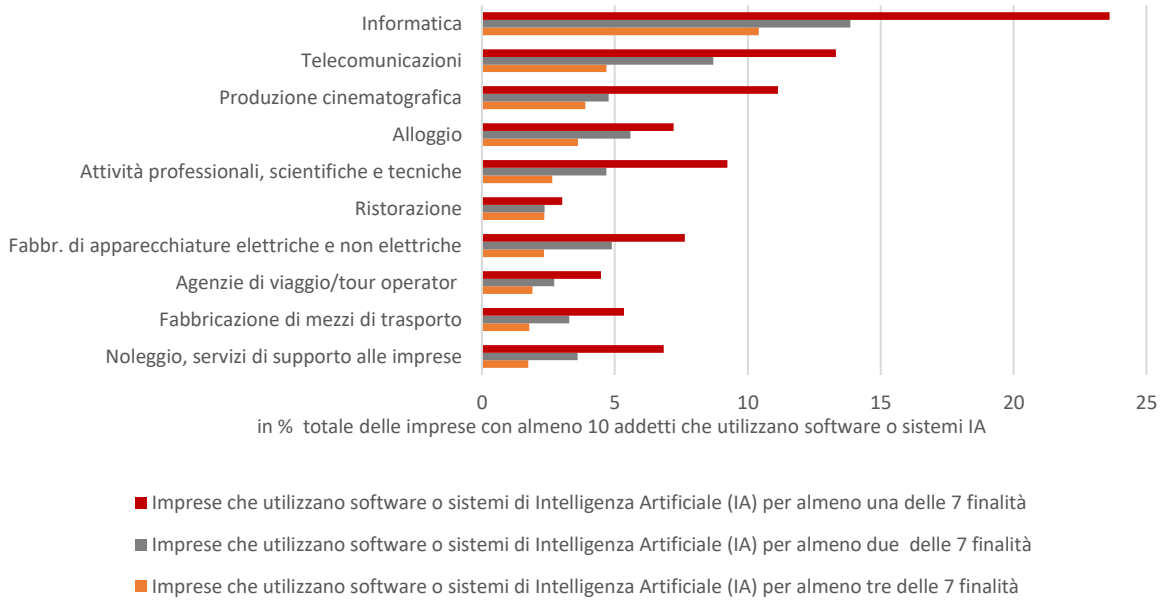
1. Estrarre conoscenza e informazione da un documento di testo (text mining);
2. Convertire la lingua parlata in un formato leggibile dal dispositivo informatico (riconoscimento vocale);
3. Generare linguaggio scritto o parlato (generazione del linguaggio naturale);
4. Identificare oggetti o persone sulla base di immagini (riconoscimento, elaborazione delle immagini);
5. Analizzare dati attraverso l'apprendimento automatico (machine learning, deep learning, reti neurali);
6. Automatizzare i flussi di lavoro o supportare nel processo decisionale (Robotic Process Automation, software robot che utilizzano tecnologie di IA per automatizzare le attività umane);
7. Consentire il movimento fisico delle macchine tramite decisioni autonome basate sull'osservazione dell'ambiente circostante (robot o droni autonomi, veicoli a guida);

Seguono poi il settore delle telecomunicazioni e della produzione cinematografica dove, rispettivamente, il 13,3% e l'11% circa delle imprese ha adottato l'IA.
Relativamente all'intensità di utilizzo, spicca nuovamente il settore dell'informatica dove più del 10% delle imprese utilizza software o sistemi IA per almeno tre finalità tra quelle sopra citate (Fig.10).

Tra i settori economici che in Italia utilizzano maggiormente le tecnologie IA spicca non a caso quello informatico, dove il 23,6% delle imprese utilizza software o sistemi di IA per almeno una delle 7 finalità più comuni

Fig.10: I primi dieci settori economici che utilizzano software o sistemi di IA (% di imprese; 2023)

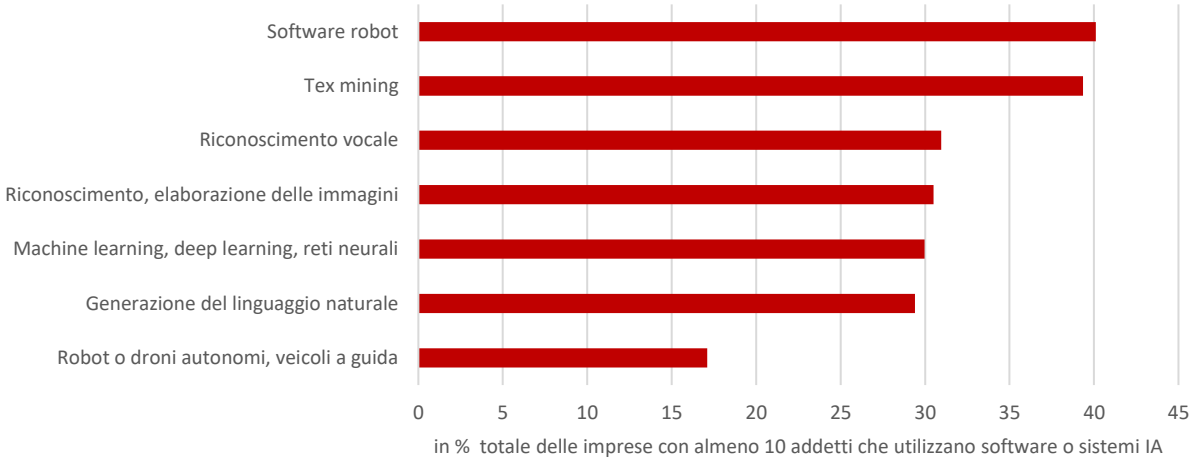
Fonte: Istat (2023)



Tra le imprese italiane che utilizzano l'IA, le tecnologie più comuni riguardano l'automatizzazione di flussi di lavoro attraverso software robot (40,1%), le applicazioni di text mining (39,3%) e il riconoscimento vocale (31%) (Fig.11).

Fig.11: Le tecnologie IA più usate dalle imprese italiane (% di imprese; 2023)

Fonte: Istat (2023)



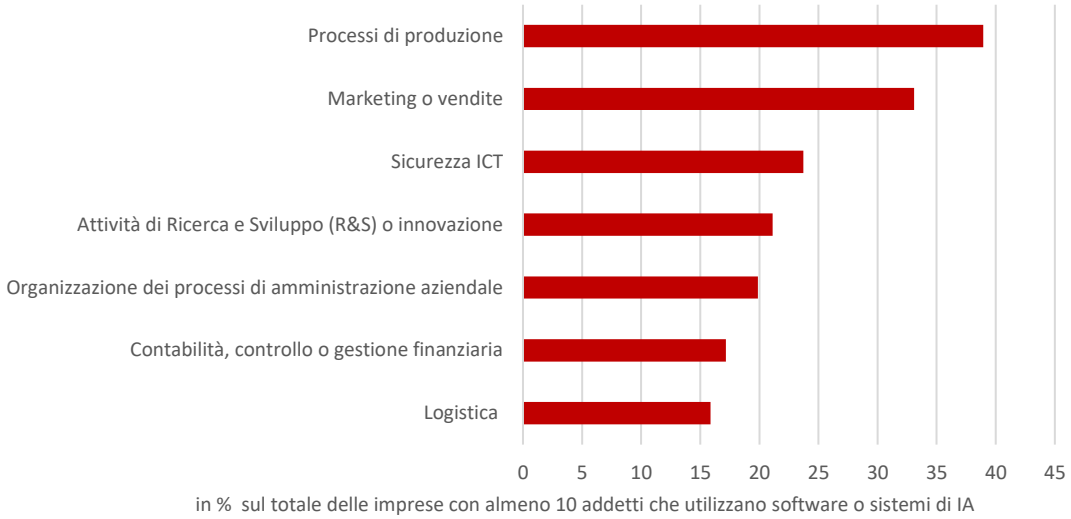
Gli ambiti aziendali in cui vengono più spesso adottati sistemi di intelligenza artificiale sono quelli relativi a processi di produzione, ad esempio per la manutenzione predittiva o il controllo qualità della produzione, in cui il 39% delle imprese italiane fa uso di questi strumenti (nel caso del settore manifatturiero si raggiunge il 52,5% delle imprese), alla funzione di marketing o vendite, ad

esempio per funzioni di assistenza ai clienti o campagne promozionali personalizzate (33,1% per il totale delle attività economiche mentre si raggiunge il 41,3% nel settore dei servizi), alla sicurezza informatica (23,7%, al 50,6% nel settore dell'energia) e alle attività di ricerca e sviluppo (R&S) o innovazione per analizzare dati, sviluppare un prodotto/servizio nuovo o significativamente migliorato (21,1%) (Fig.12)[23].

Tra le imprese italiane che utilizzano l'IA, le tecnologie più comuni riguardano l'automatizzazione di flussi di lavoro attraverso software robot (40,1%), le applicazioni di text mining (39,3%) e il riconoscimento vocale (31%)

Fig.12: Gli ambiti aziendali in cui vengono adottati i sistemi di IA (% di imprese; 2023)

Fonte: Istat (2023)



Se ci si concentra sull'IA generativa, la survey condotta nell'ambito del recente studio[24] realizzato da The European House – Ambrosetti in collaborazione con Microsoft ha rilevato come il 50,9% delle imprese del campione li stesse già utilizzando nel 2023 e a quella data solo il 21,8% non ne avesse una previsione di impiego. Questo dato così elevato si spiega in gran parte con la percentuale di grandi imprese (con oltre 250 dipendenti) presenti nel campione (il 54%) ma è certamente un segnale incoraggiante del potenziale di innovazione e dell'interesse verso la tecnologia del tessuto produttivo italiano.

[23] https://www.istat.it/it/files//2023/12/report-imprese_2023.pdf

[24] The European House – Ambrosetti e Microsoft, *AI 4 Italy: impatti e prospettive dell'intelligenza artificiale generativa per l'Italia e il Made in Italy*, 2023.

7.2. Gli impatti dell'adozione delle tecnologie IA sulla pubblica amministrazione

L'intelligenza artificiale rappresenta senz'altro un tema di grandi potenzialità anche per il settore pubblico, caratterizzato da complesse sfide e responsabilità, in cui l'adozione di soluzioni basate sull'IA potrebbe portare a vantaggi significativi. Con l'evolversi rapido di questa tecnologia, emerge dunque la necessità di una valutazione approfondita circa gli effetti dell'IA in ambito pubblico. Tuttavia, la letteratura dedicata a esplorare l'impatto dell'IA in questo contesto è ancora limitata ed è costituita principalmente da analisi di carattere teorico. L'assenza di studi empirici in grado di quantificare l'entità dell'impatto dell'IA nel settore pubblico è probabilmente dovuta alla carenza di dati sull'adozione di questa tecnologia, soprattutto nei contesti pubblici, il che rende molto difficile misurare il fenomeno e suoi possibili impatti. Questo genera, come sottolineato in alcuni recenti lavori, una situazione quasi paradossale per cui le amministrazioni pubbliche utilizzano già l'intelligenza artificiale, ma senza alcun supporto empirico da parte della ricerca scientifica (Sun e Medaglia, 2019; Tangi et al., 2022). In questo paragrafo si tenterà di riassumere brevemente la letteratura recente sull'impatto delle soluzioni di IA nel settore pubblico per poi analizzare i dati raccolti nell'ambito del servizio di studi "AI Watch" della Commissione Europea. Verranno infine considerati alcuni casi di particolare interesse nel contesto nazionale.

Anche nel settore pubblico, caratterizzato da complesse sfide e responsabilità, l'adozione di soluzioni basate sull'IA potrebbe portare a vantaggi significativi

La crescente rilevanza dell'intelligenza artificiale e il suo enorme potenziale innovativo per il settore pubblico sono oggi ampiamente riconosciuti a livello globale. Numerose amministrazioni, oltre a sondare le opportunità tramite progetti pilota, hanno avviato l'implementazione di questa tecnologia in svariati contesti e utilizzano l'IA nello svolgimento di diverse operazioni quotidiane (Sousa et al., 2019; Molinari et al., 2021). In Europa, i casi di impiego dell'AI nel settore pubblico sono in costante aumento dal 2015, ma sono soprattutto Cina e Stati Uniti ad aver mostrato maggiori progressi nell'avanzamento dell'uso dell'IA nel contesto della pubblica amministrazione (Allen, 2019).

