Le priorità della Presidenza italiana del G7 sull'Intelligenza Artificiale

Di Francesco Di Ciommo[1] e Camilla Scarpellino[2]

Policy Paper n. 03/2024

La regolazione dell'Intelligenza Artificiale costituisce la sfida della classe dirigente mondiale, come dimostra l'avvio dell'Hiroshima AI Process del G7 su iniziativa della presidenza giapponese, l'approvazione da parte del Parlamento europeo dell'Artificial Intelligence Act, la pubblicazione dell'Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence da parte della Casa Bianca. La corsa alla regolazione è tuttavia solo all'inizio, la governance dell'IA soffre ancora di lacune applicative, alcune delle quali verranno risolte dalle istituzioni attraverso la pubblicazione di ulteriori atti normativi, best practices and FAQ, altre invece richiederanno un maggiore impegno nello studio e nell'analisi del fenomeno. Benchè, infatti, possano essere rinvenute caratteristiche comuni a tutti i sistemi di IA oggi sarebbe più corretto parlare di Intelligenze Artificiali, ognuna caratterizzata da un proprio contesto operativo, finalità, ambito applicativo e analisi di impatto. Per contrastare la diffusione e la varietà del fenomeno sarebbe quindi necessario un approccio specializzato ed un corredo informativo completo sui singoli sistemi che nessuna istituzione ad oggi potrebbe possedere. Pertanto, l'Italia, a capo del G7 per l'anno 2024, dovrebbe impegnarsi nell'individuazione di soluzioni volte a ridurre il divario tra la società e gli operatori tech, per acquistare una maggiore fiducia su questa nuova tecnologia attraverso una maggiore conoscenza del fenomeno ed educazione del pubblico rispetto alle potenzialità ed ai rischi dell'IA.

[1] Membro del Comitato Scientifico del Policy Observatory, Professore Ordinario di Diritto Privato e prorettore per le relazioni con gli Alumni presso la Luiss Guido Carli;
[2] Coordinatrice del Policy Observatory, Ph.D. Candidate in Law and Business presso la Luiss Guido Carli;

Introduzione

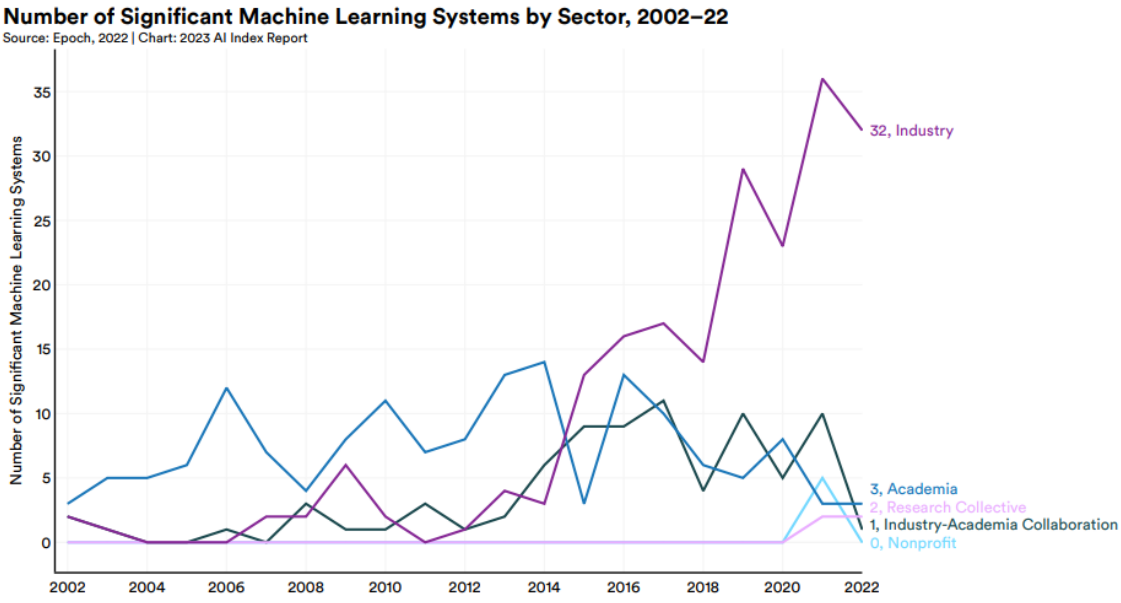
L'Intelligenza Artificiale rappresenta un'invenzione di portata rivoluzionaria, in quanto permetterebbe di sostituirsi all'uomo anche nel ragionamento, ambito del quale aveva sempre detenuto il monopolio. Di fatto, l'IA riesce ad ultimare in brevissimo tempo ricerche ed analisi elaborando conclusioni molto (spesso) corrette e risultando, quindi, un utile strumento di lavoro. La prassi economica ha, infatti, già dimostrato di voler e saper sfruttare le nuove tecnologie, ad oggi già si affida all'IA la procedura di selezione del personale o opera come prima risorsa nel customer service[3], ne sono un esempio i chatbot e gli assistenti virtuali; tuttavia, la maggior parte dei sistemi di IA sono per ora impiegati in mansioni automatiche e ripetitive.

Nondimeno, la presidenza giapponese del G7 ha avviato l'Hiroshima AI Process, un progetto che impegna le sette potenze mondiali nel promuovere uno sviluppo responsabile dell'Intelligenza Artificiale. Infatti, la dichiarazione rilasciata dai sette Capi di Stato riconosce le opportunità di innovazione e di trasformazione di questa nuova tecnologia, al contempo però ritengono necessario procedere alla definizione di un modello di governance dell'IA che gestisca i rischi e tuteli gli individui, la società e i valori condivisi dalle sette nazioni (come lo stato di diritto e la democrazia), da una prospettiva strettamente antropocentrica[4]. La governance dell'IA presenta in effetti diverse problematicità, in parte dovute alle caratteristiche intrinseche delle tecnologie smart come, ad esempio, la facile permeabilità di internet e la presenza di contenuti illeciti della rete, che potrebbero corrompere con dati erronei l'apprendimento dei sistemi di IA. Peraltro, l'estensione globale della rete rende qualsiasi soluzione "nazionale" poco efficiente, infatti le maggiori aziende tech (Apple, Microsoft, Huawei, Samsung ecc..) operano su scala internazionale. L'esigenza di un modello di governance condiviso a livello internazionale è stata espressa recentemente dalla Dichiarazione dei Ministri dell'Industria, della Tecnologia e del Digitale che ha espresso il suo appoggio a cooperazioni internazionali per lo sviluppo dell'economia digitale. I Ministri hanno anche evidenziato l'importanza di un dialogo internazionale sulla Governance dell'IA e sull'interoperabilità dei diversi modelli di governance, in modo da garantire un'accountability in grado di raccogliere i dati sul funzionamento del singolo device da ogni parte del mondo[5]. Questo obiettivo è stato per l'appunto perseguito dai membri del G7 con l'Hiroshima AI Process, nel quale i ministri competenti hanno elaborato un Codice di condotta e principi internazionali che guidino l'industria, la ricerca e le istituzioni nello sviluppo di sistemi di IA sicuri ed affidabili. I Principi Guida Internazionali pubblicati nell'Hiroshima AI Process stabiliscono innanzitutto il dovere di predisporre misure adeguate ad

[3] Alcuni dei Software di IA impiegati dalle aziende: Viso Suite Platform Software, ChatGPT Software Software, Jupyter Notebooks Software, Google Cloud AI Platform Software, Azure Machine Learning Studio Software, Infosys Nia Software, Salesforce Einstein , leggi di più al link https://viso.ai/deep-learning/ai-software/ ;
[4] G7 Leaders' Statement on the Hiroshima AI Process October 30, 2023 in https://www.soumu.go.jp/hiroshimaaiprocess/en/index.html ;
[5] Dipartimento per la Trasformazione Digitale, G7 Industry, Technology And Digital Ministerial Meeting, Verona and Trento 14-15 March 2024, https://innovazione.gov.it/notizie/articoli/en/g7-ministerial-declaration-on-industry-technology-and-digital/ ;

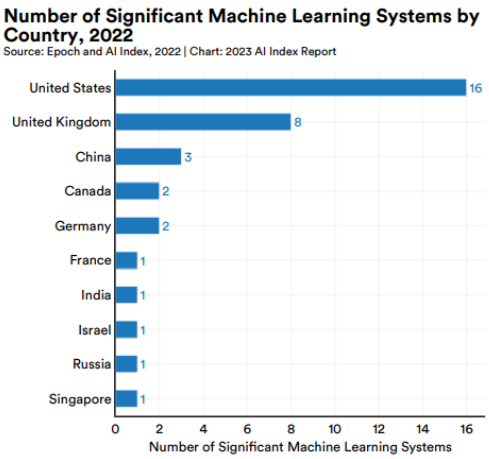
«identificare, valutare e mitigare i rischi durante tutto il ciclo di vita dell'intelligenza artificiale»[6], questo approccio coincide con la politica sull'IA inaugurata dall'Unione Europea con l'AI Act[7]. Infatti, il Regolamento UE sull'Intelligenza Artificiale, attualmente in attesa di voto, fonda il proprio modello di governance sull'identificazione dei sistemi di IA ad alto rischio, ai quali impone obblighi di compliance volti a delineare e analizzare «i rischi noti e ragionevolmente prevedibili che il sistema di IA ad alto rischio può porre per la salute, la sicurezza e i diritti fondamentali»[8].