Wirtz et al. (2019), attingendo da precedenti studi e informazioni su programmi governativi, sono stati tra i primi a identificare e categorizzare i diversi casi di impiego dell'AI nel settore pubblico. La loro ricerca ha fornito una panoramica completa delle applicazioni di intelligenza artificiale in tale settore e ha portato all'identificazione di dieci distinte categorie di applicazioni:

1. Software di gestione della conoscenza
2. Sistemi di automazione dei processi
3. Agenti virtuali
4. Analisi predittiva e data visualization
5. Analisi dell'identità
6. Robotica cognitiva e sistemi autonomi

7. Sistemi di raccomandazione
8. Assistenti digitali intelligenti (IDA)
9. Analisi del linguaggio
10. Analisi cognitiva della sicurezza e intelligence sulle minacce

Gli autori sostengono che alcune di queste applicazioni, come ad esempio i sistemi di automazione, possono contribuire in modo significativo ad aumentare l'efficienza dei processi, migliorare l'allocazione delle risorse e ridurre costi e oneri amministrativi. Inoltre, poiché nel settore privato sono già ben noti e applicati tutti i casi d'uso dell'intelligenza artificiale menzionati, Wirtz et al. (2019) suggeriscono che sarebbe auspicabile il trasferimento del know-how e dell'esperienza dal settore privato a quello pubblico.

Gli autori sostengono che alcune di queste applicazioni, come ad esempio i sistemi di automazione, possono contribuire in modo significativo ad aumentare l'efficienza dei processi, migliorare l'allocazione delle risorse e ridurre costi e oneri amministrativi

Esistono anche diversi studi che sottolineano come le soluzioni basate sull'IA nella gestione e nell'erogazione dei servizi pubblici, laddove implementate, stanno già offrendo notevoli vantaggi e valore pubblico ai cittadini.

Tangi et al. (2022), partendo da un database di 686 casi di impiego dell'IA nel settore pubblico dei paesi europei, hanno classificato tali casi in base al loro contributo al valore pubblico. La ricerca ha evidenziato che il maggiore contributo si registra nell'ambito dei servizi pubblici, con il 53% dei casi che contribuisce al miglioramento della gestione o dell'erogazione dei servizi pubblici. Nel 47% dei casi, l'IA ha un impatto positivo sull'efficienza amministrativa, mentre nel 12% dei casi si registrano impatti positivi sulla trasparenza e comunicazione delle politiche governative.

Tangi et al. (2022), partendo da un database di 686 casi di impiego dell'IA nel settore pubblico dei paesi europei, hanno classificato tali casi in base al loro contributo al valore pubblico. La ricerca ha evidenziato che il maggiore contributo si registra nell'ambito dei servizi pubblici, con il 53% dei casi che contribuisce al miglioramento della gestione o dell'erogazione dei servizi pubblici

Alla luce anche della recente pandemia da Covid-19, diversi studi hanno approfondito le potenzialità dell'IA nell'ambito del settore sanitario. Jungwirth e Haluza (2023), ad esempio, mostrano come l'IA possa contribuire in modo positivo alle attività di ricerca nel campo della salute pubblica. Altri accademici, invece, si sono concentrati sul ruolo dell'intelligenza artificiale nella creazione e nell'evoluzione delle Smart Cities, finalizzate a ottimizzare le risorse pubbliche e

affrontare sfide urbane che vanno dalla sicurezza pubblica alla gestione dei sistemi di trasporto (Allam e Dhunny, 2019; Ullah et al, 2020).
Dunque, per quanto possa essere difficile quantificare l'impatto dell'IA sul settore pubblico nel suo complesso, gli studi condotti finora dimostrano che l'IA può offrire, e in alcuni casi sta già offrendo, valore tangibile ai cittadini.
Nel 2018, la Commissione Europea ha lanciato l'*AI Watch*, per monitorare lo sviluppo, l'adozione e l'impatto dell'intelligenza artificiale in Europa. Al fine di rispondere a domande fondamentali sull'adozione dell'IA nel settore pubblico e fornire a ricercatori e decisori di policy una solida base per la ricerca è stato pubblicato un database sui casi di impiego dell'IA in Europa. Il database è costituito da un inventario di 686 casi di utilizzo di intelligenza artificiale nel settore pubblico, che coprono tutti i 27 Stati membri dell'UE, oltre ad alcuni altri Paesi europei.

Nel 2018, la Commissione Europea ha lanciato l'AI Watch, per monitorare lo sviluppo, l'adozione e l'impatto dell'intelligenza artificiale in Europa

La raccolta dei casi ha avuto inizio nel dicembre 2019 ed è stata effettuata utilizzando diverse fonti, da articoli e notizie raccolti attraverso la rete ad un'indagine *ad hoc* sulle sfide e le opportunità dell'IA. La raccolta è proseguita fino a dicembre 2021 e ha portato alla creazione di un database che rappresenta il primo tentativo a livello europeo di offrire una panoramica della situazione attuale rispetto allo sviluppo e all'uso dell'IA nel settore pubblico dei paesi dell'Unione. Nelle pagine che seguono, verranno analizzati in maniera aggregata i casi di impiego contenuti nel database, concentrandosi in particolare sulla situazione italiana. Prima di esaminare i risultati dell'analisi, è importante sottolineare che – come evidenziato in Tangi et al. (2022) – le informazioni contenute nel database non sono e non mirano a essere rappresentative della situazione in Europa e pertanto non sono rilevanti per trarre conclusioni o confrontare il livello di maturità dei paesi europei circa l'adozione dell'IA. Ciò nonostante, l'analisi può offrire un'interessante panoramica e contribuire a una più ampia comprensione delle dinamiche nell'adozione dell'IA nel settore pubblico.

La figura 13 mostra il numero dei casi di impiego dell'intelligenza artificiale nel settore pubblico per anno di inizio progetto, nel periodo dal 2010 al 2021 in Europa. La tendenza che emerge dai dati è chiaramente positiva, e mostra un aumento costante nel numero di casi nel corso degli anni, con un picco di 167 casi nel 2021.

Considerando la distribuzione dei casi tra i paesi europei, emerge un quadro molto eterogeneo. Il numero più elevato di casi di impiego di intelligenza artificiale si registra nei Paesi Bassi (124 casi registrati) – molto probabilmente a causa dell'esistenza di un registro nazionale di casi – seguito dall'Italia (75 casi registrati) e Portogallo (60 casi registrati)(Fig.14). A tal proposito, tuttavia, si richiama quanto appena detto circa l'impossibilità di confrontare il livello di adozione dell'intelligenza artificiale dei singoli paesi europei utilizzando i dati a disposizione.

Fig.13: Casi di impiego dell'IA nel settore pubblico per anno di inizio progetto (2010-2021)

Fonte: Joint Research Center della Comissione Europea

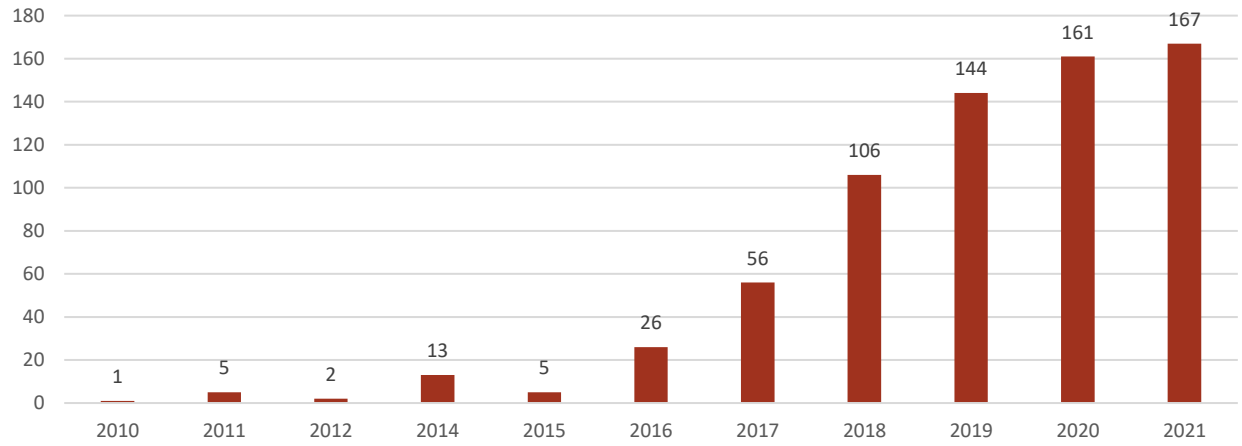
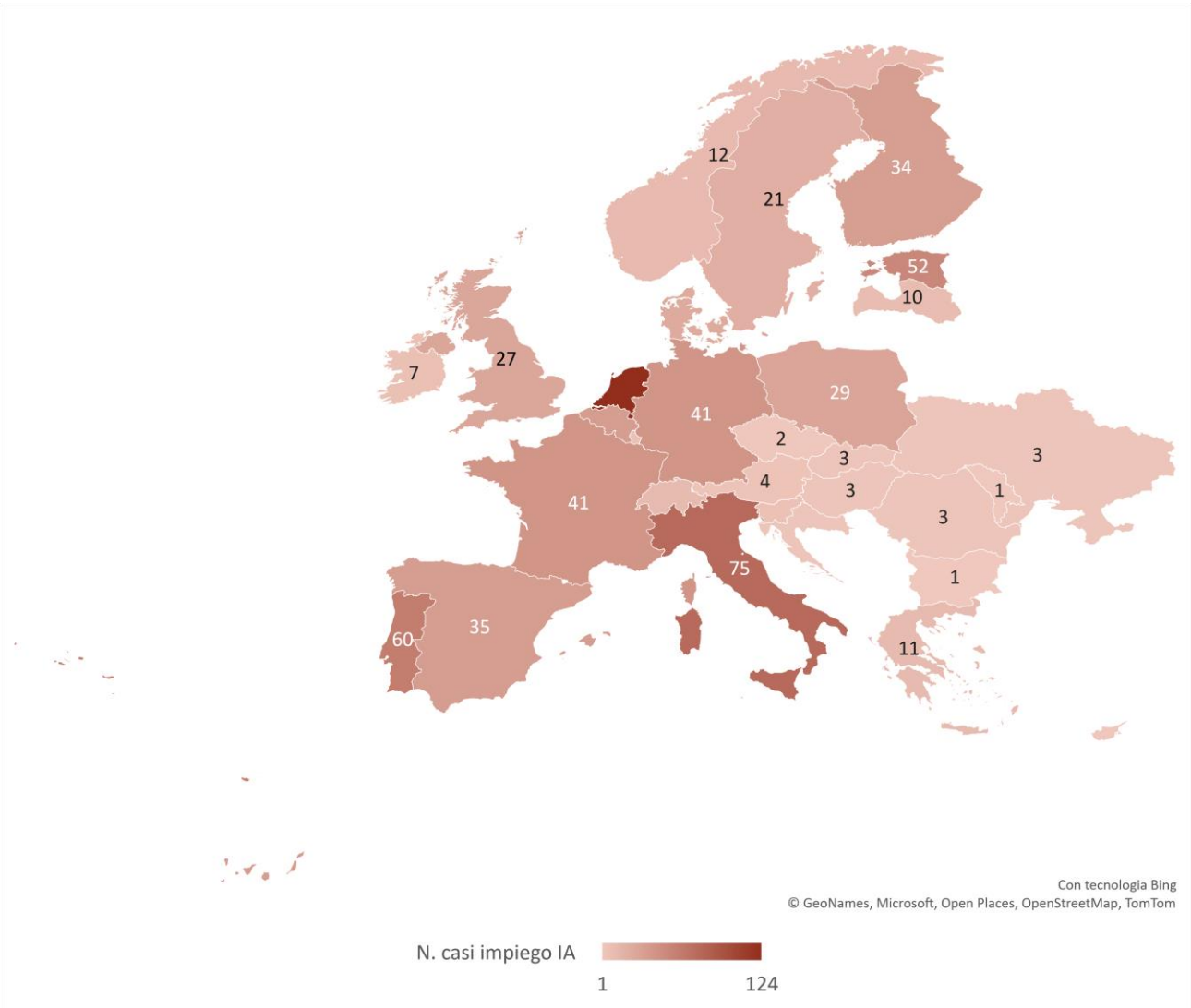


Fig.14: Numero di casi di impiego di IA per paese (2021)

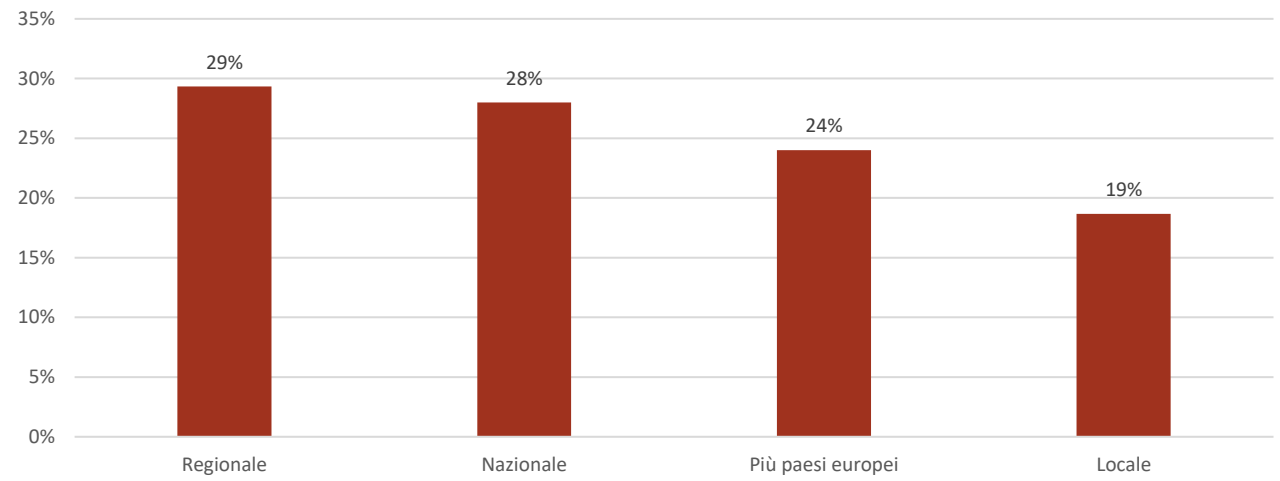
Fonte: Joint Research Center della Comissione Europea



Passando a considerare esclusivamente la situazione italiana, notiamo che un numero considerevole di casi (il 29%), è costituito da iniziative lanciate a livello regionale, seguite da iniziate nazionali (28%), iniziative che coinvolgono più paesi europei (24%) e iniziative lanciate a livello locale (19%) (Fig.15). Mentre a livello europeo lo sviluppo dell'IA nel settore pubblico sembra essere guidato principalmente dai governi nazionali, in Italia notiamo un notevole coinvolgimento delle amministrazioni regionali e locali nel promuovere lo sviluppo e l'uso delle soluzioni di intelligenza artificiale. Ciò evidenzia il fatto che le regioni, come anche le città e i piccoli comuni, possono contribuire alla diffusione di questa tecnologia in qualità di attori chiave.

Fig.15: Casi di impiego dell'IA per livello amministrativo coinvolto, in Italia (2021)

Fonte: Joint Research Center della Comissione Europea



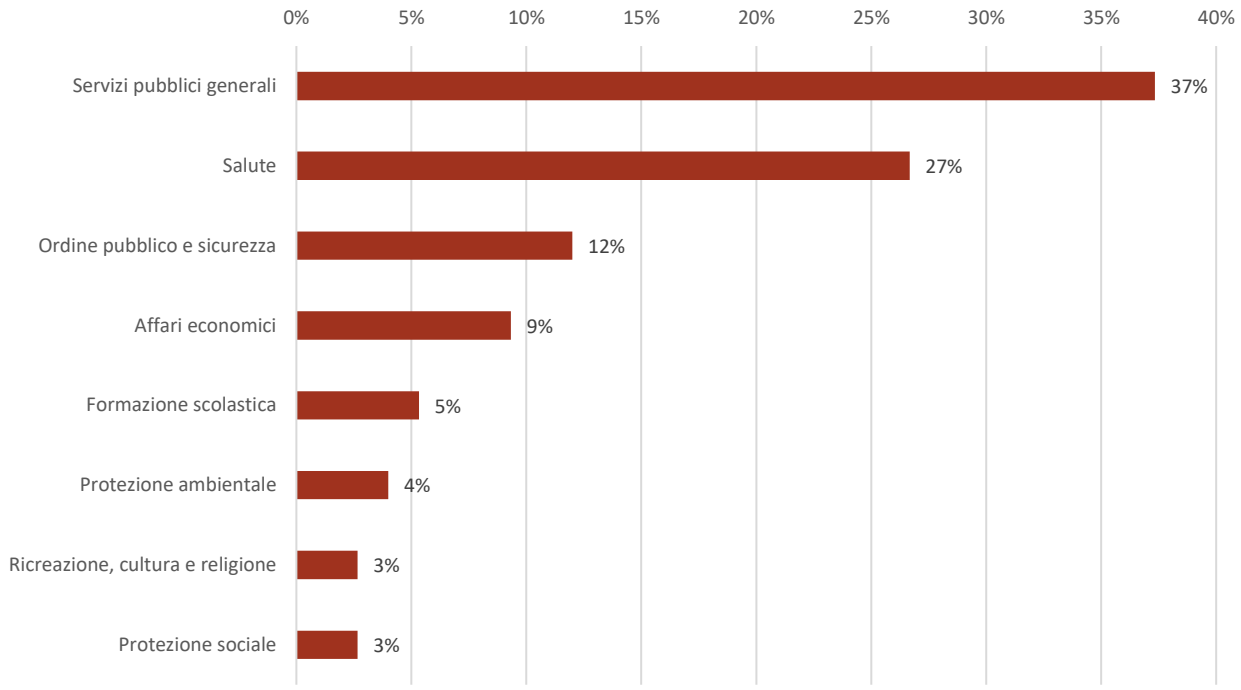
Mentre a livello europeo lo sviluppo dell'IA sembra essere guidato principalmente dai governi nazionali, in Italia notiamo un notevole coinvolgimento delle amministrazioni regionali e locali nel promuovere lo sviluppo e l'uso delle soluzioni di intelligenza artificiale

Per quanto concerne lo specifico settore di impiego dell'IA, notiamo che le applicazioni di IA sono usate principalmente nell'ambito dei servizi pubblici (37%). Ciò non sorprende in quanto è in quest'area che le amministrazioni pubbliche sono solitamente più coinvolte (Fig.16). La categoria dei servizi pubblici generali comprende diverse applicazioni: chatbot e assistenti virtuali utilizzati per interagire con i cittadini e le aziende, nonché per accelerare i processi interni; monitoraggio di vari tipi di spazi pubblici tramite telecamere, microfoni o altri sensori; rilevamento e gestione delle informazioni errate; classificazione, archiviazione e ricerca di documenti (anche scritti a mano), video e/o discorsi registrati con estrazione automatica di metadati e informazioni; rilevamento di anomalie nei dati o di potenziali frodi. Con riferimento alle aree di impiego, troviamo in seguito il settore sanitario (27%) e quello dell'ordine pubblico e della sicurezza (12%), mentre l'utilizzo dell'IA sembra essere meno diffuso nei settori della formazione scolastica, della protezione ambientale, della cultura e religione, e della protezione sociale. Ciò potrebbe dipendere anche da un maggior coinvolgimento del settore privato in queste ultime aree.

Per quanto concerne lo specifico settore di impiego dell'IA, notiamo che le applicazioni di IA sono usate principalmente nell'ambito dei servizi pubblici (37%). Ciò non sorprende in quanto è in quest'area che le amministrazioni pubbliche sono solitamente più coinvolte

Fig.16: Casi di impiego di IA per area del settore pubblico, in Italia (2021)

Fonte: Joint Research Center della Comissione Europea



Guardando agli aspetti tecnologici, il 25% dei casi di impiego di IA in Italia si riferisce all'utilizzo del machine learning, che nel settore pubblico trova applicazione in diverse aree, quali il rilevamento di frodi, l'ottimizzazione della qualità dei documenti, la previsione basata sui dati disponibili e l'automazione di compiti ripetitivi.

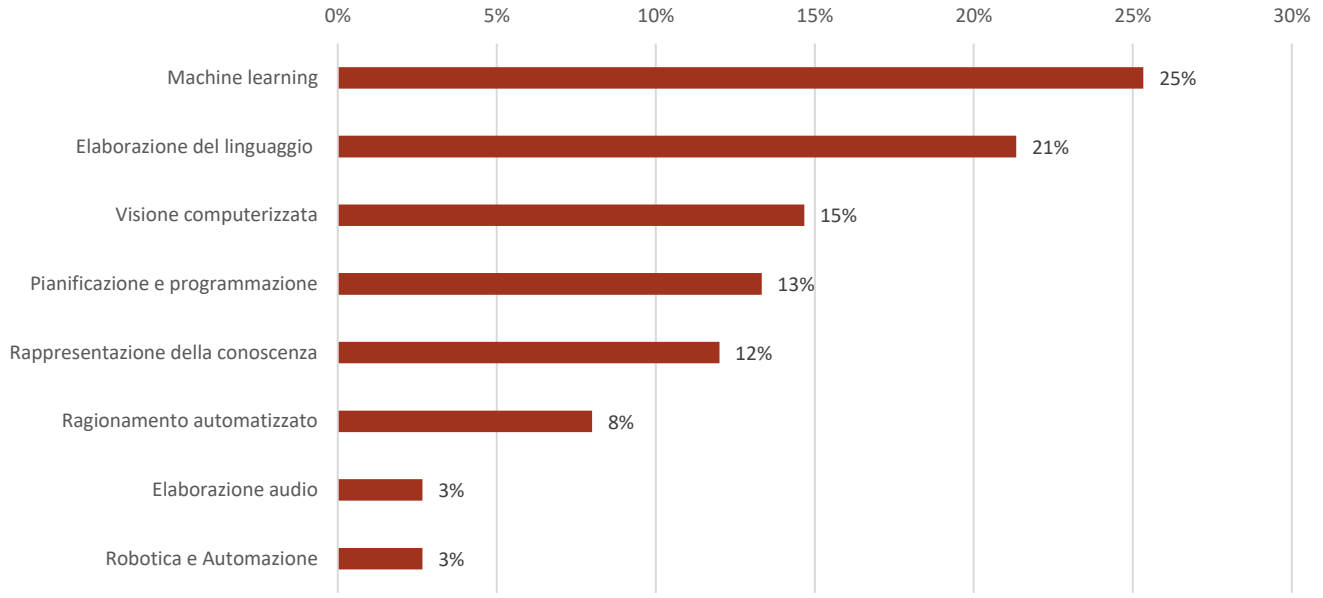
Guardando agli aspetti tecnologici, il 25% dei casi di impiego di IA in Italia si riferisce all'utilizzo del machine learning, una branca dell'intelligenza artificiale che si occupa dello sviluppo di algoritmi e modelli che consentono ai computer di apprendere da dati passati o esperienze senza essere esplicitamente programmati

Il 21% dei casi coinvolge invece l'uso di tecniche di elaborazione del linguaggio, ovvero un tipo di intelligenza artificiale che conferisce ai sistemi la capacità di identificare, elaborare, comprendere e/o generare informazioni nelle comunicazioni umane scritte e parlate. Alcuni esempi nei casi

raccolti includono servizi come chatbot e assistenti virtuali. Sono invece meno diffuse, quantomeno nel settore pubblico, le tecniche di elaborazione audio (3% dei casi) e la robotica e l'automazione (3% dei casi).

Fig.17: Casi di impiego dell'IA per tecnologia

Fonte: Joint Research Center della Comissione Europea



Tra i casi di impiego di IA più interessanti in Italia e che fanno uso del machine learning, vi è sicuramente quello del progetto *"Giurisprudenza Predittiva"* del Tribunale di Pisa. Si tratta del primo esperimento di decisioni automatizzate all'interno del sistema giudiziario, avviato nel 2019 e sviluppato presso il LiDER Lab della Scuola Superiore Sant'Anna, in collaborazione con il dipartimento di eccellenza EMbeDS, il Tribunale di Genova e il Tribunale di Pisa. L'ambizioso progetto è attualmente in via di sviluppo e prevede la creazione di una banca dati aperta in grado di aiutare la giustizia, fornendo supporto a cittadini, professionisti e giudici. Da un lato, i cittadini e i professionisti potranno avvalersi delle informazioni per approfondire le loro conoscenze giuridiche e per valutare le probabilità di successo di un determinato caso. D'altro lato, i giudici, ha sottolineato la Presidente del Tribunale di Pisa Maria Giuliana Civinini, "potranno conseguire una conoscenza più approfondita, perché analizzabile in modo aggregato, per ragioni delle decisioni e per soluzioni, della giurisprudenza in materie sensibili come quelle relative all'assegno di mantenimento in caso di separazione e divorzio o alla liquidazione del danno alla salute. Questo consentirà di adottare decisioni più consapevoli che garantiscano e realizzino al massimo grado i principi di equità, uguaglianza di trattamento e corretta presa in considerazione delle problematiche coinvolte."

Tra i casi di impiego di IA più interessanti e che fanno uso del machine learning, vi è sicuramente quello del progetto "Giurisprudenza Predittiva" del Tribunale di Pisa. Si tratta del primo esperimento di decisioni automatizzate all'interno del sistema giudiziario

A livello nazionale, merita di essere menzionato il progetto dell'INPS per la classificazione e smistamento automatico delle PEC, premiato nella Top 10 mondiale dell'Intelligenza Artificiale di IRCAI UNESCO. Il progetto nasce con l'obiettivo di efficientare il flusso di comunicazione via PEC tra INPS e cittadini. Ciò è reso possibile grazie ad un sistema basato sull'IA in grado di comprendere il contenuto delle PEC inviate all'INPS e di indirizzarle automaticamente all'operatore più appropriato. Con questo sistema l'INPS abbatte i tempi di risposta alle PEC e ottimizza il lavoro degli operatori, consentendo loro di concentrarsi su attività a maggior valore aggiunto verso gli utenti.