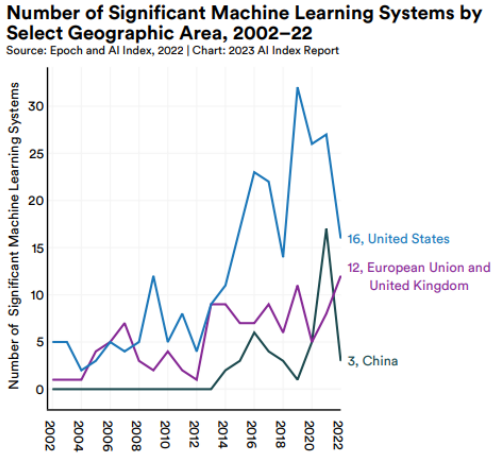
I leader hanno poi invocato la collaborazione dei privati e della ricerca nell'applicazione di queste raccomandazioni, il loro ruolo sarà inoltre decisivo nell'aggiornamento di questi documenti, che richiederanno consultazioni e audizioni per conoscere la prospettiva degli stakeholders sulle esigenze e problematiche del processo di produzione e sviluppo dei software. Alcune iniziative in tal senso sono state promosse dall'OCSE, dal Global Partnership on Artificial Intelligence e dall'UNESCO, nel Global Challenge on Trust in the Age of Generative AI[9]. In effetti sino al 2014 sono state per lo più le Università e gli enti di ricerca a sviluppare i sistemi di machine learning, mentre al 2022 sono stati 32 i progetti su IA condotti dalle industrie.



[6] Cit. Principle 1 of I Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System https://www.mofa.go.jp/files/100573471.pdf .
[7] L'ultima versione disponibile sul documento approvato dal Parlamento europeo "Emendamenti Del Parlamento Europeo alla proposta della Commissione" il 13 marzo 2024 è disponibile al link https://www.europarl.europa.eu/plenary/en/report-details.html?reference=A9-0188-2023 ;
[8] Cit. Art. 9 "Sistema di gestione dei rischi" del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione";
[9] Maggiori informazioni sono disponibili nel sito globalchallenge.ai ;

L'affiliazione geografica di questi sistemi è tutt'altro che variegata, infatti 16 machine learning systems, più della metà dei software, è stata sviluppata negli Stati Uniti, seguiti dal Regno Unito dal quale ne provengono 8 e dalla Cina con tre.



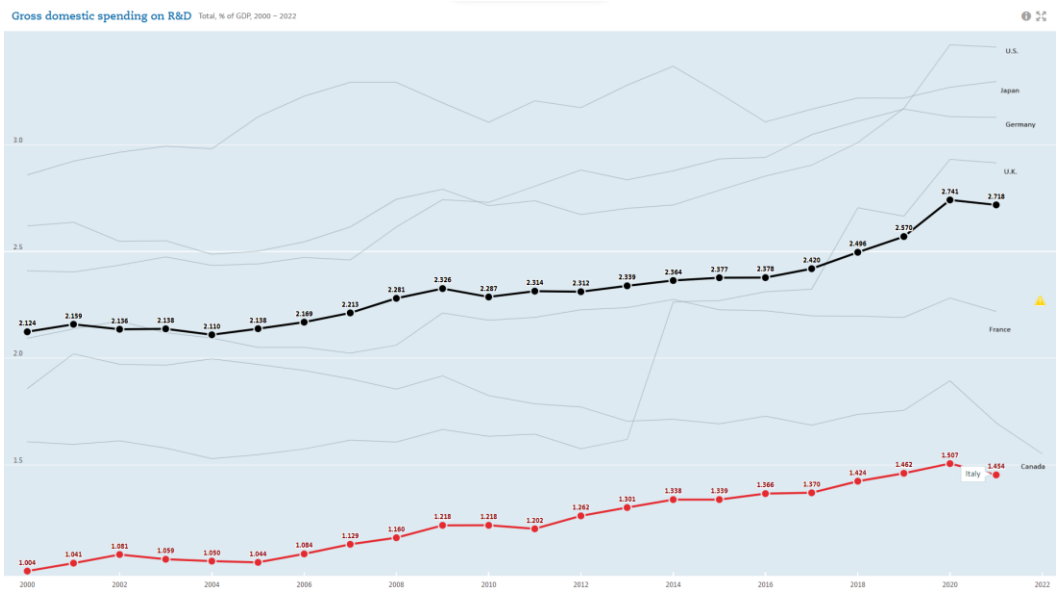


L'Italia quindi non rappresenta per ora un punto di riferimento per lo sviluppo di nuovi AI tools, tuttavia potrà presentarsi come precursore nella definizione di un quadro normativo adeguato alla gestione e sviluppo dell'IA. Questo obiettivo impegnerà il Governo italiano sia sul fronte interno che sul fronte internazionale quale nuovo presidente in carica del G7. Infatti, proprio nel recente incontro dei Ministri dell'industria e della Tecnologia del G7 è stata confermata la volontà di proseguire il percorso di governance and compliance avviato con l'Hiroshima Process predisponendo un meccanismo che consenta di monitorare l'applicazione del Codice di Condotta e dei Principi Guida Internazionali da parte di organizzazioni internazionali ed imprese[10]. Infine, ha annunciato l'intenzione di redigere un Compendio dei Servizi Digitali Governativi con particolare focus sull'identità digitale al fine di stimolare la discussione internazionale sull'efficientamento e la resilienza della pubblica amministrazione raggiunta attraverso strumenti di e-governance[11].

[10] Dipartimento per la trasformazione digitale, Op. Cit. punti 52 e 53;
[11] Dipartimento per la trasformazione digitale, Op. Cit. punti Annex 4;

La strategia dell'Italia per l'Intelligenza Artificiale

Lo sviluppo dell'Intelligenza Artificiale cd. Generale e della Intelligenza Artificiale Generativa hanno reso più tangibili le preoccupazioni legate all'IA quale disruptive technology per la struttura economica e sociale del nostro secolo. Infatti, una tecnologia in grado di competere a livello intellettivo con l'essere umano, presenta un numero ampio di rischi di lesione dei diritti individuali e collettivi. I settori maggiormente a rischio sono quelli direttamente connessi alla tutela e realizzazione della persona e che possiamo individuare nel mercato del lavoro, istruzione, sanità, immigrazione, pubblica assistenza, amministrazione della giustizia e processi democratici. Tuttavia, se da un lato preoccupa l'impatto delle nuove tecnologie perché capaci di arrecare pregiudizi ai diritti fondamentali, il rovescio della medaglia è molto più accattivante, in quanto le applicazioni dell'IA consentirebbero un più largo e facile accesso ai servizi della salute e dell'educazione su scala globale.
Al momento però l'Italia non è in testa nella corsa alle nuove tecnologie, il ritardo può essere addebitato a diversi fattori, tra i quali figura in particolar modo il ridotto investimento sulla ricerca e sviluppo dell'IA da parte di aziende, istituti di ricerca, laboratori universitari e governativi.



Secondo le statistiche OCSE riportate in figura, L'Italia è ultima tra i paesi del G7 per investimenti su ricerca e sviluppo, i dati prendono in considerazione anche gli investimenti esteri, sono invece esclusi i finanziamenti nazionali svolti al di fuori dell'economia domestica.

Per recuperare il gap il "Programma Strategico Intelligenza Artificiale 2022-2024" varato dal Governo italiano si è concentrato su tre linee di intervento:

- rafforzare **le competenze e attrarre talenti** per sviluppare un ecosistema dell'intelligenza artificiale in Italia;
- aumentare i **finanziamenti** per la ricerca avanzata nell'IA;
- incentivare l'adozione dell'IA e delle sue applicazioni, sia nella pubblica amministrazione (PA) che nei settori produttivi in generale[12].