A livello nazionale, merita di essere menzionato il progetto dell'INPS per la classificazione e smistamento automatico delle PEC, premiato nella Top 10 mondiale dell'Intelligenza Artificiale di IRCAI UNESCO

7.3. Public procurement: dalle linee guida internazionali alla proposta di clausole contrattuali tipo della Commissione europea e lo stato dell'arte in Italia

Alla luce delle potenzialità derivanti dall'evoluzione dell'intelligenza artificiale per il miglioramento delle operazioni governative e il soddisfacimento delle esigenze dei cittadini, appare di rilievo porre l'attenzione circa l'aspetto dei processi di acquisizione delle moderne soluzioni di IA.
In virtù del fatto che tale pratica risulta spesso sfavorita a causa della particolare cautela che molte istituzioni pubbliche pongono nella scelta di impiegare questa nuova tecnologia, una prima forma di incentivo si rivede nelle linee guida internazionali per l'approvvigionamento dell'IA, sviluppate nel 2019 dal World Economic Forum.

Queste ultime esortano i governi all'elaborazione di considerazioni fondamentali prima dell'acquisizione e della distribuzione delle soluzioni e dei servizi di IA, aiutando i responsabili politici ad accelerare il raggiungimento dei loro obiettivi, i funzionari addetti agli acquisti e i team commerciali a sviluppare proposte e a gestire i processi di approvvigionamento, nonché supportando i professionisti dei dati (ad esempio, statistici, scienziati, esperti di tecnologia e digitale) nel salvaguardare il beneficio pubblico e gestire i potenziali rischi derivanti dall'algoritmo e, inoltre, offrendo mezzi adeguati ai fornitori di soluzioni di IA per comprendere meglio le aspettative di base legate ai progetti governativi in materia e al fine di allineare le loro proposte con gli standard concernenti gli appalti pubblici.

Nello specifico, le linee guida consistono in dieci raccomandazioni ordinate in termini di rilevanza per il processo di approvvigionamento, ossia:

1. Utilizzare procedure di approvvigionamento che non si concentrino sulla prescrizione di una soluzione specifica, ma piuttosto sul delineare i problemi e le opportunità e lasciare spazio all'iterazione;
2. Definire il beneficio pubblico e valutare i rischi dell'utilizzo dell'IA;

3. Cercare di inserire l'approvvigionamento all'interno di una strategia per l'adozione dell'IA valida per tutta la pubblica amministrazione;
4. Assicurarsi che la legislazione e i codici di pratica siano osservati nella richiesta di proposta;
5. Articolare la fattibilità tecnica e le considerazioni della governance per ottenere i dati rilevanti;
6. Evidenziare i limiti tecnici ed etici dell'uso dei dati per evitare problemi come i pregiudizi;
7. Lavorare con un team multidisciplinare;
8. Concentrarsi durante l'intero processo sui meccanismi di *accountability* e sulle norme di trasparenza;
9. Implementare un'interlocuzione continua tra il fornitore e l'entità acquirente per il trasferimento di conoscenze e la valutazione del rischio a lungo termine;
10. Garantire condizioni di parità tra i fornitori di soluzioni di IA.

Ad ognuna di queste linee guida sono attribuiti più principi che ne approfondiscono il contenuto, invogliando i destinatari di riferimento alla trasparenza, alla cooperazione, all'esecuzione di una previa valutazione dei rischi, alla compliance rispetto agli atti normativi vigenti in ogni Stato coinvolto, alla responsabilità, alla rendicontabilità delle decisioni assunte e all'obiettivo comune dei governi di soddisfare le aspettative dell'opinione pubblica rispetto ad una supervisione sia esperta che democratica del processo decisionale algoritmico, in assenza della quale potrebbero generarsi danni non di lieve portata.

A livello UE, il 29 settembre scorso sono state pubblicate nell'ambito della piattaforma "*Procurement of AI community*", presente sul sito web della Commissione Europea, due proposte di clausole standard per l'acquisto da parte delle pubbliche amministrazioni di sistemi di intelligenza artificiale, in conformità con le previsioni regolamentari dell'AI Act. Va innanzitutto premesso che non si tratta di un atto ufficiale dell'UE, poiché esse sono state redatte unicamente ai fini di discussione e per raccogliere i riscontri iniziali dei portatori di interessi.

*Il 29 settembre scorso sono state pubblicate nell'ambito della piattaforma "*Procurement of AI community*", presente sul sito web della Commissione Europea, due proposte di clausole standard per l'acquisto da parte delle pubbliche amministrazioni di sistemi di intelligenza artificiale*

Le clausole si ispirano a quelle già elaborate dalla città di Amsterdam nel 2018 per l'acquisto di sistemi algoritmici e sono suddivise in due documenti che si distinguono in base all'acquisizione delle due diverse categorie di sistemi di intelligenza artificiale: quelli "ad alto rischio" e quelli che non presentano un rischio elevato.

Tra le previsioni analoghe più rilevanti di entrambi gli atti vi è, *in primis*, una specifica sezione (Sezione B) denominata "Requisiti essenziali in relazione al sistema IA", che stabilisce obblighi a carico del fornitore circa il sistema di intelligenza artificiale in fase di acquisizione da parte dell'entità pubblica. Il fornitore deve garantire che prima della consegna sia istituito e attuato un sistema di gestione dei rischi basato su più fasi che si sostanziano nell'individuazione, nella stima e

nella valutazione dei rischi noti e ragionevolmente prevedibili per la salute, la sicurezza e i diritti fondamentali dell'Unione europea; nella valutazione di altri rischi che potrebbero presentarsi e nell'adozione di misure adeguate e mirate a gestirli. Peraltro, è previsto che il fornitore dichiari contrattualmente l'accettabilità del rischio residuo complessivo**,** a condizione che il sistema di IA sia usato conformemente alla finalità prevista o in condizioni di uso improprio ragionevolmente prevedibili. Inoltre, tale sistema deve essere testato prima della consegna e il fornitore deve rendersi disponibile anche ad effettuare ulteriori controlli presso i locali della committente.

È di rilievo anche l'aspetto inerente la governance dei dati, per cui l'articolo 3 dispone che il fornitore sia tenuto a adottare specifiche misure volte ad assicurare:

- La trasparenza per quanto riguarda la finalità originaria della raccolta dei dati;
- Scelte progettuali pertinenti;
- Procedure di raccolta dei dati;
- La preparazione dei dati ai fini delle operazioni di trattamento, quali annotazione, etichettatura, pulizia, arricchimento e aggregazione;
- La formulazione di ipotesi pertinenti in merito alle informazioni che i dati dovrebbero misurare e rappresentare;
- La valutazione, l'individuazione, la prevenzione di possibili distorsioni che rischiano di compromettere la salute e la sicurezza delle persone fisiche o comportare discriminazioni vietate dalla legislazione dell'Unione europea;
- L'individuazione e la risoluzione di lacune o carenze nei dati che impediscono il rispetto delle presenti clausole.

I set di dati devono rispondere ad alcuni requisiti fondamentali, ossia essere pertinenti, rappresentativi, esatti e completi rispetto alle finalità previste. Il fornitore ha l'onere di assicurare che le banche dati possiedano le proprietà statistiche appropriate, anche, ove applicabile, per quanto riguarda le persone o i gruppi di persone sui quali il sistema di IA è destinato a essere usato e che tengano conto delle caratteristiche o degli elementi particolari dello specifico contesto geografico, comportamentale o funzionale all'interno del quale l'IA è destinata a essere impiegata. Detti obblighi devono essere adempiuti sia nella fase di sviluppo del sistema che durante il suo utilizzo.

Sono inoltre riconosciuti, con modalità differenti, i rispettivi diritti di utilizzo dei set di dati coinvolti per l'addestramento del sistema di intelligenza artificiale, compresi quelli di proprietà intellettuale, a seconda che questi appartengano all'organizzazione pubblica, al fornitore o a terzi. Nella prima ipotesi vi è un divieto per il fornitore di utilizzare i dati per finalità diverse dall'esecuzione del contratto, salvo specifiche previsioni contrarie, mentre, nel secondo caso, viene previsto un diritto di uso non esclusivo degli stessi che sia sufficiente per l'osservanza delle disposizioni dell'accordo, clausole incluse, salvo disposizioni contrarie. È anche riconosciuta la possibilità di inserire una clausola, non imponibile al fornitore, che consenta alla committente di utilizzare tali dati per ulteriori sviluppi di sistemi di intelligenza artificiale.

Sono inoltre riconosciuti, con modalità differenti, i rispettivi diritti di utilizzo dei set di dati coinvolti per l'addestramento del sistema di intelligenza artificiale, compresi quelli di proprietà intellettuale, a seconda che questi appartengano all'organizzazione pubblica, al fornitore o a terzi

Altra previsione è quella di fornire, insieme alle apposite istruzioni per l'uso, anche la documentazione tecnica attestante la conformità del sistema di IA alle disposizioni contrattuali, stabilendo un obbligo di aggiornamento della stessa, che comprenda l'inserimento di ogni modifica sostanziale apportata durante il periodo di validità dell'accordo.

La garanzia di trasparenza dei sistemi di IA è espressamente disciplinata dall'art. 6, che la declina anche in termini di ragionevole comprensibilità per l'organizzazione pubblica del funzionamento e dei dati trattati da tale sistema. Accanto a tale previsione, l'art. 13 delle clausole standard per l'acquisto di sistemi IA ad alto rischio prescrive che "durante il periodo di validità dell'accordo il fornitore sia tenuto ad assistere l'organizzazione pubblica, su sua prima richiesta, per spiegare alle persone o al gruppo di persone su cui il sistema di IA è (destinato a essere) utilizzato in che modo il sistema di IA è giunto a una determinata decisione o a un determinato risultato. Tale assistenza comprenderà, come minimo, una chiara indicazione dei fattori chiave che hanno indotto il sistema di IA a giungere a un determinato risultato e delle modifiche da apportare agli input per giungere a un risultato diverso".

In aggiunta, il fornitore ha il dovere di assicurare che, prima della consegna, nel sistema di IA siano integrate misure adeguate, volte a garantire la sorveglianza umana. Queste misure devono permettere di monitorare e comprendere le capacità e i limiti del sistema, impedire l'eccessivo affidamento sui risultati dello stesso ai soggetti che lo utilizzano, essere in grado di interpretare correttamente l'output prodotto, decidere in qualsiasi situazione particolare di non usarlo o, altrimenti, di ignorarne, annullarne o ribaltarne l'output e di intervenire sul funzionamento del sistema di IA.

In ultimo, all'interno delle clausole viene richiesta accuratezza, robustezza e cybersecurity del sistema in questione, riconoscendo anche obblighi di audit che, diversamente, non sono previsti nelle clausole standard applicate per l'acquisto di sistemi IA non ad alto rischio.

In ambito nazionale, recente novità risiede nella pubblicazione ad opera dell'Agenzia per l'Italia Digitale del Piano Triennale per l'Informatica nella Pubblica Amministrazione 2024-2026, frutto di un'attività di concertazione tra amministrazioni e soggetti istituzionali. Nello specifico, il documento di programmazione strategica delinea alcuni principi generali che le pubbliche amministrazioni dovranno adottare e applicare in risposta alle sfide derivanti dall'impiego dell'IA.

Il documento di programmazione strategica delinea alcuni principi generali che le pubbliche amministrazioni dovranno adottare e applicare in risposta alle sfide derivanti dall'impiego dell'IA

Innanzitutto, gli investimenti in tale ambito dovranno puntare al miglioramento dei servizi istituzionali obbligatori e del funzionamento dell'apparato amministrativo, automatizzandone i compiti ripetitivi e implementando meccanismi di proattività. Altra prerogativa è quella dell'analisi dei rischi, per cui si prevede che le PA debbano adottare la classificazione dei sistemi di IA secondo le categorie di rischio definite dall'AI Act, al fine di evitare che questi possano provocare violazioni dei diritti fondamentali della persona o altri danni rilevanti. Con l'intento di garantire la responsabilità e la rendicontazione delle decisioni assunte con il supporto di tecnologie di IA, le amministrazioni devono concentrarsi particolarmente sui requisiti di trasparenza e interpretabilità dei modelli coinvolti, fornendo agli utenti informazioni adeguate che permettano di sviluppare scelte consapevoli. Anche l'inclusività, l'accessibilità e la sostenibilità costituiscono tre punti fondamentali del Piano strategico, da esercitare nell'osservanza dei principi di equità, trasparenza e non discriminazione. Ulteriore investimento di rilievo è quello nella formazione e nello sviluppo di competenze necessarie per gestire l'intelligenza artificiale in maniera efficace nell'ambito dei servizi pubblici. Inoltre, con riferimento alla standardizzazione, il documento richiede alle amministrazioni di tenere in considerazione le attività di normazione tecnica in corso a livello internazionale e a livello europeo da CEN e CENELEC, rifacendosi ai requisiti definiti dall'AI Act e, con riguardo all'adozione di modelli fondazionali "ad alto impatto", di accertare che tali sistemi prevedano misure di trasparenza adeguate volte a chiarire l'attribuzione delle responsabilità e dei ruoli, soprattutto dei fornitori e degli utenti della soluzione di IA in questione. Per quanto concerne i dati, le PA che acquistano servizi di intelligenza artificiale tramite API, devono valutare le modalità e le condizioni con le quali il fornitore del servizio li gestisce, tutelandone la proprietà, la privacy e la sicurezza.

Allo scopo di accrescere la consapevolezza della PA nel procurement delle tecnologie di intelligenza artificiale, tra le iniziative del Piano rilevano le seguenti:

1) redazione "linee guida per promuovere l'adozione dell'IA nella Pubblica Amministrazione", che forniranno strumenti di valutazione sull'uso dell'IA come risposta alle esigenze delle amministrazioni, illustrando casi d'uso e promuovendo buone pratiche (dicembre 2024 e conseguente adozione entro dicembre 2025);
2) redazione "linee guida per il procurement di IA nella Pubblica amministrazione", che orienteranno la scelta delle procedure di approvvigionamento e la definizione delle specifiche funzionali e non funzionali delle forniture (dicembre 2024 e conseguente adozione entro dicembre 2025);
3) redazione "Linee guida per lo sviluppo di applicazioni di IA per la Pubblica Amministrazione", che forniranno gli strumenti metodologici necessari per affrontare progetti di sviluppo di soluzioni IA, compresa la creazione di soluzioni basate su *foundation models* (dicembre 2024 e conseguente adozione entro dicembre 2025);
4) adozione, entro dicembre 2026, di applicazioni di IA a valenza nazionale e di basi di dati nazionali strategiche.

In conclusione, appare evidente come in un'era in cui l'IA si sta rapidamente integrando nelle strutture della PA, sia necessario dare impulso a uno sviluppo e un'adozione sulla base di criteri cautele efficaci tra le quali, a titolo di esempio, va inclusa la predisposizione di strumenti per l'analisi del rischio, nonché, lo sviluppo di metodologie e procedure di valutazione per applicazioni AI, la creazione e l'implementazione di linee guida sulla raccolta e il trattamento di dati all'interno

dell'amministrazione, la progettazione e l'adozione di un piano di competenze per l'AI e di un piano dei fabbisogni.

7.4. Gli incentivi per l'adozione dell'IA nel settore privato

Come appena illustrato per il settore pubblico, anche in ambito privato appare necessario delineare strumenti idonei a incentivare l'approvvigionamento di sistemi e servizi di intelligenza artificiale, che possano migliorare e rendere più efficienti i processi che aumentano la competitività di un'organizzazione e dell'intero sistema Paese. La descrizione delle forme a sostegno della Transizione 4.0 affonda le radici nel 2016, anno in cui è stato lanciato il Piano Nazionale Industria 4.0, con il quale si mirava a sostenere ed incentivare l'innovazione tecnologica del tessuto imprenditoriale italiano, caratterizzato per la maggior parte da piccole e medie imprese operanti nel settore manifatturiero e da una bassa crescita della produttività, seguendo tre linee guida principali: 1) operare in una logica di neutralità tecnologica; 2) intervenire con azioni orizzontali e non verticali o settoriali; 3) agire su fattori abilitanti.

I principali strumenti operativi erano:

- Iper e superammortamento**:** iperammortamento del 250% nel caso di acquisto di beni materiali nuovi, dispositivi e tecnologie abilitanti la trasformazione in chiave 4.0, inseriti nell'Allegato A della Legge di Stabilità 2017 - Legge n. 232/2016 (inclusi i beni che già comprendono un software necessario per il loro funzionamento) e superammortamento del 140% per l'acquisto di beni immateriali (software, sistemi e *system integration*, piattaforme e applicazioni) inseriti nell'All. B alla stessa legge, compiuti da imprese che già hanno beneficiato dell'iperammortamento;
- Beni Strumentali (Nuova Sabatini**)**: finalizzata a migliorare l'accesso al credito delle micro, piccole e medie imprese per l'acquisto di nuovi macchinari, impianti e attrezzature, consente alle imprese di ottenere un contributo a parziale copertura degli interessi su finanziamenti bancari di importo compreso tra €20 mila e €2 milioni, concessi da istituti bancari convenzionati con il MISE. Il contributo era calcolato sulla base di un piano di ammortamento convenzionale di 5 anni con un tasso d'interesse del 2,75% annuo ed era maggiorato del 30% per investimenti in tecnologie Industria 4.0;
- Credito d'imposta R&D: credito d'imposta del 50% su spese incrementali in Ricerca e Sviluppo. È riconosciuto fino a un massimo annuale di 20 mln di €/anno per beneficiario e computato su una base fissa data dalla media delle spese in Ricerca e Sviluppo negli anni 2012-2014. Rientrano nel beneficio tutte le spese relative a ricerca fondamentale, ricerca industriale e sviluppo sperimentale e, dunque, quelle relative all'assunzione di personale altamente qualificato e tecnico, contratti di ricerca con università, enti di ricerca, imprese, startup e PMI innovative, quote di ammortamento di strumenti e attrezzature di laboratorio, privative industriali, ecc.;
- Patent Box: riduzione delle aliquote IRES e IRAP del 50% dal 2017 in poi, sui redditi d'impresa connessi all'uso diretto o indiretto (ovvero in licenza d'uso) di beni immateriali sia nei confronti di controparti terze che di controparti correlate (società infragruppo) a patto che il contribuente conduca attività di R&S connesse allo sviluppo e al mantenimento dei beni immateriali;
- Misure a favore di Startup e PMI innovative: previsione di una serie di vantaggi in modo da sostenere le imprese innovative in tutte le fasi del loro ciclo di vita (es. nuova modalità di

costituzione digitale e gratuita, *Equity crowdfunding* per la raccolta di nuovi capitali di rischio, esonero dalla disciplina fallimentare ordinaria, incentivi agli investimenti in capitale di rischio quali detrazione IRPEF - per investimenti fino a €1 milione - o deduzione dell'imponibile IRES - fino a €1,8 milioni - pari al 30%).

Se questo era il punto di partenza, sono ampie e numerose le modifiche intervenute nel tempo e che hanno condotto al superamento del Piano Nazionale Industria 4.0, dapprima in favore del Piano Nazionale Impresa 4.0 (ad opera della legge di bilancio 2018) e poi del Piano Transizione 4.0, che da ultimo è stato rimodulato ad opera del PNRR, mentre all'orizzonte vi è l'evoluzione verso il paradigma 5.0.

Sono ampie e numerose le modifiche intervenute nel tempo e che hanno condotto al superamento del Piano Nazionale Industria 4.0, dapprima in favore del Piano Nazionale Impresa 4.0 (ad opera della legge di bilancio 2018) e poi del Piano Transizione 4.0, che da ultimo è stato rimodulato ad opera del PNRR, mentre all'orizzonte vi è l'evoluzione verso il paradigma 5.0

Più nel dettaglio, il piano Transizione 4.0 è stato finanziato nell'ambito della Missione 1 – Componente 2 "Digitalizzazione, innovazione e competitività del sistema produttivo", con una dotazione finanziaria di €13,381 miliardi (a cui si aggiungono €5,08 miliardi del Fondo complementare) e l'obiettivo di sostenere la trasformazione digitale delle imprese incentivando gli investimenti privati in beni e attività a sostegno della digitalizzazione attraverso il riconoscimento di un credito di imposta a fronte di: acquisto di beni materiali; acquisto di beni immateriali 4.0 (es. software avanzati); acquisto di beni immateriali tradizionali (es. software di base); attività di R&D&I; attività di formazione 4.0. I beni materiali e immateriali 4.0 soggetti al regime di incentivazione sono specificati nei due Allegati (A e B) predisposti dall'allora Ministero dello Sviluppo economico.

Dal primo gennaio 2023 è scaduto sia il regime di favore per l'acquisto di beni materiali e immateriali tradizionali che per le attività di formazione 4.0. Allo stesso tempo, sono stati previsti tagli significativi per l'acquisto di beni strumentali 4.0 (sia materiali che immateriali) così come per le attività di ricerca, sviluppo e innovazione. In particolare, per i beni materiali 4.0, si è stabilito un dimezzamento per tutte le classi di investimento: dal 40% al 20% fino a €2,5 milioni; dal 20% al 10% da €2,5 milioni a €10 milioni e dal 10% al 5% da €10 milioni a €20 milioni (che è il tetto massimo ammissibile). Il taglio è stato ancora maggiore per i beni immateriali 4.0, dal 50 al 20% (fino a un tetto di €1 milione). Mentre, è intervenuta ancora la regola del dimezzamento (dal 20 al 10%) per le attività di ricerca di base, industriale e sperimentale. La diminuzione è stata più lieve solo per le attività di innovazione tecnologica "green", alle quali si applica un credito d'imposta sceso dal 15% al 10%, prefigurando in un certo senso il cambio di paradigma verso Transizione 5.0.

Ciò premesso, sarebbe opportuno dare massima priorità ai beni immateriali 4.0, alla ricerca e sviluppo e alla formazione 4.0, trattandosi peraltro di strumenti che assumono un'utilità rafforzata qualora considerati congiuntamente. Il fatto che siano tuttora pochissime le aziende in grado di lavorare con i dati e applicare soluzioni di intelligenza artificiale è più da ascrivere alla mancanza di

competenze che alla disponibilità dei dati (o di modelli di IA), per cui se si aiutano le competenze sia internamente (es. formazione) che esternamente (es. accordi con università e centri di ricerca) e, allo stesso tempo, si favorisce l'acquisto di software per l'analisi dei dati, si agevola un salto importante all'innovazione delle imprese.

Il fatto che siano tuttora pochissime le aziende in grado di lavorare con i dati e applicare soluzioni di intelligenza artificiale è più da ascrivere alla mancanza di competenze che alla disponibilità dei dati (o di modelli di IA)

Andrebbero altresì concentrati gli sforzi sulle piccole e medie imprese e, al contempo, incoraggiata la crescita dimensionale delle micro-imprese, che, come noto, evidenziano il massimo deficit di produttività con gli altri Paesi più avanzati. Sarebbe inoltre da valutare anche una revisione delle tecnologie incluse negli allegati A e B ammessi ai crediti d'imposta (tra le quali non figurano le infrastrutture di rete abilitanti le comunicazioni tra dispositivi). Sul punto, si tratta di un vulnus da non sottovalutare ove si consideri la crescente importanza assunta da nuove tecnologie, inclusa l'IA, ma anche con riguardo all'IoT, realtà virtuale e realtà aumentata, in grado di promuovere l'efficienza operativa e ottimizzare i processi aziendali.

Andrebbero altresì concentrati gli sforzi sulle piccole e medie imprese e, al contempo, incoraggiata la crescita dimensionale delle micro-imprese, che, come noto, evidenziano il massimo deficit di produttività con gli altri Paesi più avanzati. Sarebbe inoltre da valutare anche una revisione delle tecnologie incluse negli allegati A e B ammessi ai crediti d'imposta (tra le quali non figurano le infrastrutture di rete abilitanti le comunicazioni tra dispositivi)

Le PMI italiane in questi anni hanno ricevuto molti fondi sotto forma soprattutto di crediti fiscali per l'acquisto di strumenti digitali, hardware e, con qualche difficoltà in più, software. In alcuni anni con maggiore generosità, soprattutto i primi dopo l'introduzione del piano Industria 4.0 nel 2015, grazie soprattutto a iper- e super-ammortamento. Ma, a parte i roadshow previsti da quel piano e altre iniziative che si tennero a latere, non si è data a quelle imprese, soprattutto alle più piccole, e ai loro proprietari e manager la possibilità di orientarsi prima di procedere a investire in una o più tecnologie, a partire da un'analisi dei loro fabbisogni per essere competitive rispetto alla concorrenza nazionale e internazionale.

Data la modesta entità delle spese per i software di IA, in particolare quelli di IA generativa, a fronte dei sensibili aumenti di produttività da questi resi possibili per le PMI italiane, unammisura senz'altro utile sarebbe quella di prevedere un voucher per audit tecnologici svolti da attori riconosciuti (centri di competenza, digital innovation hub, altri soggetti pubblici o privati qualificati). In questo modo si coglierebbero due piccioni con una fava: si darebbe una spinta concreta alle PMI, propedeutica all'individuazione degli investimenti più adeguati a caratteristiche

e posizione di mercato e si contribuirebbe anche a far decollare un mercato italiano dei servizi di innovazione tecnologica che stenta ad avviarsi.