Gli strumenti messi in campo per l'attuazione di queste politiche prevedono: crediti d'imposta o voucher per l'assunzione di profili STEM, appalti pubblici alle start-up per l'acquisto di beni e servizi, strumenti di sperimentazione normativa, campagne di informazione sull'IA per le imprese. Sul fronte della preparazione e ricerca dei talenti invece ha finanziato un Dottorato Nazionale in "Intelligenza Artificiale" (PhD-AI.it) e promuove corsi e carriere nelle materie STEM. Inoltre, ha previsto meccanismi di sostegno alla ricerca tramite bandi di ricerca-innovazione IA per collaborazioni pubblico-private, ha istituito cattedre dedicate alla ricerca sull'IA e ha bandito iniziative IA-PRIN. Questi progetti trovano quale ulteriore fonte di finanziamento i fondi del Piano Nazionale di Ripresa e Resilienza la cui Missione 4, alla quale sono stati assegnati 30, 9 miliardi di euro, persegue l'obiettivo di ridurre le carenze strutturali e sistemiche dell'apparato di istruzione di tutti gli ordini e gradi. Inoltre, lo sviluppo dell'IA potrebbe senz'altro trovare fondi nella Missione 1 del PNRR sezione "Digitalizzazione, innovazione e competitività nel sistema produttivo" (M1C2), alla quale sono stati assegnati complessivamente ulteriori 30,57 miliardi di euro alla[13]. Infine, si menziona l'istituzione di un Fondo Nazionale di Innovazione da parte del Ministero delle Imprese e del Made in Italy, il fondo è un soggetto (SGR) multifondo che opera esclusivamente attraverso metodologie venture capital[14]. Il Ministero ha individuato CDP Venture quale ente di gestore del fondo al quale destinare la somma di 800 milioni di euro per il sostegno e la promozione di start-up attive sull'IA[15]. Accanto a queste forme di finanziamento si auspica lo sviluppo della digital industry assicurando una catena produttiva globale e resiliente, non solo nel settore dell'IA ma si riconosce un ruolo chiave anche ai semiconduttori ed alle tecnologie quantistiche. Per un concreto sviluppo

[12] Cit. B. Caputo, I. Castiglioni, M. Conti, R. Cucchiara, Juan Carlos de Martin, Fosca Giannotti, Giuseppe Magnifico, Michela Milano, Giovanni Miragliotta. ,Programma strategico Intelligenza Artificiale 2022-2024, a cura del Ministero dell'Università e della Ricerca, del Ministero dello Sviluppo Economico e del Ministro per l'Innovazione tecnologica e la Transizione Digitale, 24 novembre 2021, https://assets.innovazione.gov.it/1637777289-programma-strategico-iaweb.pdf ;

[13] Governo Italiano Presidenza del Consiglio dei Ministri, PNRR: digitalizzazione, innovazione, competitività, cultura e turismo, in https://www.governo.it/it/approfondimento/digitalizzazione-innovazione-competitivit-e-cultura/16701 ;

[14] Programma strategico Intelligenza Artificiale 2022-2024, cit p. 30 ;

[15] C. Fotina, Intelligenza artificiale, il governo parte da Authority e start up, in Sole 24 Ore, 2 novembre 2023, p. 8 ;

di questo settore produttivo è necessaria una rete internet aperta, libera, globale, interoperabile e sicura[16], pertanto si dovrà potenziare la rete di infrastrutture di comunicazione internazionali come i cavi sottomarini transoceanici.

Viceversa, l'Italia ha riscontrato maggior successo nelle politiche normative sull'IA, posizionandosi quarta per numero di atti normativi che menzionano l'Intelligenza Artificiale.

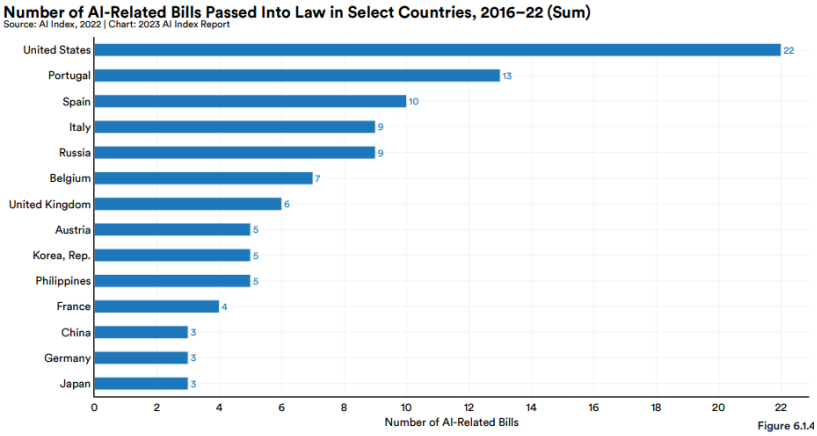


Figure 6.1.4

La governance dell'Intelligenza Artificiale dovrà coinvolgere diverse expertise in considerazione degli infiniti settori applicativi e della complessità dei singoli software. L'estensione del fenomeno e la varietà di utilizzi rende difficile concepire un quadro normativo completo e uniforme, infatti ad oggi possiamo enumerare già vari regolamenti e direttive provenienti dall'Unione Europea che intervengono incidentalmente sui software di IA come il GDPR, il DSA e il DMA; nonché altri atti intervenuti a regolare i servizi della società dell'informazione[17]. A tali iniziative vanno aggiunte: la legislazione di settore sui singoli prodotti e servizi, l'attuale AI Act in discussione, le proposte direttive sulla responsabilità civile dei sistemi di Intelligenza Artificiale e sulla riforma della direttiva sulla responsabilità da prodotto difettoso. Ad oggi, quindi, non solo sarebbe necessario una regolazione specifica sull'IA, ma anche la raccolta e sistematizzazione di tutti gli atti intervenuti sul cd. Diritto di Internet e che saranno applicabili anche ai software di IA. Proprio al fine di delineare un quadro certo di regole il Sottosegretario Alessio Butti ha firmato il decreto di nomina dei 13 esperti che compongono il **Comitato di Coordinamento**, incaricato di analizzare l'impatto dell'IA sul tessuto socioeconomico del paese ed elaborare la nuova strategia italiana per l'IA. L'obiettivo sarà quello di guidare le imprese e la pubblica amministrazione nello sviluppo e impiego delle nuove tecnologie con l'unico obiettivo di massimizzare i benefici e ridurne i rischi[18].

[16] Cit. Dipartimento per la Trasformazione Digitale, G7 Industry, Annex 1 Joint Statement on Cable Connectivity for Secure and Resilient Digital Communications Networks contenuto nel G7 Industry, Technology And Digital Ministerial Meeting, Verona and Trento 14-15 March 2024, https://innovazione.gov.it/notizie/articoli/en/g7-ministerial-declaration-on-industry-technology-and-digital/ ;

[17] Ai sensi della Direttiva 2015/1535, che prevede una procedura d'informazione nel settore delle regolamentazioni tecniche e delle regole relative ai servizi della società dell'informazione, si intende per società dell'informazione qualsiasi servizio prestato normalmente dietro retribuzione, a distanza, per via elettronica e a richiesta individuale di un destinatario di servizi

[18] AGID, Intelligenza Artificiale: presentato il Comitato di Coordinamento per l'aggiornamento della strategia nazionale, 9 novembre 2023, in https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2023/11/09/intelligenza-artificiale-presentato-il-comitato-coordinamento-laggiornamento ;

Il prossimo passo si concentrerà sulla nomina dell'Authority, prevista dall'AI Act come centro di coordinamento degli organismi di valutazione che si occuperanno di attestare la sicurezza dei sistemi di IA, prima dell'ingresso nel mercato europeo. L'autorità avrà il compito di coordinare le linee guida e le buone pratiche che emergeranno in relazione all'utilizzo dei software di IA in ciascun settore, come ad esempio quello dell'istruzione o della sanità. Il coordinamento in un settore così vasto sarà indispensabile per evitare che la normativa disordinata e contraddittoria diventi un ostacolo amministrativo, piuttosto che un fattore di semplificazione per tutti gli operatori del settore tech. Ad oggi sembra che la scelta dell'Italia ricadrà sull'Agenzia per l'Italia Digitale, alla quale verrà affidato il compito di coordinare ed attuare la regolazione sull'IA. Altri paesi, invece, hanno preferito investire il proprio Garante della Privacy, la cui competenza sul trattamento dati ben si sovrappone alle esigenze dell'Intelligenza Artificiale che nasce e cresce attraverso i dati[19].

Il risk based approach in Europa e nel mondo

Il Regolamento EU sull'Intelligenza Artificiale (AI Act), approvato dal Parlamento europeo il 13 marzo 2024, costituisce un primo esempio di bilanciamento tra i rischi e i benefici implicati dai nuovi sistemi di IA. Lo scopo del Regolamento, infatti, è quello di promuovere lo sviluppo e il commercio di tecnologie intelligenti nel mercato unico, ed al contempo assicurare ai cittadini dell'UE software affidabili e sicuri ed un alto livello di protezione dei diritti fondamentali[20]. La strategia normativa prescelta ripropone i principi di accountability e privacy by design and by default inaugurati nel 2016 con l'emanazione del GDPR[21].

In particolare, il concetto di privacy by design e by default impone ai fornitori di sistemi di IA:

- l'identificazione dei rischi ragionevolmente prevedibili derivanti dall'utilizzo del sistema di IA nel contesto di riferimento e in base alle istruzioni impartire dal produttore;
- la predisposizione di misure tecniche e organizzative finalizzate ad eliminare, ridurre la probabilità che il rischio si verifichi;
- la predisposizione di procedure o prassi che consentano di ridurre i danni causati dall'evento lesivo.

Invece il principio di accountability richiede che il soggetto si assuma la responsabilità del proprio operato e che ne dia conto, dimostrando di aver diligentemente rispettato (e come) lo schema di compliance by design and by default sopra descritto. Questo approccio ha dimostrato di essere l'unico effettivamente praticabile per

[19] E. Raffiotta, Quale autorità governerà l'intelligenza artificiale?, in Il Sole 24 Ore, 27 marzo 2023, https://www.ilsole24ore.com/art/quale-autorita-governera-l-intelligenza-artificiale-AEVKPYAD ;

[20] Considerando (1) del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione";

[21] Commissione UE, What does data protection 'by design' and 'by default' mean?, in https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en ;

governare l'infinita varietà di software, che non permette la costruzione di una fattispecie unica e onnicomprensiva. Inoltre, la definizione di fattispecie troppo rigide richiederebbe continui e costosi aggiornamenti per includere nel testo di legge i nuovi ritrovati della tecnica. Queste considerazioni devono peraltro coordinarsi con il principio di libera circolazione delle merci, che permette restrizioni al libero mercato solo se giustificate da motivi imperativi di interesse generale e proporzionate al livello di protezione perseguito.

Sulla base di questi assunti le istituzioni europee hanno proposto un sistema di governance fondato sul risk based approach, costituto idealmente da uno schema piramidale dove il valore più alto è rappresentato dai sistemi che implicano pratiche di IA vietate (art.5), seguiti dai sistemi ad alto rischio (art. 6) e dai General-Purpose AI models con rischio sistemico (art. 51), i GPAI che non presentano rischio sistemico (art. 53) ed infine i sistemi con rischio trascurabile (art. 50). I sistemi ad alto rischio si dividono tra i sistemi per i quali sono già stati imposti standard di sicurezza da precedenti legislazioni (richiamate nell'Allegato II) e per i quali era già richiesta un'attestazione di conformità proveniente da una parte terza, ed infine quelli che ricadono nelle particolari aree di applicazione dell'Allegato III. I modelli di GPAI (modelli di IA invece costituiscono la cd. Intelligenza Artificiale generativa ossia quella «che sia caratterizzato una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle»[22]. Anche in questa seconda categoria troviamo una classificazione in base al rischio, infatti in presenza di rischio sistematico hanno imposto maggiori obblighi[23].

La politica normativa orientata sul rischio è stata peraltro avallata anche dagli Stati Uniti, che nell'Ordine Esecutivo del Presidente degli Stati Uniti per uno sviluppo di un'Intelligenza Artificiale sicura e affidabile si sofferma proprio sul rischio quando afferma che per sfruttare al meglio l'intelligenza artificiale, e per trarne tutti i vantaggi possibili, è necessario mitigarne i rischi più gravi. Peraltro, proprio pochi mesi prima il Presidente degli Stati Uniti aveva firmato un ordine esecutivo che per la prima volta vieta al Governo statunitense di utilizzare spyware commerciali che «pongano rischi per la sicurezza nazionale o sono stati utilizzati in modo improprio da operatori stranieri per perpetrare violazioni dei diritti umani in tutto il mondo»[24]. Si osserva pertanto un allineamento tra le prospettive degli Stati Uniti e dell'Europa sulla tutela dei

[22] Cit. Art. 3 del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione" ;

[23] Art. 55 del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione". «In aggiunta agli obblighi di cui all'articolo 53, i fornitori di modelli di IA per finalità generali con rischio sistemico: a) effettuano una valutazione dei modelli in conformità di protocolli e strumenti standardizzati che rispecchino lo stato dell'arte, anche svolgendo e documentando il test contraddittorio (adversarial testing) del modello al fine di individuare e attenuare il rischio sistemico; b) valutano e attenuano i possibili rischi sistemici a livello dell'Unione, comprese le loro fonti, che possono derivare dallo sviluppo, dall'immissione sul mercato o dall'uso di modelli di IA per finalità generali con rischio sistemico c) tengono traccia, documentano e riferiscono senza indebito ritardo all'ufficio per l'IA e, se del caso, alle autorità nazionali competenti, le informazioni pertinenti su incidenti gravi ed eventuali misure correttive per porvi rimedio; d) garantiscono un livello adeguato di protezione della cibersicurezza per quanto riguarda il modello di IA per finalità generali con rischio sistemico e l'infrastruttura fisica del modello.;

[24] Cit. The White House, FACT SHEET: President Biden Signs Executive Order to Prohibit U.S. Government Use of Commercial Spyware that Poses Risks to National Security, 27 marzo 2023, in https://www.whitehouse.gov/briefing-room/statements-releases/2023/03/27/fact-sheet-president-biden-signs-executive-order-to-prohibit-u-s-government-use-of-commercial-spyware-that-poses-risks-to-national-security/ ; The White House, Executive Order on Prohibition on Use by the United States Government of Commercial Spyware that Poses Risks to

diritti e della dignità dei cittadini, che potrebbe suggerire un riavvicinamento tra i due continenti dopo la rottura del Privacy Shield[25].

I passi avanti degli Stati Uniti sulla protezione della privacy potrebbe in effetti suggerire una soluzione di policy condivisa tra i due continenti. Infatti, solo una risposta congiunta che coinvolga più giurisdizioni potrebbe garantire l'effettivo raggiungimento degli obiettivi auspicati dall'Hiroshima Process, il quale attualmente sollecita «tutti gli operatori AI e le organizzazioni che sviluppano sistemi di intelligenza artificiale avanzati» ad applicare i Principi Guida Internazionali. Una soluzione di policy non condivisa potrebbe risultare non efficace a fronte della facile permeabilità di internet, che costituisce ancora la principale fonte di dati usata per allenare i sistemi di IA, e la possibilità che i sistemi di IA comunichino tra loro diffondendo conclusioni biased. Pertanto, anche se un sistema di IA corrotto non venisse offerto sul mercato europeo, nessuno potrebbe impedirgli di attingere e trattare i dati degli utenti europei con tutti i rischi che potrebbero derivare da definizioni e conclusioni erronee e discriminatorie sulla popolazione e la struttura economica e sociale europea.

Solo un sistema di policy condiviso permetterebbe la costruzione di un mercato tecnologico fluido e competitivo. Di converso se venissero imposti regole e standard specifici per ogni nazione o continente verrebbero di conseguenza aumentati i costi amministrativi delle industrie tech per conformarsi ad ogni quadro normativo. Inoltre, un assetto di linee guida precise e condivise consentirebbe ai sistemi di IA di circolare oltreoceano attingendo ad un patrimonio informativo più ampio, dati di prima mano su usi, costumi e caratteristiche della popolazione mondiale, evitando in questo modo lo sviluppo di IA con un panorama di conoscenza limitato che presenterebbe un maggior rischio di bias. Bisogna infatti iniziare a pensare i sistemi di IA non come strumenti finiti e perfezionati, anche dopo essere messi sul mercato essi raccolgono dati e li classificano secondo i parametri imposti dal produttore, quelli che gli verranno impartiti tramite gli aggiornamenti ed infine quelli autonomamente.