Prima ancora delle attività di formazione, un progetto di audit 4.0 o IA, che per ciascuna impresa indichi il posizionamento rispetto ai benchmark innovativi del rispettivo settore (o di settori limitrofi) e suggerisca gli interventi più adatti per ridurre i gap esistenti, in termini sia di investimenti sia di interventi di formazione, serve a prioritizzare le azioni necessarie in organizzazioni prive di sufficiente expertise *in-house* – che è poi la situazione più ricorrente specie nelle piccole imprese –, evitando sprechi e aumentando così il ritorno degli investimenti sia pubblici sia privati. Con conseguente impatto positivo su produttività e crescita.

Data la modesta entità delle spese per i software di IA a fronte dei sensibili aumenti di produttività da questi resi possibili per le PMI italiane, una misura senz'altro utile sarebbe quella di prevedere un voucher per audit tecnologici svolti da attori riconosciuti. In questo modo si coglierebbero due piccioni con una fava: si darebbe una spinta concreta alle PMI, propedeutica all'individuazione degli investimenti più adeguati a caratteristiche e posizione di mercato, e si contribuirebbe anche a far decollare un mercato italiano dei servizi di innovazione tecnologica che stenta ad avviarsi

Per aumentare l'impegno delle imprese a sfruttare l'opportunità in maniera efficiente, si potrebbe immaginare una qualche forma di compartecipazione al costo delle attività finanziate (che copra per esempio il 10 o il 20% del costo dell'intervento). Inoltre, sempre per evitare un utilizzo non efficiente dei fondi da devolvere alla misura, le imprese che hanno goduto in un dato anno del voucher non ne avrebbero più diritto in quelli successivi (per esempio nei due esercizi seguenti); in questo modo, oltre a evitare sprechi, si amplierebbe la platea di aziende potenzialmente finanziabili a parità di spesa. Questa misura potrebbe non valere solo per l'IA ed estendersi anche a tecnologie correlate.

Accanto ai voucher per la valutazione e l'orientamento tecnologico delle imprese, in particolare di quelle piccole, si dovrebbero prevedere crediti d'imposta per la formazione IA del personale, senza particolari distinzioni dimensionali, sia per corsi da effettuare in-house che per programmi executive da svolgere in Italia (anche in questo caso con un impatto prevedibile sia sulla domanda che sull'offerta di formazione).

Su questa linea, quella del rafforzamento delle competenze di base, un'iniziativa che, per rispondere alle legittime preoccupazioni della premier Meloni sulle conseguenze occupazionali dell'IA, varrebbe la pena lanciare in occasione del G7 presieduto dall'Italia potrebbe riguardare l'organizzazione di corsi gratuiti online con la collaborazione dei centri di ricerca più importanti e delle principali aziende tecnologiche dei Paesi più industrializzati, nelle rispettive lingue. Una Academy nata sotto gli auspici del G7 e che sotto una regia centrale che monitori qualità e fabbisogni potrebbe essere alimentata attraverso donazioni di know-how, tempo e piattaforme multimediali.

Tornando all'ambito squisitamente nazionale, desta qualche preoccupazione l'alveo nel quale si muoverà il nuovo programma Formazione 5.0, che succede a quello 4.0, venuto meno alla fine del 2022. Tra le spese ammissibili per godere di crediti d'imposta nell'ambito di Transizione 5.0, dovrebbero infatti rientrare quelle per la formazione del personale ma entro il limite del 5% dell'investimento complessivo e solo per l'acquisizione o il rafforzamento delle skill richieste per la transizione ecologica.

Seppur la transizione ecologica e digitale siano spesso definite gemelle e certamente sono caratterizzate da complementarietà, appare evidente che ciascuna presupponga competenze specifiche, solo in piccola parte sovrapponibili. Non riconoscerlo o, peggio, non prevedere coscientemente strumenti che accompagnino il *reskilling* e *upskilling* dei lavoratori in ambito ICT sarebbe uno sbaglio enorme che ignorerebbe gli ultimi posti in classifica occupati dall'Italia a livello europeo sia sulle competenze digitali di base che in quelle avanzate.

Per sviluppare le competenze di base e specialistiche, centrale deve essere la collaborazione pubblico-privato diffusa territorialmente attraverso centri di competenza e digital innovation hub, con un ruolo cruciale specie nelle realtà provinciali più periferiche svolto dalle camere di commercio e dalle associazioni datoriali insieme a università e centri di ricerca, anche per mappare relativi fabbisogni. Ma per consentire alle Pmi di diventare credibili protagoniste dei processi di innovazione è essenziale incentivare da un lato la condivisione delle necessarie risorse di input (dati, calcolo, personale di ricerca), dall'altro processi di adozione e accesso alle infrastrutture di realtà più grandi pubbliche e private.

8. ANALISI DELLE SFIDE DA AFFRONTARE

Tante sono le opportunità legate allo sviluppo e all'adozione dell'intelligenza artificiale. Tuttavia, la tecnologia solleva anche diverse sfide, di ordine etico e legale, che vanno gestite sia a livello istituzionale ma anche per iniziativa delle stesse aziende perché la sua adozione non provochi crisi di rigetto o costi sostanziali per la società nel suo insieme.
Sappiamo ad esempio che, pur essendo strumenti estremamente sofisticati rispetto ai propri predecessori, i tool di AI generativa possono produrre a livello testuale le cosiddette «allucinazioni», come spesso sono chiamate, o con termine tecnico più corretto «confabulazioni». Anche se per qualcuno sarebbe meglio limitarsi a parlare di semplici errori, dato che chi li commette sono pur sempre delle macchine e non persone. Errori fattuali, come dati biografici, riferimenti bibliografici o giurisprudenziali del tutto inventati. Tanto più insidiosi perché appaiono all'occhio non esperto del tema specifico del tutto verosimili, traendo così facilmente in errore. Se dunque tool di IA potessero essere manomessi per dire falsità sul conto di personaggi pubblici potremmo assistere a una disinformazione molto più potente di quella fin qui vista. E in grado di riprodursi molto più velocemente.

Stesso discorso per i cosiddetti deepfakes, cioè immagini, audio e video che grazie all'IA riescono a impersonare una figura pubblica con un grado di realismo senza precedenti. Finché in questa maniera il papa viene vestito da Balenciaga o Trump corre inseguito da poliziotti, siamo più nel campo dei meme (anche se molto realistici) che di qualcosa di realmente serio. Diverso è stato il caso del falso bombardamento del Pentagono, diffuso la mattina del 22 maggio 2023, che ha fatto crollare i mercati borsistici[25]. Fortunatamente per poco tempo, ma tanto basterebbe a uno speculatore di borsa che volesse sfruttare a proprio vantaggio operazioni di questo tipo. Per fortuna l'IA, in questo come in altri casi, produce però anche i propri anticorpi, con strumenti che sono stati sviluppati per consentire di smascherare simili operazioni.

Come avviene anche nella sicurezza informatica, dove l'IA gioca sia in attacco che in difesa, nell'auspicio che quest'ultima, grazie all'azione delle istituzioni ma anche agli investimenti dei privati, possa essere il più possibile avanti rispetto agli attori malintenzionati che la mettono a rischio. Le preoccupazioni relative alla sicurezza informatica sono su vari fronti. Oltre alle minacce avanzate di ingegneria sociale e phishing, gli aggressori potrebbero utilizzare questi strumenti per generare più facilmente codici malevoli. Per molti esperti informatici, qui si nasconde la vulnerabilità principale dei modelli di IA generativa. Poiché i modelli linguistici di grande dimensione sono in grado di produrre una grande quantità di contenuti discutibili, i lavori recenti si sono concentrati sull'allineamento di questi modelli con principi etici corretti nel tentativo di prevenire generazioni indesiderate.

In un recente paper scientifico[26] viene notato che, sebbene ci siano stati alcuni successi nel bypassare i blocchi di sicurezza predisposti dalle aziende che hanno rilasciato i modelli – in gergo tecnico questi tentativi sono chiamati jailbreak a indicare la «liberazione» da misure di restrizione –, questi attacchi sembrerebbero difficili da replicare e hanno bisogno di notevole ingegno umano.

[25] *Bond, «Fake viral images of an explosion at the Pentagon were probably created by AI», NPR, 22 maggio 2023.*

[26] *. Zou, Z. Wang, J.Z. Kolter, M. Fredrikson, «Universal and Transferable Adversarial Attacks on Aligned Language Models», ArXiv, 28 luglio 2023*

Anche i tentativi di generazione automatica di richieste avversarie (in inglese adversarial attacks), cioè input tesi a ingannare un modello di apprendimento automatico, hanno avuto fortunatamente un successo limitato finora. Gli autori, tra i quali ci sono diversi informatici della Carnegie Mellon University, propongono invece un semplice ed efficace metodo di attacco che induce modelli linguistici allineati con valori e obiettivi corretti a generare comportamenti discutibili. Nello specifico, l'approccio in questione trova un suffisso che, quando associato a una vasta gamma di query, punta a massimizzare la probabilità che il modello produca una risposta affermativa (anziché rifiutarsi di rispondere). Gli autori hanno scoperto che le richieste avversarie generate da questo approccio sono piuttosto trasferibili da un modello all'altro, incluso agli LLM pubblicamente rilasciati. Il paper solleva dunque domande importanti su come si possa prevenire che tali sistemi, che sono stati preliminarmente allineati dai loro creatori proprio per impedire tale tipo di manipolazioni, producano informazioni discutibili. La domanda naturale che si pongono gli autori in conclusione è se i modelli di IA generativa possano essere esplicitamente affinati per evitare questo tipo di attacchi (che certamente richiedono un alto livello di competenza informatica ma sono assolutamente possibili). Questa è precisamente la strategia dell'addestramento avversario (adversarial machine learning), il mezzo empiricamente più efficace per rendere più robusti i modelli di apprendimento automatico a questo tipo di pericoli: durante questa fase o quella di ottimizzazione, i modelli di apprendimento vengono attaccati con uno di questi metodi per poi essere addestrati iterativamente sulla risposta «corretta» alla query potenzialmente dannosa (e preferibilmente, per aumentare ulteriormente la robustezza, anche su ulteriori richieste non potenzialmente dannose ma in qualche modo correlate).
I timori su questo versante sono così forti da stimolare sforzi senza precedenti dei governi e delle istituzioni preposte, in alleanza con il settore privato. Come è accaduto recentemente ngli USA su input diretto della stessa Casa Bianca in occasione dell'ultima DEF CON, celeberrima conferenza annuale di hacker, quando alcuni dei più potenti sistemi di IA al mondo hanno subito un attacco simultaneo da parte di un piccolo esercito di informatici intenzionati a scoprirne le vulnerabilità nascoste[27]. La Casa Bianca non solo era a conoscenza di tutto ma ha spinto perché l'happening si tenesse e le principali aziende del settore vi partecipassero, sottoponendosi all'assalto informatico. Non è certo una novità che un'istituzione pubblica promuova collaborazioni di questo tipo: succede abitualmente negli USA così come in Europa. Ma una collaborazione che parte direttamente da un ufficio politico, per giunta a capo della nazione più potente del mondo, è stato un fatto decisamente più insolito. A dimostrazione delle preoccupazioni che possono derivare da strumenti molto potenti come quelli messi a disposizione dall'IA generativa ma anche delle potenzialità della collaborazione pubblico-privato opportunamente gestita.

Su un altro versante, un'altra potenziale sfida deriva dalla necessità di garantire livelli elevati di privacy, per evitare in particolare che informazioni personali sensibili finiscano nelle mani sbagliate. Con un provvedimento che, come tutto ciò che ha riguardato l'IA generativa nell'ultimo anno, ha avuto una forte eco internazionale, il Garante italiano per la protezione dei dati personali è intervenuto il 30 marzo 2023, disponendo la limitazione provvisoria del trattamento dei dati degli utenti italiani nei confronti di OpenAI, di fatto determinando uno stop a ChatGPT, a cui venivano contestate una serie di violazioni o quantomeno ambiguità in tema privacy[28]. Dopo alcune settimane di intenso confronto, alla fine di aprile OpenAI ha riattivato per gli utenti italiani

27 M. Chatterjee, «White House sends hackers against the most powerful AIs», Politico, 10 agosto 2023.

[28] *https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847*

il servizio del proprio chatbot, introducendo una serie di misure richieste dal Garante, tra le quali un'informativa rivolta a tutti gli utenti e non, in Europa e nel resto del mondo, per illustrare quali dati personali e con quali modalità sono trattati per l'addestramento degli algoritmi e per ricordare che chiunque ha diritto di opporsi a tale trattamento in modalità facilmente accessibili. In questo modo, per chi esercita il proprio diritto, le conversazioni e la relativa cronologia possono essere escluse dal training degli algoritmi. OpenAI ha inoltre previsto per gli interessati la possibilità di far cancellare le informazioni ritenute errate, pur dichiarandosi, allo stato, tecnicamente impossibilitata a correggere gli errori.
Il lavoro e il confronto tra aziende e authority della privacy sono destinati a evolversi e a conoscere probabilmente altre tensioni. Ma le criticità appaiono risolvibili con il necessario impegno di tutti i soggetti coinvolti.

Più complesse e difficili da gestire con il quadro normativo attuale, peraltro rivisto da poco nell'Unione Europea, appaiono invece le questioni del copyright. Sotto due aspetti principali: i materiali di input impiegati per l'addestramento dei modelli e gli output prodotti dagli strumenti di IA.
Rispetto al primo profilo, i chatbot basati sull'IA generativa sono addestrati su una grande quantità di dati provenienti da internet, che possono includere materiale protetto da copyright. Sappiamo che i principali LLM usano web crawler, versioni filtrate della rete, composte da quantità enormi di token di testo. C'è molto dibattito sul fatto che questo procedimento sia lecito, senza che i detentori di diritti siano pagati, e già sono state intentate diverse cause da parte dei produttori tradizionali di contenuti. Al tempo stesso, sono state già siglate molte partnership tra aziende tecnologiche e detentori dei diritti e c'è da scommettere che molte altre saranno sottoscritte in futuro.

I timori su possibili conseguenze occupazionali negative dell'IA sono stati avanzati ben prima che i modelli generativi e l'architettura transformer sui quali sono basati fossero ideati. Anzi, se non altro l'IA generativa potrebbe essere la prima tipologia di automazione in grado di ridurre l'ineguaglianza anziché aumentarla, proprio perché si basa sul linguaggio e dunque è in grado di imitare abilità più elevate rispetto alle precedenti ondate di innovazione

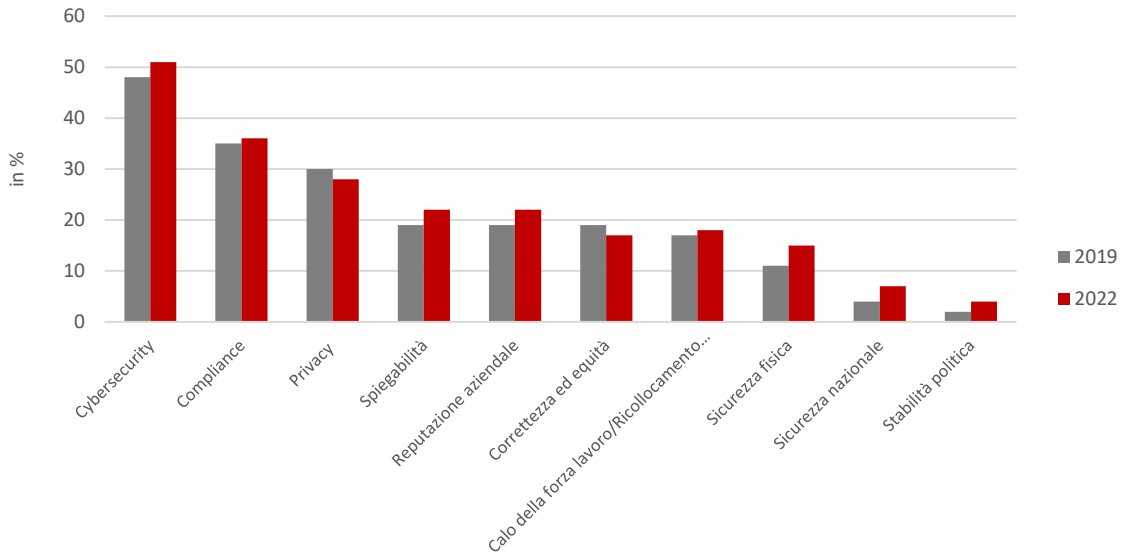
Infine, l'IA può avere implicazioni sull'evoluzione del mercato del lavoro in ragione della crescente automazione di alcune task lavorative. I timori su possibili conseguenze occupazionali negative dell'IA sono stati avanzati ben prima che i modelli generativi e l'architettura transformer sui quali sono basati fossero ideati. Anzi, se non altro l'IA generativa potrebbe essere la prima tipologia di automazione in grado di ridurre l'ineguaglianza anziché aumentarla, proprio perché si basa sul linguaggio e dunque è in grado di imitare abilità più elevate rispetto alle precedenti ondate di innovazione. Recenti studi che analizzano i possibili impatti dell'IA generativa sul mercato del lavoro, come ad esempio quello appena uscito realizzato dal dipartimento ricerca del Fondo Monetario Internazionale, guardano a due parametri chiave, esposizione e complementarietà rispetto all'IA, come fattori chiave perché gli effetti siano positivi o negativi. Se a prevalere è l'esposizione, il segno sarà necessariamente negativo. Ma se il valore aggiunto assicurato dai

lavoratori umani sarà elevato e diffuso nei diversi settori e Paesi allora il risultato netto potrà essere positivo, portando anche a un aumento dei salari grazie agli incrementi di produttività. In ultima analisi, oltre a contare sul senso etico di manager e imprenditori, per evitare che la ricerca del profitto immediato prevalga su altri fattori, bisognerà mettere i lavoratori nelle migliori condizioni per utilizzare al meglio i nuovi strumenti offerti dall'IA.

Secondo i dati rilevati da McKinsey[29], tra i principali rischi IA che le organizzazioni considerano rilevanti rientrano quelli legati alla sicurezza informatica. Oltre il 50% degli intervistati infatti ha timore di minacce informatiche correlate all'uso dell'IA. A questi poi si aggiungono i timori relativi alla privacy e alla sicurezza dei dati personali nonché rischi reputazionali (Fig.18).

Fig.18: I principali rischi IA secondo le imprese di tutto il mondo

Fonte: McKinsey & Comany (2022)

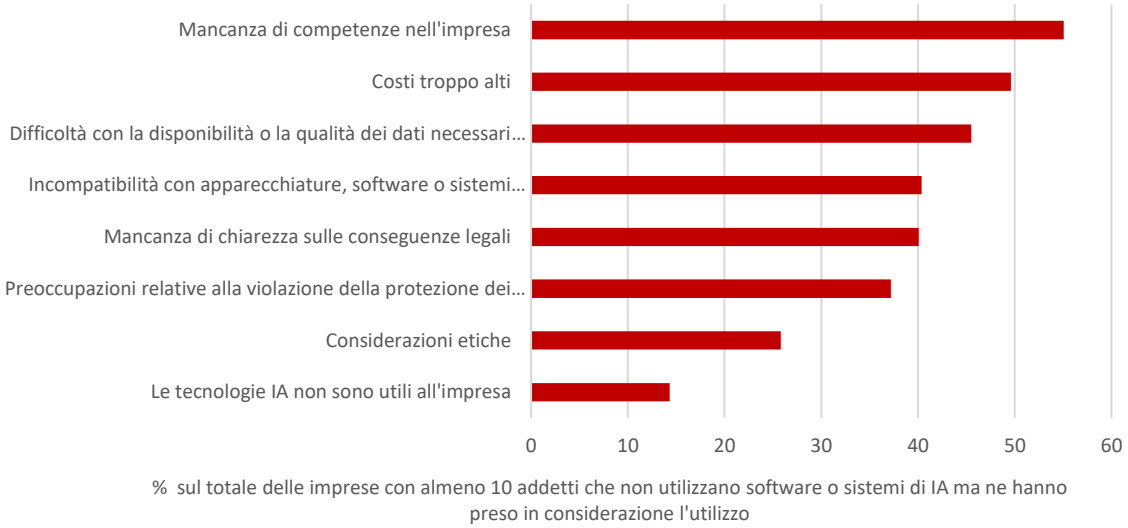


Relativamente al contesto italiano, il principale fattore frenante appare però la mancanza di competenze. La mancanza di skill costituisce un ostacolo per oltre la metà delle imprese italiane che non utilizzano ancora l'IA ma vorrebbero farlo; i costi troppo elevati sono, invece, una preoccupazione del 49,6% di imprese.
I timori relativi a violazione della privacy e alla protezione dei dati personali riguardano il 37,2% di imprese interessate a utilizzare tali tecnologie (Fig.19).

[29] *https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review#/*

Fig.19: Ostacoli all'utilizzo dell'IA da parte delle imprese italiane (% di imprese; 2023)

Fonte: Istat (2023)



Relativamente al solo settore pubblico, oltre ad indentificare i casi di impiego citati nel paragrafo 6.2, Wirtz et al. (2019) si sono anche concentrati sui possibili rischi derivanti dall'utilizzo di questa tecnologia in un contesto pubblico, identificando quattro principali categorie di sfide: sfide che derivano dall'implementazione della tecnologia nel contesto pubblico e che hanno a che fare con questioni quali la sicurezza, la sostenibilità finanziaria, l'integrazione dei dati e la specializzazione dei dipendenti; sfide regolamentari, legate a questioni legali come la privacy dei dati; sfide etiche, che fanno riferimento alle conseguenze etiche e morali dell'implementazione della tecnologia e alla questione di come incorporare principi etici nei sistemi di IA; e infine sfide sociali, ovvero legate all'accettazione di questa tecnologia da parte della società. Più recentemente, Mikalef et al. (2022) hanno focalizzato l'attenzione sulle sfide di tipo organizzativo legate all'implementazione dell'IA nelle organizzazioni pubbliche, evidenziando l'importante ruolo che il governo può ricoprire nel promuovere presso le organizzazioni l'adozione di soluzioni di intelligenza artificiale, anche attraverso finanziamenti pubblici e incentivi.

Mikalef et al. (2022) hanno focalizzato l'attenzione sulle sfide di tipo organizzativo legate all'implementazione dell'IA nelle organizzazioni pubbliche, evidenziando l'importante ruolo che il governo può ricoprire nel promuovere presso le organizzazioni l'adozione di soluzioni di intelligenza artificiale, anche attraverso finanziamenti pubblici e incentivi

L'analisi delle applicazioni, dei rischi e delle sfide organizzative, unita ai benefici sottolineati da studi specifici, sottolinea la necessità di un approccio bilanciato per massimizzare i benefici dell'intelligenza artificiale in ambito pubblico ma anche privato, garantendo al contempo la tutela dei valori fondamentali della società.

D'altronde, accanto alla regolamentazione, appare importante il coinvolgimento attivo del settore privato per internalizzare nei processi di business un approccio all'IA basato su principi di responsabilità, trasparenza e centralità delle persone. Questo processo di coinvolgimento è già in corso ed in alcuni casi è partito decisamente prima che il dibattito regolamentare entrasse nel vivo[30]. E naturalmente ha subito un'accelerazione con l'esplosione dell'IA generativa nell'ultimo anno, anche indotta dagli impegni volontari sottoscritti dalle principali aziende tecnologiche su iniziativa della Casa Bianca.
Molto lavoro, ad esempio, si sta facendo sulla provenienza e autenticità dei contenuti, che ha già portato allo sviluppo di strumenti di *watermark* per quelli sviluppati da AI generativa, come nel caso della Coalizione per la provenienza e l'autenticità dei contenuti, promossa da grandi attori tecnologici come Adobe e Microsoft ma anche da produttori di contenuti come la BBC[31] oppure il prodotto sviluppato da Google DeepMind[32].
Sui profili più generali della responsabilità, sono state promosse svariate iniziative, come ad esempio i principi formulati da Microsoft, pubblicati la scorsa estate[33], oppure l'Alleanza per un IA aperta, sicura e responsabile, lanciata verso la fine del 2023 da IBM e Meta e che include tra i membri fondatori altre aziende, startup, insieme a una vasta pluralità di istituzioni di ricerca e università[34]. La consapevolezza è che la stessa regolamentazione possa avere un reale impatto positivo solo attraverso un coinvolgimento degli attori privati, nel rispetto evidentemente dei ruoli di ciascuno.

.

[30] Ne è un esempio importante, anche se di certo non l'unico, la Call for AI Ethics, documento firmato dall'Accademia Pontificia per la Vita, Microsoft, IBM, FAO e Ministero dell'Innovazione per promuovere un approccio etico all'IA, che comprende tre aree di impatto (etica, istruzione e diritti) e sei principi (trasparenza, inclusione, responsabilità, imparzialità, affidabilità, sicurezza & privacy.
[31] https://c2pa.org/post/contentcredentials/
[32] https://deepmind.google/technologies/synthid/
[33] https://blogs.microsoft.com/eupolicy/2023/06/29/advancing-ai-governance-europe-brad-smith/
[34]https://ai.meta.com/blog/ai-alliance/

CONCLUSIONI E SPUNTI DI POLICY

I progressi nel campo dell'intelligenza artificiale hanno spinto i governi nazionali, direttamente o attraverso le organizzazioni di cui fanno parte, ad istituire commissioni e fora internazionali per occuparsi sempre di più di questa materia con riferimento alle relative opportunità, ma sollevando anche alcune preoccupazioni in merito all'etica, alla privacy, alla sicurezza e, più in generale, all'impatto sociale dell'IA e delle decisioni intraprese con l'ausilio – più o meno ampio – di questa tecnologia. L'analisi ad oggetto del presente studio si è innanzitutto soffermata sui documenti programmatici e le linee guida dei principali organismi e organizzazioni internazionali, al fine di fornire una panoramica di insieme sui numerosi contributi che cercano di guidare e indirizzare la regolazione (non vincolante) dell'IA. In questo modo, è stato possibile cogliere i punti di contatto e di potenziale divergenza tra queste iniziative.