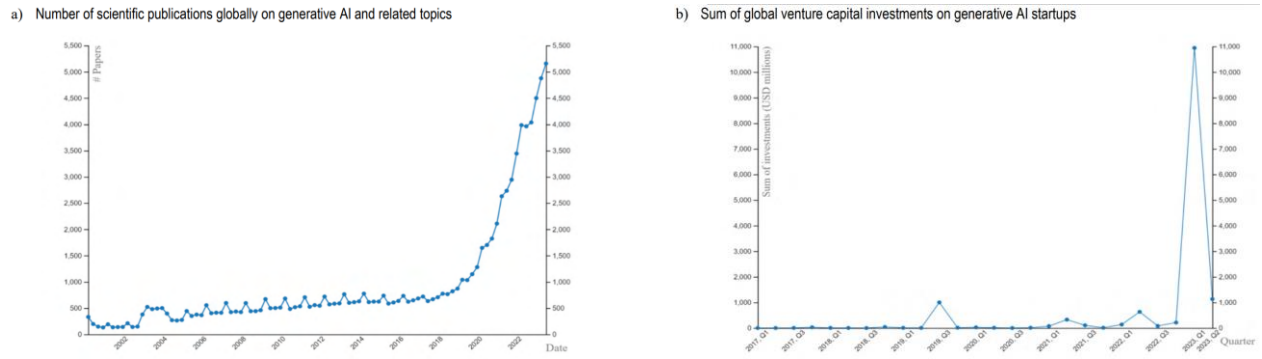
La minaccia dell'Intelligenza Artificiale Generativa

L'intelligenza Artificiale generativa costituisce l'esempio lampante dell'alto livello di obsolescenza delle norme che regolano le nuove tecnologie, essa infatti non era contemplata nella prima versione dell'AI Act, oggi invece costituisce il principale oggetto di attenzione di giudici, legislatori e organizzazioni internazionali. L'IA generativa si distingue dall'IA "tradizionale" sotto diversi profili: viene allenata su vari tipi di input per generare una risposta linguistica, procede quindi in modo inverso rispetto al classico apprendimento tramite feedback

National Security, 27 marzo 2023, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/03/27/executive-order-on-prohibition-on-use-by-the-united-states-government-of-commercial-spyware-that-poses-risks-to-national-security/ ;

[25] Per maggiori osservazioni sull'Executive Order ed i rapporti USA-Europa sul trattamento dati dei cittadini si rinvia a Guido Scorza, Scorza: "Executive order di Biden, ecco i punti critici, quelli positivi e quelli da chiarire", 10 ottobre 2022, in Agenda Digitale, https://www.agendadigitale.eu/sicurezza/privacy/scorza-executive-order-di-biden-ecco-i-punti-critici-quelli-positivi-e-quelli-da-chiarire/, l'articolo commenta il FACT SHEET: President Biden Signs Executive Order to Implement the European Union-U.S. Data Privacy Framework ;

su cui si basano i metodi di regressione e classificazione[26]. Questi sistemi dotati di machine learning system sono addestrati su un enorme quantità di dati, e si caratterizzano per la capacità di generare nuovi contenuti digitali sotto forma di testo, immagini, audio, video o codice software. Insomma, oggi l'IA generativa è il sistema che più si avvicina all'Intelligenza Artificiale forte (o generale), ossia quel sistema in grado di replicare perfettamente le capacità cognitive e creative umane. Queste caratteristiche hanno reso da subito l'IA generativa uno dei prodotti più attrattivi del 2023 per la ricerca scientifica e per gli investimenti privati, nel 2022 infatti si osserva un aumento dei modelli di open source dedicati allo sviluppo dell'IA generativa.



Fonte: G7 Hiroshima Process on Generative Artificial Intelligence (AI) towards a G7 common understanding on Generative AI.

Una macchina in grado di replicare il pensiero umano rivoluzionerebbe il sistema socio-economico che conosciamo, per questo motivo è necessario monitorare il fenomeno in modo che non costituisca una disruptive innovation, ma solo un ulteriore avanzamento della tecnica al servizio della società. Le potenzialità applicative ed i rischi connessi all'IA generativa hanno perciò imposto a tutti policy maker lo studio e la gestione del fenomeno. Pertanto, il legislatore europeo ha inserito all'interno dell'AI Act la categoria dei "'general purpose AI models" (GPAI). Il regime imposto dall'AI Act per i fornitori di modelli di IA generativa a fini generali stabilisce una serie di obblighi specifici volti a garantire trasparenza, tracciabilità e sicurezza nell'utilizzo di tali tecnologie. In primo luogo, i fornitori sono tenuti a redigere e mantenere aggiornata la documentazione tecnica del modello, che descriva dettagliatamente il processo di addestramento, i test di sicurezza effettuati, nonché i risultati della valutazione. I fornitori sono anche tenuti a mettere a disposizione informazioni e documentazione del sistema di IA generativa all'acquirente che intenda integrare questi modelli nei propri sistemi. La documentazione fornita deve consentire ai fornitori di sistemi di IA di comprendere a fondo le capacità e i limiti del modello, supportandoli nell'adempiere ai propri obblighi di conformità rispetto al regolamento; il tutto nel rispetto dei diritti di proprietà intellettuale, delle informazioni riservate e dei segreti commerciali. L'AI Act prevede anche la categoria di GPAI con rischio sistemico a livello

[26] Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity (January 14, 2024). Disponibile su SSRN: https://ssrn.com/abstract=4694565 or http://dx.doi.org/10.2139/ssrn.4694565 ;

comunitario, sono tali i sistemi che hanno un impatto significativo sul mercato dell'Unione a causa della loro portata o degli effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore[27]. Il rischio è considerato sistemico qualora presenti anche l'attitudine ad estendersi su larga scala e coinvolgere l'insieme delle attività e processi aziendali necessari per creare e vendere un prodotto o un servizio, quindi coinvolgendo tutte le diverse fasi di produzione. Per questi sistemi di IA sono richieste specifiche valutazioni del rischio sistemico e misure per mitigare il rischio individuato, è inoltre consigliato di affidarsi a codici di condotta che verranno elaborati dalla tecnica e dalle prassi industriali.

Anche gli Stati Uniti sono intervenuti sul fenomeno annunciando il 21 luglio 2023 che otto compagnie - Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, e Stability - hanno volontariamente aderito ad un accordo di autoregolamentazione per promuovere uno sviluppo sicuro, protetto e trasparente dei modelli di intelligenza artificiale generativa. In particolare, le aziende si sono impegnate a condividere le informazioni con il governo sui rischi per la sicurezza nazionale, la loro evoluzione e le eventuali violazioni alle misure di sicurezza; ad investire nella sicurezza informatica e nelle misure a salvaguardia dei diritti di copyright; sviluppare meccanismi che consentano al pubblico di riconoscere la provenienza artificiale dei contenuti attraverso un watermark. Inoltre, l'accordo prevede che vengano pubblicati report sulle capacità, i limiti, gli impieghi inappropriati ed i rischi sociali, come danni o discriminazioni; infine, si impegnano a promuovere la ricerca sui rischi sociali connessi ai GPAI ed educare l'IA generativa nella soluzione delle grandi sfide dell'umanità[28].

Entrambi gli approcci sono diretti a stimolare lo sviluppo dei GPAI in forza delle opportunità connesse a questa tecnologia, e degli importanti risultati che potrebbe raggiungere nei campi della sanità, transizione verde, educazione, previdenza, industria e commercio. Allo stesso tempo però l'impiego incontrollato di queste tecnologie potrebbe portare a delle derive difficili da arrestare perché imprevedibili ex ante sia nell'***an*** che nel ***quantum***. Uno dei settori maggiormente sensibili è quello della comunicazione, ci sono forti timori legati alla disinformazione aggravata dai deep fake, e soprattutto alla manipolazione dell'opinione pubblica; si aggiungono inoltre le preoccupazioni per la protezione dei dati personali e la compatibilità della tecnologia con i principi del GDPR (minimizzazione del trattamento, trasparenza, circoscritta finalità e durata del trattamento). Il Garante per la protezione dei dati personali italiano aveva infatti censurato Chat-GPT a causa dell'insufficiente sistema adottato per accertare l'età degli utenti, che non consentiva di tutelare a dovere i dati dei minori[29].

[27] Cit. num 64 dell'Articolo 3 del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione".;
[28] The White House, FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI, in FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI | The White House ;
[29] Novelli, Claudio and Casolari, Federico and Hacker, Philipp and Spedicato, Giorgio and Floridi, Luciano, op cit. p. 15;

Alla luce delle particolari caratteristiche che rendono l'IA generativa più pericolosa e più potente dei sistemi di IA tradizionali i leader del G7 riconoscono il bisogno di sottoporla a regolazione e controllo attraverso un quadro di principi internazionali uniformi e una maggiore interazione tra le differenti giurisdizioni. Quindi il primo passo si concentrerà sulla definizione di principi etici relativi all'utilizzo dell'IA generativa in generale e principi specifici per ciascuna area di implementazione (es. militare, sanitaria, ambientale ecc...). Questi principi possono indurre le compagnie tech a sviluppare IA votate al bene comune, inteso quale arricchimento dell'intera comunità, inclusi i paesi sin ora esclusi dalla corsa alle nuove tecnologie. Asia e Africa non possono essere tagliate fuori dalla politica di incentivi e promozione dell'IA in primis perché potrebbe aggravare il divario digitale esistente, ed in secundis, un'IA generativa che non comprenda anche i dati dell'economia e della società di questi continenti potrebbe determinare ulteriori discriminazioni e bias, così aggravando la polarizzazione mondiale tra occidente e oriente[30].