In merito ai primi è emerso che allo stato attuale si registra un diffuso consenso sui principi etici e di sicurezza da applicare all'IA, oltre che in termini di cooperazione internazionale e collaborazione multi-stakeholder. Fa ben sperare, peraltro, che vi sia quantomeno unità di intenti nel riconoscere l'importante legame tra governance dell'IA e governance dei dati, nonché dell'importanza degli investimenti per accompagnare lo sviluppo delle tecnologie di IA con opportune competenze e infrastrutture necessarie al suo impiego a beneficio di tutti. D'altra parte, come anticipato, sono emersi alcuni punti di contrasto (o di divergenza), su cui sarà necessario continuare a discutere, e che fanno riferimento – in primo luogo – alla potenziale valenza abilitante dei modelli di IA *open-source*, i quali potrebbero contribuire a uno sviluppo dell'IA libero, aperto e trasparente, oppure se prevarrà la linea che vede questo approccio particolarmente rischioso in questo ambito e quindi si raccomanderanno (o imporranno) determinati modelli piuttosto che altri. In secondo luogo, l'approccio verso una regolamentazione rigida o, viceversa, flessibile sarà un altro aspetto dirimente per lo sviluppo e l'utilizzo dell'IA. In terzo luogo, soprattutto in seno all'ONU, sta emergendo una maggiore attenzione sull'impatto ambientale ed energetico delle soluzioni di IA, su cui non pare si sia sviluppato ancora un consenso sufficiente, come dimostrano formulazioni abbastanza generiche negli altri report esaminati. Inoltre, sarà interessante comprendere nel prossimo futuro la direzione che si intraprenderà circa le considerazioni etiche e legali applicabili specificamente all'intelligenza artificiale generativa, attualmente a uno stadio piuttosto preliminare.

La prima convenzione internazionale sull'intelligenza artificiale, attualmente in discussione presso il Comitato sull'IA (CAI) del Consiglio d'Europa, può rappresentare un passaggio storico nella regolamentazione dell'IA sul piano internazionale, anche se va evidenziato che la maggiore o minore significatività e autorevolezza di un simile quadro normativo dipenderà da quello che sarà il testo finale e, in particolare, dall'inclusione o dall'esclusione del settore privato dal rispettivo ambito di applicazione.

La maggiore o minore significatività e autorevolezza di un simile quadro normativo dipenderà da quello che sarà il testo finale e, in particolare, dall'inclusione o dall'esclusione del settore privato dal rispettivo ambito di applicazione

Seppur in un contesto di crescente interesse generale, sono diversi gli approcci attraverso i quali i vari paesi – innanzitutto Cina e USA, leader nel settore dell'IA – stanno affrontando le sfide regolamentari che i sistemi di IA pongono. Ed infatti, se la Cina si è affermata come uno dei primi paesi ad adottare una legislazione specifica attraverso cui, in linea con il regime politico vigente, governare, controllare ed indirizzare lo sviluppo e l'impiego dell'IA, gli USA, seppur con l'inaugurazione di una nuova tendenza che si è manifestata nell'adozione dell'executive order presidenziale dell'ottobre scorso, hanno tradizionalmente rinunciato all'adozione di prescrizioni, incentivando, al contrario, l'assunzione volontaria di impegni da parte delle imprese attive nel settore dell'IA nel tentativo di non ostacolare o rallentare l'innovazione.

Se la Cina si è affermata come uno dei primi paesi ad adottare una regolamentazione specifica attraverso cui, in linea con il regime politico vigente, governare, controllare ed indirizzare lo sviluppo e l'impiego dell'IA, gli USA, seppur con l'inaugurazione di una nuova tendenza che si è manifestata nell'adozione dell'executive order presidenziale dell'ottobre scorso, hanno tradizionalmente rinunciato all'adozione di prescrizioni, incentivando, al contrario, l'assunzione volontaria di impegni da parte delle imprese attive nel settore dell'IA nel tentativo di non ostacolare o rallentare l'innovazione

Nella partita globale, l'UE, con l'adozione dell'AI Act, di cui si attende la pubblicazione del testo definitivo, punta ad acquisire un ruolo di guida nella regolamentazione così come accaduto con il GDPR per la tutela dei dati personali. Si tratta di una scelta, quella di adottare un regolamento, che se da un lato sicuramente favorirà la certezza del diritto garantendo uno sviluppo armonico dell'IA ed un'efficace tutela dei diritti fondamentali degli individui, dall'altro si scontra con la rapidità dell'evoluzione tecnologica che pone continue sfide con una velocità senza precedenti (si pensi ad esempio al fenomeno dell'IA generativa che non esisteva nella proposta della Commissione europea).
Anche l'Italia si trova a giocare la propria partita nel tentativo di non restare indietro. Ci troviamo in un momento cruciale in cui il Paese continua a soffrire un ritardo nelle competenze e nella consapevolezza - soprattutto per quanto concerne le PMI - circa i benefici del digitale in generale e dell'IA in particolare e si trova impegnato nella posa di reti fisse e mobili performanti in grado di assicurare una connettività adeguata alla complessità dei sistemi di IA. In questo contesto, il Programma Strategico per l'Intelligenza Artificiale (IA) 2022-2024 ha mostrato i suoi limiti ed in particolare, il possesso di un orizzonte temporale e spaziale troppo ridotti (tre anni e un focus preponderante sulla ricerca e sviluppo) e la mancanza di un budget e di una governance dedicati, che l'attuale Governo auspicabilmente cercherà di superare con un intervento tempestivo ed efficace in grado di dare, finalmente, al paese quell'impulso necessario a diventare un'eccellenza nel settore dell'IA.

Il Programma Strategico per l'Intelligenza Artificiale (IA) 2022-2024 ha mostrato i suoi limiti ed in particolare, il possesso di un orizzonte temporale e spaziale troppo ridotti (tre anni e un focus preponderante sulla ricerca e sviluppo) e la mancanza di un budget e di una governance dedicati, che l'attuale Governo auspicabilmente cercherà di superare con un intervento tempestivo ed efficace in grado di dare, finalmente, al paese quell'impulso necessario a diventare un'eccellenza nel settore dell'IA

D'altronde l'attenzione ed il dibattito sull'IA sono al centro dell'agenda politica e dell'attività anche del Parlamento, dove il principale partito di opposizione ha presentato nei mesi scorsi un ddl che persegue il condivisibile fine di assicurare, in una logica di tutela degli individui, la chiara identificazione dei contenuti prodotti dall'IA.

L'impiego dell'IA nel settore pubblico può agevolare il raggiungimento degli obiettivi principali della pubblica amministrazione, ossia rendere le operazioni governative più efficienti e soddisfare al meglio le esigenze dei cittadini. Ma anche stimolare il settore privato, attraverso forme di procurement innovativo, con una quota (anche in partnership) riservata a startup e PMI innovative, a proporre soluzioni che siano successivamente scalabili a livello nazionale o internazionale. Pertanto, è necessario incentivare le procedure di approvvigionamento di questa nuova tecnologia.

L'impiego dell'IA nel settore pubblico può agevolare il raggiungimento degli obiettivi principali della pubblica amministrazione, ossia rendere le operazioni governative più efficienti e soddisfare al meglio le esigenze dei cittadini

In ambito internazionale, detto obiettivo prende forma grazie alle linee guida per l'approvvigionamento dell'IA, sviluppate dal World Economic Forum, che mirano a sostenere tutte le parti coinvolte nel ciclo di vita del procurement, tra cui funzionari politici, responsabili degli acquisti, scienziati dei dati, fornitori di tecnologia e i loro dirigenti. In ambito UE, sono state pubblicate due proposte di clausole standard, che hanno tentato di anticipare i requisiti ed obblighi stabiliti dal Regolamento sull'intelligenza artificiale, con previsioni differenziate in base all'acquisto di sistemi di IA ad alto rischio e che non presentano un rischio elevato. A livello nazionale, Il 12 febbraio 2024 l'Agenzia per l'Italia Digitale ha pubblicato il Piano Triennale per l'Informatica nella Pubblica Amministrazione 2024-2026 che, tra le varie previsioni, detta i principi generali che le pubbliche amministrazioni dovranno adottare e applicare in risposta alle sfide derivanti dall'impiego dell'IA, attribuendo particolare rilievo alla valutazione dei rischi, alla governance dei dati e alle iniziative di formazione rivolte ai soggetti coinvolti.

Data l'importanza che l'intelligenza artificiale ha assunto nelle dinamiche di mercato, appare necessario incentivarne l'approvvigionamento da parte delle singole organizzazioni private,

affiancando all'acquisizione un'idonea formazione che garantisca un impiego consapevole e competente di questa nuova tecnologia. A partire da un necessario orientamento che consenta alle imprese più piccole di investire consapevolmente in nuove tecnologie, in base alle proprie esigenze.

Data l'importanza che l'intelligenza artificiale ha assunto nelle dinamiche di mercato, appare necessario incentivarne l'approvvigionamento da parte delle singole organizzazioni private, affiancando all'acquisizione un'idonea formazione che garantisca un impiego consapevole e competente di questa nuova tecnologia. A partire da un necessario orientamento che consenta alle imprese più piccole di investire consapevolmente in nuove tecnologie, in base alle proprie esigenze

Andrebbero dunque previste e rese strutturali iniziative come un voucher per l'acquisto di servizi di consulenza per l'innovazione da soggetti accreditati, a partire da un audit 4.0 (o a questo punto 5.0) basato su un benchmarking dello stato delle tecnologie in uso rispetto ai migliori competitor del proprio settore e un piano concreto per scalare il proprio livello tecnologico.

Per aumentare l'impegno delle imprese a sfruttare l'opportunità in maniera efficiente, si potrebbe immaginare una qualche forma di compartecipazione al costo delle attività finanziate (che copra per esempio il 10 o il 20% del costo dell'intervento). Inoltre, sempre per evitare un utilizzo non efficiente dei fondi da devolvere alla misura, le imprese che hanno goduto in un dato anno del voucher non ne avrebbero più diritto in quelli successivi (per esempio nei due esercizi seguenti); in questo modo, oltre a evitare sprechi, si amplierebbe la platea di aziende potenzialmente finanziabili a parità di spesa. Questa misura potrebbe non valere solo per l'IA ed estendersi anche a tecnologie correlate.

Accanto ai voucher per la valutazione e l'orientamento tecnologico delle imprese, in particolare di quelle piccole, si dovrebbero prevedere crediti d'imposta per la formazione IA del personale, senza particolari distinzioni dimensionali, sia per corsi da effettuare in-house che per programmi executive da svolgere in Italia (anche in questo caso con un impatto prevedibile sia sulla domanda che sull'offerta di formazione).

Su questa linea, quella del rafforzamento delle competenze di base, un'iniziativa che, per rispondere alle legittime preoccupazioni della premier Meloni sulle conseguenze occupazionali dell'IA, varrebbe la pena lanciare in occasione del G7 presieduto dall'Italia potrebbe riguardare l'organizzazione di corsi gratuiti online con la collaborazione dei centri di ricerca più importanti e delle principali aziende tecnologiche dei Paesi più industrializzati, nelle rispettive lingue. Una Academy nata sotto gli auspici del G7 e che sotto una regia centrale che monitori qualità e fabbisogni potrebbe essere alimentata attraverso donazioni di know-how, tempo e piattaforme multimediali.

Per sviluppare le competenze di base e specialistiche, centrale deve essere infatti anche in una prospettiva nazionale la collaborazione pubblico-privato, diffusa il più possibile omogeneamente dal punto di vista territoriale attraverso un'adeguata pervasività di centri di competenza e digital innovation hub, con un ruolo cruciale specie nelle realtà provinciali più periferiche svolto dalle camere di commercio e dalle associazioni datoriali insieme a università e centri di ricerca, anche per mappare relativi fabbisogni. Ma per consentire alle Pmi di diventare credibili protagoniste dei processi di innovazione è essenziale incentivare da un lato la condivisione delle necessarie risorse di input (dati, calcolo, personale di ricerca), dall'altro processi di adozione e accesso alle infrastrutture di realtà più grandi pubbliche e private. Solo la sperimentazione di nuovi sentieri di sviluppo può portare l'Italia su un percorso di modernizzazione. Rispetto al quale il PNRR non può che essere un primo passo di una strada difficile ma al tempo stesso necessaria.