Strumenti per un'AI affidabile: Technical, Procedural ed Educational approaches

La regolazione di un nuovo fenomeno richiede prima di tutto un nucleo di principi condivisi e stabiliti in relazione agli obiettivi da raggiungere e beni da tutelare. Il 30 ottobre i leader del G7 hanno sottoscritto i nuovi principi guida internazionali sull'intelligenza artificiale ed un codice di condotta volontario per lo sviluppo di un IA sicura e affidabile, con l'auspicio che vengano implementati e rispettati da enti pubblici e privati. Il documento, infatti, espressamente invita i «principali stakeholder del settore a seguire tali azioni, in linea con un **approccio basato sul rischio**, mentre i governi svilupperanno modelli regolatori e di governance più duraturi e dettagliati»[31]. La lista – non esaustiva – di azioni deve essere attuata non solo a seguito della immissione nel mercato dei sistemi ma sin dal loro sviluppo e per tutto il ciclo di vita del sistema:

1) la predisposizione di misure che riducano il rischio di bias e danni alla collettività;
2) Identificare e ridurre le vulnerabilità del sistema e, ove appropriato, gli incidenti e i modelli di abuso;
3) Pubblicare tramite report le capacità, i limiti, i settori di applicazione dei sistemi di IA e la destinazione d'uso in modo da assicurare una trasparenza sufficiente incrementando l'accountability;
4) Condivisione tra istituzioni e privati delle informazioni relative agli incidenti causati da IA;
5) Sviluppare, implementare e divulgare politiche di governance dell'IA e di gestione del rischio, dirette a ridurre il rischio di discriminazioni e bias e tutelare la privacy;
6) Investire e implementare controlli di sicurezza, per tutelare la sicurezza informatica e fisica, nonché tutele contro le minacce interne;

[30] OECD, G7 Hiroshima Process on Generative Artificial Intelligence (Ai) towards A G7 common understanding on Generative AI, 7 settembre 2023, https://www.oecd-ilibrary.org/science-and-technology/g7-hiroshima-process-on-generative-artificial-intelligence-ai_bf3c0c60-en ;
[31] Minister of Foreign Affairs of Japan, Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI system, https://www.mofa.go.jp/files/100573471.pdf;

7) Sviluppare e implementare meccanismi affidabili di autenticazione e di definizione della provenienza dei contenuti, ove tecnicamente fattibile, come watermarking o altre tecniche per consentire agli utenti per identificare i contenuti generati dall'intelligenza artificiale;
8) Investire sulla ricerca di misure dirette a minimizzare i rischi individuali e sociali;
9) Dare priorità allo sviluppo di sistemi di intelligenza artificiale per rispondere alle sfide globali, quali ma non solo, la crisi climatica, la salute e l'istruzione dei paesi sottosviluppati;
10) Promuovere l'adozione di standard tecnici internazionali;
11) Implementare misure che garantiscano set di dati (input) di qualità, tutele per i dati personali e la proprietà intellettuale.

In effetti i principi appena elencati possono trovare già ampia condivisione, in quanto coincidono con le azioni previste dalle principali strategie sull'IA adottate da governi, centri di ricerca e organizzazioni internazionali per contrastare i potenziali effetti pregiudizievoli della Black Box. Infatti, proprio l'autonomia e predittività che fa dell'IA una tecnologia rivoluzionaria costituiscono i suoi principali elementi di rischio. Infatti, i sistemi di IA al termine di ogni processo automatizzato non rappresentano i parametri ed i dati presi in considerazione per il responso, pertanto non è possibile indagare la ragionevolezza dei risultati suggeriti dall'IA. In questo contesto servono linee guida per orientare gli stakeholders verso un'IA più intellegibile e giusta.
La corsa alla regolazione dell'IA vede, per ora, in testa l'Unione Europea che già nell'aprile 2019 varava gli "Orientamenti etici per una intelligenza artificiale affidabile"[32], mentre oggi si appresta ad emanare il primo atto normativo al mondo sull'intelligenza artificiale. L'AI Act propone, per l'appunto, un approccio normativo basato sul rischio inteso quale « rischio significativo di danno per la salute, la sicurezza o i diritti fondamentali delle persone fisiche, anche nel senso di non influenzare materialmente il risultato del processo decisionale»[33]. Allo stesso modo il Blueprint for an AI Bill of Rights pubblicato dalla Casa Bianca esprime timori per gli effetti negativi che gli automated systems hanno sui diritti civili e libertà democratiche; infatti esordisce asserendo che «i sistemi dovrebbero essere sottoposti a test di pre-implementazione, identificazione e mitigazione del rischio, e ad un monitoraggio continuo che dimostri la sicurezza e l'efficacia in base all'uso previsto, riducendo gli outcome non sicuri, inclusi quelli che fuoriescono dalla destinazione d'uso del sistema e dagli standard previsti»[34]. Questa linea politica è stata recentemente implementata con l'"Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence"[35] precedentemente citato.

[32] Commissione EU, Orientamenti etici per un'IA affidabile, in https://digital-strategy.ec.europa.eu/it/library/ethics-guidelines-trustworthy-ai ;
[33] Cit. Art. 6 del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione";
[34] Cit. Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ ;
[35] The White House, Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ ;

Dalla lettura comparativa dei due atti normativi possiamo delineare un nucleo di obiettivi condivisi sui quali predisporre un comune progetto di policy diretto a **tutelare la privacy** dei cittadini e ad incentivare un **responsabile utilizzo** dei sistemi di IA nella **sanità** e nell'**istruzione**. Particolare attenzione dovrà essere prestata al **mercato del lavoro** che è già stato rivoluzionato dalla tecnica introducendo l'utilizzo dell'IA nelle procedure di reclutamento e selezione del personale, nella gestione dei dipendenti, demandandogli scelte su promozioni o licenziamenti, valutazione del comportamento individuale e del rendimento. Questi sistemi sono infatti stati classificati quali sistemi ad alto rischio a causa dell'impatto che le valutazioni automatizzate avranno sul futuro professionale e prospettive economiche del lavoratore[36]. Infine, si promuove l'utilizzo di sistemi di IA nell'amministrazione della giustizia e nei processi democratici in nome dell'efficienza, al contempo classificandoli quali sistemi ad alto rischio a causa dei gravi effetti che potrebbero avere sulla democrazia, stato di diritto e sulle libertà individuali[37]. Alla luce di questi rilievi l'OCSE suggerisce strumenti di natura tecnica, procedurale o educativi garantire un'IA sicura, affidabile e accurata:

- **strumenti tecnici** possono consistere in toolkits, software tools, documentazione tecnica o certificazioni attestanti la sicurezza del prodotto e la conformità alle norme. Alcune aziende tech hanno infatti già sviluppato software in grado di rintracciare il pericolo di bias discriminatori, strumenti open-source che curano l'equità e la correttezza dei risultati automatizzati. Così come sono stati sviluppati test per misurare il grado di trasparenza dell'IA, come software che aiutino ad individuare i dati su cui i sistemi di IA abbiano preso in considerazione in un determinato processo di funzionamento e algoritmi che aiutino a comprendere gli schemi di predizione dell'IA[38]:
- **strumenti procedurali** che guidino lo sviluppo dei sistemi e la produzione degli automi e che possono consistere in linee guida, modelli di governance, codici di condotta, standard per programmatori ed ingegneri, programmi di certificazione;
- **strumenti educativi** invece comprendono meccanismi per informare tutti gli operatori del settore sui rischi e sulle migliori misure per ridurli. Gli stakeholder di miglioramento delle competenze coinvolti o interessati dall'implementazione di un sistema di IA, esempi di questi strumenti sono stati dei videogame che istruivano gli utenti sui rischi implicati dall'IA, lezioni e seminari per istruire i lavoratori e managers aziendali ai rischi connessi all'IA che utilizzano, corsi on line ed esercizi per imparare quale impatto può arrecare un determinato sistema di IA.

[36] Considerando 57, del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione";
[37] Considerando 61 del testo "Emendamenti Del Parlamento Europeo alla proposta della Commissione";
[38] OECD, Tools for trustworthy ai a framework to compare implementation tools for trustworthy ai systems OECD digital economy papers, June 2021 No. 312;

Infine, si consiglia la tenuta di un database aggiornato che raccolga tutti gli strumenti utilizzati dalle imprese per rendere i loro sistemi di IA più sicuri, accurati e trasparenti, in modo da offrire un panorama dello stato della tecnica a tutti gli operatori di mercato, nonché stimolare un circolo virtuoso di concorrenza e competizione tra gli stakeholders nel dotarsi di misure e tools sempre più avanzati.

Spazi di sperimentazione normativa

I progressi della tecnica richiedono spesso adeguamenti normativi per regolare i nuovi prodotti, servizi e modelli di business resi possibili proprio dalle nuove tecnologie; tuttavia, gli atti normativi tardano ad arrivare o richiedono lunghi lavori parlamentari a causa delle poche informazioni sulle funzionalità tecniche dei nuovi strumenti. Infatti, la cd. "better regulation" richiede l'approfondimento di vari aspetti del fenomeno da regolare: una definizione del problema, un'effettiva cognizione del rapporto di causalità, una proposta normativa ragionevole e proporzionale agli obiettivi di tutela previsti, l'analisi d'impatto della norma sull'ambiente e valutare i benefici apportati dalla nuova soluzione di policy[39].

Le nuove tecnologie, e soprattutto il livello di automatizzazione raggiunta, hanno reso più complessa la rilevazione dei nessi causali e dei danni correlati ai nuovi prodotti. La black box ha, di fatto, oscurato il nesso logico che intercorre tra l'input dato al software e l'output da questo generato, mentre i sistemi di apprendimento automatico permettono all'IA di imparare dall'esperienza e di generare nuovi e originali contenuti; per questi motivi, l'operato dei sistemi di IA non è direttamente e strettamente riconducibile alle direttive del programmatore. Inoltre, anche la percezione del danno causato dalla macchina è molto attenuata. Infatti, eccezion fatta per i danni arrecati a persone o cose, la società non è allenata a distinguere le discriminazioni generate dai bias del software, soprattutto perché si associa alla macchina l'idea di neutralità e correttezza tipica di una calcolatrice senza considerare che gli elementi su cui ragiona l'IA non sempre aderiscono ad una legge matematica, ma spesso sono valutati con le medesime logiche umane. A ciò si aggiunge la trasversalità dell'intelligenza artificiale, applicata in molteplici settori (assicurativo, finanziario, sanitario, immobiliare, fiscale), richiederà di conseguenza la definizione di specifici requisiti e misure di sicurezza previsti per ciascuna destinazione d'uso dei software[40].

Quindi, l'attuale contesto vede da un lato il mondo dell'industria, del commercio e dei servizi fare largo uso di sistemi di IA, mettendo in serio rischio i diritti degli individui privi di una normativa a tutelarli, dall'altro un mercato tech globale e competitivo dal quale l'Europa non potrebbe semplicemente uscire con un semplice divieto di implementazione dei software, nell'attesa di una legge che li normi. Pertanto, sono state elaborate

[39] Commissione UE, Better Regulation July 2023, https://commission.europa.eu/law/law-making-process/planning-and-proposing-law/better-regulation/better-regulation-guidelines-and-toolbox_en ;
[40] OECD, Regulatory Sandboxes in Artificial Intelligence, OECD Digital Economy Papers July 2023 No. 356;

forme di sperimentazione normativa che consentono di coniugare la necessità di fattispecie chiare con un approccio empirico che permetta di provare la regolazione sul campo. Infatti, molto spesso l'innovazione appare disruptive perché entra in collisione con il quadro normativo di riferimento, pertanto sono elaborati strumenti di indagine come le regulatory sandboxes, che permettono agli operatori di testare il loro servizio o prodotto tech in un ambiente controllato e circoscritto. Infatti, le regulatory sandboxes consistono in regimi giuridici sperimentali che derogano o modificano temporaneamente la normativa nazionale a vantaggio di uno o più operatori. Al termine dell'esperimento gli operatori responsabili predispongono un progetto di policy per la propria tecnologia innovativa basato sulle rilevazioni ottenute durante l'esperimento normativo; in questo modo avranno la possibilità di influenzare il dibattito utilizzando i dati e ricerche ottenuti tramite la sandbox.
La necessità di strumenti normativi agevoli e flessibili è stata recepita dal legislatore europeo che ha previsto al titolo V dell'AI Act spazi di sperimentazione normativa in ogni stato membro, che dovranno essere istituiti e organizzati dall'autorità amministrativa competente con lo scopo di raggiungere i seguenti obiettivi:

- incrementare la certezza del diritto assicurando la conformità della tecnologia con le disposizioni europee e il quadro normativo dello stato membro;
- supportare la condivisione delle best practice attraverso un maggior coordinamento con le autorità;
- promuovere l'innovazione, la competizione e lo sviluppo di un ecosistema per l'IA;
- contribuire ad un processo legislativo basato sulle rilevazioni empiriche;
- facilitare ed accelerare l'accesso dei sistemi di IA nel mercato europeo[41].

Per evitare un'applicazione frammentata dell'istituto, la Commissione europea interverrà con atti separati per regolare in dettaglio l'istituzione, implementazione, funzionamento e supervisione dell'AI sandbox. In particolare, dovrà stabilire i requisiti e la procedura amministrativa per accedere allo strumento; tali informazioni dovranno essere fornite in forma chiara e agevole tramite un'interfaccia web dedicata, in modo da coinvolgere SME e start-up che hanno minori risorse legali e amministrative ridotte rispetto alle Big Tech[42].
Attualmente in Italia vediamo due sandbox attive, la prima è stata istituita tramite il decreto-legge n. 34/2019 e destinata al settore finanziario, bancario e assicurativo, opera sotto la supervisione della Banca d'Italia, IVASS e Consob[43]. Per accedere a questa Fintech Sandbox italiana gli operatori dei settori elencati devono presentare un'attività "significativamente innovativa", che apporti valore aggiunto ai consumatori o una maggiore efficienza al sistema economico-finanziario, inoltre deve essere economicamente sostenibile ed aver

[41] Cit. Art.57 ;
[42] Art. 58 ;
[43] Maggiori informazioni disponibili al sito https://www.bancaditalia.it/focus/sandbox/index.html?com.dotmarketing.htmlpage.language=1&dotcache=refresh ;

raggiunto uno sviluppo "sufficientemente avanzato per la sperimentazione"[44]. La seconda regulatory sandbox, dal nome **Sperimentazione Italia** è invece una sandbox di carattere generale, destinata alle innovazioni di qualsiasi settore, è gestita dal Dipartimento per la Trasformazione digitale e dal MIMIT. Questa rappresenta la prima sandbox a scopo generale in Europa; in tal senso, potrebbe rappresentare una porta di ingresso importante per le start-up e le imprese che vogliano entrare nel mercato europeo e per la sua regolamentazione. Tuttavia, le sandboxes avviate tramite Sperimentazione Italia sino ad ora ammontano a due, l'irrisoria adesione all'istituto potrebbe essere addebitato innanzitutto alla carente divulgazione pubblica dell'esistenza e delle caratteristiche della sandbox. Infatti, sulla pagina web di Sperimentazione Italia troviamo indicazioni precise sulle fasi e i tempi della procedura amministrativa, ma non sui requisiti di partecipazione, criteri di valutazione e, soprattutto, sui precedenti progetti avviati[45]. Questo difetto informativo scoraggia gli operatori di mercato dall'intraprendere una procedura burocratica dagli esiti troppo incerti.

In conclusione, le regulatory sandboxes ricopriranno un ruolo centrale nella regolazione delle nuove tecnologie, in quanto permette un dialogo ed un reciproco scambio di informazioni tra la pubblica amministrazione ed i privati sulle tecnologie impiegate dall'impresa, apportando benefici ad entrambi. Le autorità non dovranno attendere che le informazioni sui rischi e sulle caratteristiche del sistema innovativo siano esibite in giudizio da un cittadino leso, ma potranno conoscerle in anticipo collaborando con il privato che intenda investire sulla sua applicazione in territorio europeo. Pertanto, sarà necessario perfezionare Sperimentazione Italia fornendo maggiori informazioni, linee guida sulla modalità di selezione per ciascuna destinazione d'uso o area di applicazione ed in molteplici lingue, per attrarre investimenti internazionali. Inoltre, al fine di avvicinare l'amministrazione agli imprenditori interessati potrebbero essere forniti i contatti degli uffici competenti per ciascun settore economico, offrendo la possibilità di avviare un dialogo proficuo anche prima dell'invio della domanda.

Verso un'Intelligenza Artificiale Affidabile e Responsabile

Una volta stabilito un elenco di principi condivisi per lo sviluppo di un'IA sicura ed affidabile è necessario predisporre un sistema che ne assicuri la concreta applicazione. In effetti, lo stesso principio 1.5 previsto dall'OCSE per un'IA sicura ed affidabile richiede che «AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of art»[46]. Pertanto, richiama tutti gli attori del ciclo di produzione e sviluppo dell'IA a rispondere del rispetto dei requisiti e principi di riduzione del rischio previsti dall'IA Act. Infatti, un buon

[44] Cit. Ministero dell'Economia E Delle Finanze, Decreto 30 aprile 2021, n. 100, art. 6 ; Cfr. Sofia Ranchordas, Valeria Vinci, Regulatory sandboxes and innovation-friendly regulation : between collaboration and capture, in https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4696442 ;
[45] Cfr, Sofia Ranchordas, Valeria Vinci cit. p. 14;
[46] OECD.AI Policy Advisory, Accountability (Principle 1.5), https://oecd.ai/en/dashboards/ai-principles/P9 ;

sistema di governance deve poter contare su una forte accountability, intesa quale capacità del soggetto di rendere conto del proprio operato e della sua conformità alle norme. L'accountability, intesa quale attitudine a rispondere delle proprie azioni, necessita di tre fondamenti: un'autorità riconosciuta, attività di indagine ed esercizio del potere [47]. Il sistema descritto sembra corrispondere al canonico rapporto tra l'autorità amministrativa di settore e i soggetti sottoposti al suo controllo, tuttavia possiamo distinguere diverse fasi di accountability in base allo scopo perseguito dall'attività di policy.

Possiamo infatti distinguere le fasi di: compliance, relazione, controllo e applicazione. La prima si concretizza nella redazione di linee guida, codici di condotta e standard tecnici che aiutino gli operatori a disegnare un'IA conforme alle disposizioni di legge. La seconda riguarda il momento dialogico tra la singola impresa e l'autorità, nel quale l'operatore chiarisce come sono stati soddisfatti i requisiti legali. Infine, vi sarà la fase di indagine e comminazione della sanzione qualora si rilevi una difformità nel software rispetto ai requisiti ed adempimenti richiesti dall'AI Act e dalle disposizioni di settore applicabili al singolo device[48].Un sistema di governance completo prevede strumenti appartenenti a tutte queste fasi, contemplando forme proattive e reattive di accountability; in particolare, forme proattive sono previste nelle fasi di compliance e controllo mentre la richiesta di relazioni e la comminazione di sanzioni costituisce la fase reattiva dell'ordinamento.

L'AI Act mette in campo una proposta di governance che rispecchia i requisiti previsti; infatti, il sistema prevede l'istituzione di un'autorità di notifica in ogni stato membro, che dovrà selezionare, valutare e coordinare gli organismi di valutazione che si occuperanno, in prima istanza, di indagare e dichiarare la conformità al regolamento europeo del sistema di IA, attraverso un attestato di conformità. Gli organismi di valutazione dovranno soddisfare alcuni requisiti, quali:

- possedere le risorse tecniche professionali per valutare i sistemi di IA dei settori di cui si ritengono competenti;
- indipendenza, terzietà ed imparzialità;
- garantire la confidenzialità delle informazioni apprese durante lo svolgimento dell'incarico;
- devono possedere personale specializzato con esperienza e competenza relativa al sistema di IA da valutare e ai dati ed elaborazioni di dati effettuate del software[49];

Gli organismi di valutazione costituiranno gli intermediari specializzati che aiuteranno i fornitori tech ad accertare e provare che il proprio prodotto intelligente è sicuro, affidabile e conforme ai requisiti previsti dall'AI

[47] Novelli Claudio, Taddeo Mariarosaria, Floridi Luciano, Accountability in Artificial Intelligence: What It Is and How It Works (August 3, 2022). AI & Society: Journal of Knowledge, Culture and Communication - Springer https://doi.org/10.1007/s00146-023-01635-y , Available at SSRN: https://ssrn.com/abstract=4180366 or http://dx.doi.org/10.2139/ssrn.4180366 ;
[48] Novelli Claudio, Taddeo Mariarosaria, Floridi Luciano, Op. cit. p. 12.
[49] Capo 4 del

Act. I sistemi di IA che entreranno nel mercato europeo dovranno garantire un'informativa completa sul contesto e lo scopo al quale è destinato il sistema, sulla modalità con cui effettua le proprie analisi e inferenze, sul set di dati utilizzato per allenare il software; dovranno inoltre essere incluse informazioni sull'attività del fornitore, sui potenziali utilizzatori del sistema. Infatti, il sistema di governance elaborato dall'AI Act prevede che i produttori conducano una valutazione di impatto della singola IA sui diritti fondamentali che descriva gli specifici rischi legati all'utilizzo del prodotto o del software indicando le misure adottate per ridurre o eliminare il rischio rilevato. La valutazione di impatto dovrà essere sempre aggiornata durante tutto il ciclo di vita dell'automa; pertanto, se una delle informazioni inserite non dovesse più corrispondere al vero, a causa della capacità evolutive e di apprendimento del sistema, il fornitore dovrà intervenire e completare il resoconto. La valutazione di impatto costituirà quindi il rapporto con il quale i fornitori dimostreranno di aver sviluppato un sistema di IA equo, giusto e rettificabile, o per lo meno governabile, grazie al pronto intervento dell'essere umano. Non esiste quindi una risposta univoca e valida per tutti i software, ogni prodotto comporta un autonomo e singolare rischio per cui non è possibile imporre un insieme di misure obbligatorie per tutti; sarà, bensì, necessario che ciascun titolare dell'IA disegni un modello di azioni proporzionate e adeguate per tutelare i diritti fondamentali esposti. Dinnanzi le specificità delle nuove tecnologie la legge generale ed astratta deve essere integrata dalla tecnica e dalle sue invenzioni per elaborare un'IA più trasparente, accurata e gestibile. Tali strumenti, di tipo tecnico, procedurale o educativo, dovranno essere combinati caso per caso in base al contesto, diritti individuali e sociali coinvolti e probabilità che si verifichi il rischio. Infine, dovrà essere messo in campo un adeguato ed efficiente sistema di responsabilità civile, che supporti e ripaghi gli eventuali soggetti danneggiati dal funzionamento di un'IA, accompagnando il soggetto dalla scoperta sino alla prova del malfunzionamento della macchina.

Oltre all'impegno delle autorità nell'assicurare un IA affidabile e sicura, l'incontro a Verona e Trento avvenuto lo scorso marzo tra i Ministri dell'industria del G7 ha evidenziato l'importanza del coinvolgimento del settore privato, quale principale attore nell'analisi di standard e best practice. Infatti, al momento sono solo le Big Tech a detenere le ultime ed aggiornate informazioni sugli ultimi ritrovati della tecnica, sul loro funzionamento e data set; inoltre, sono sempre loro che dovranno in prima istanza analizzare l'impatto delle nuove tecnologie sull'ambiente. Soltanto una approfondita conoscenza del fenomeno garantisce un quadro regolatorio efficiente, pertanto il regime di responsabilità per i danni, che inevitabilmente verranno arrecati dai prodotti intelligenti o dai responsi automatizzati. Gli stakeholders del settore hanno infatti evidenziato la necessità di strutturare l'approccio regolatorio seguendo i diversi livelli ed elementi impiegati per l'implementazione e l'esecuzione dell'IA, quali ad esempio e la raccolta e la preparazione dei dati, l'apprendimento automatico (machine learning), la gestione delle infrastrutture di calcolo, l'implementazione di algoritmi specifici e così

via. Infatti, queste diverse fasi del processo di sviluppo e produzione della macchina intelligente[50] possono essere affrontate tutte nella stessa azienda o essere affidate a soggetti esterni, per cui gli obblighi e le responsabilità di ciascuna fase di sviluppo dovrebbero ricadere sul soggetto che non ha adempiuto alle best practices previste per la singola tappa di produzione. Tuttavia, in molti casi l'analisi sull'impatto della tecnologia e individuazione dei rischi, può essere condotta soltanto nella fase ultima ed applicativa del sistema; bisognerà, quindi, attendere che venga implementata l'IA nella forma di prodotto o servizio con la quale verrà resa disponibile al pubblico perché si possano identificare i pericoli determinati dalla macchina. Una particolare metodologia è, ad esempio, l'utilizzo di Red teaming AI systems che simulano l'attacco di un hacker o un nemico per scoprire le debolezze del sistema. Una volta identificati i punti deboli è possibile intervenire con misure specifiche ed, infine, trarre le conclusioni all'interno di report, che illustrino come l'insieme di misure adottate riesca a portare il software all'interno del livello di rischio ritenuto accettabile. Questa stretta procedura di monitoraggio e raccolta dei dati è l'unica strada per stimolare lo sviluppo delle nuove tecnologie, senza perdere il controllo sulle conseguenze sociali ed economiche dei sistemi di IA. Tali obiettivi potrebbero essere perseguiti più velocemente grazie al coinvolgimento delle Università, alle quali dovrebbe essere ampliato l'accesso alle risorse ed alle informazioni sull'IA; in questo modo i centri di ricerca potrebbero sostenere l'industria ed il governo nello sviluppo di nuove soluzioni tecniche per minimizzare i rischi dei software[51].

[50] Secondo l'OECD «le fasi del ciclo di vita del sistema di intelligenza artificiale sono: i) pianificazione e progettazione, raccolta ed elaborazione dei dati e costruzione del modello ed interpretazione; ii) verifica e convalida; iii) dispiegamento iv) funzionamento e monitoraggio.» in OECD Multilingual Summaries Artificial Intelligence in Society, https://www.oecd-ilibrary.org/sites/b60a092a-it/index.html?itemId=/content/component/b60a092a-it#:~:text=Le%20fasi%20del%20ciclo%20di,dispiegamento%20iv)%20funzionamento%20e%20monitoraggio.

[51] Cfr. Microsoft, Governing AI: A Blueprint for the Future, in query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14Gtw ;