ISSN 2723-9195

CERIDAP

Rivista Interdisciplinare sul Diritto delle Amministrazioni Pubbliche

Estratto

Fascicolo
4 / 2023

OTTOBRE - DICEMBRE

L'Artificial Intelligence Act Proposal e la regolamentazione degli algoritmi predittivi: luci e ombre

Elena Falletti

DOI: 10.13130/2723-9195/2023-4-19

Lo scopo di questo articolo è analizzare sommariamente le fasi del procedimento di approvazione dell'Artificial Intelligence Act Proposal, in particolare per quel che concerne gli algoritmi predittivi ad alto rischio, per valutare la variabilità degli effetti delle modifiche in materia.

The Artificial Intelligence Act Proposal and the regulation of predictive algorithms: lights and shadows
This article summarises the approval process of the Artificial Intelligence Act Proposal, particularly concerning high-risk predictive algorithms, to assess the effect of this legislative change.

1. Introduzione

Uno dei pregi della letteratura riguarda la sua natura di ponte con le generazioni

che ci hanno preceduto per comprendere come chi apparteneva a quelle passate immaginava il futuro, cioè il nostro presente. Questo paragone è assai utile per comprendere l'impatto dell'evoluzione culturale avvenuta tra il passato e il tempo attuale, nonché misurare quanto l'immaginario e il suo bagaglio di timori e paure abbiano costruito la nostra mentalità.

Per lo scopo di quanto si va ad affrontare Isaac Asimov è il testimone perfetto[1] perché la sua opera ha introdotto i lettori ad una realtà futuribile popolata di robot positronici (cioè automi obbedienti alle tre leggi della robotica)[2] o automi artificiali in grado di interagire con gli esseri umani seguendone le logiche. Al contempo è assai distante dalla nostra realtà, perché seppure questo scrittore sia stato profetico, non lo è stato in senso realistico, almeno al momento, in quanto non è (ancora) possibile creare macchine effettivamente in grado di interagire da pari a pari con gli esseri umani, come avviene in "Io, Robot", mentre i programmi di intelligenza artificiale sono sovente utilizzati dagli esseri umani in senso discriminatorio o prevaricante nei confronti di altri esseri umani.

È possibile osservare che nell'esperienza giurisprudenziale ciò sia già accaduto in precise fattispecie quali nell'erogazione di benefici tributari[3] o nello svolgimento di particolari forme di lavoro come il food delivery[4] o nell'assegnazione delle cattedre di insegnamento nelle scuole pubbliche[5] oppure ancora in caso di erogazione di fondi pubblici[6]; mentre alcuni enti pubblici, come la Banca d'Italia[7], hanno adottato precisi regolamenti interni.

In siffatto contesto, la Commissione europea ha elaborato la bozza del Regolamento sull'intelligenza artificiale (d'ora in poi *AIA Proposal*)[8] pubblicata e sottoposta alla pubblica discussione il 21 aprile 2021. Successivamente il Parlamento Europeo ne ha approvato una prima versione ampiamente emendata il 12 giugno 2023[9]. Dopo di che il *Proposal* è stato sottoposto alla fase di discussione nel trilogo con il Consiglio Europeo[10], a partire dal 18 luglio 2023[11].

Gli obiettivi che l'*AIA Proposal* intende raggiungere sono molto ambiziosi ed il suo iter di approvazione si è rivelato essere assai tortuoso nel tentativo di bilanciare la contrapposizione tra gli interessi in gioco, rischiando di affievolirne sia la coerenza sia il rigore.

Infatti, da un lato è emersa, specie in sede di approvazione[12], la pressione esercitata dagli operatori economici, soprattutto extraeuropei, che hanno inteso sfruttare il loro vantaggio competitivo in ambito tecnologico, ripetendo l'esperienza già

accaduta con Internet e le piattaforme *social network*[13]. Dall'altro lato gli interessi da proteggere sono duplici, cioè quelli a garanzia dei cittadini, dei consumatori tesi a rafforzare le tutele dei diritti fondamentali quali dignità, *privacy*, riservatezza e autodeterminazione informativa[14], nonché quelli degli operatori economici europei.

Tale disciplina inquadra lo sviluppo dell'IA presente, nel senso che deve assicurare la tutela dei diritti fondamentali e della *rule of law* ed insieme indirizzare il progresso di quella futura e quindi essere così flessibile da adattarsi a cambiamenti tecnologici che al momento non sono ancora prevedibili. In altri termini, viene richiesto all'*AIA Proposal* un doppio obiettivo: *in primis* di porre in essere una "quadratura del cerchio" tra le esigenze a vantaggio dell'innovazione tecnologica e il riconoscimento delle garanzie a favore dei cittadini, in particolare di quelli vulnerabili. *In secundis,* l'*AIA Proposal* è investito dell'importante compito di fungere quale modello regolatorio giuridico a livello globale, come già avvenuto con il GDPR[15].

Pertanto, l'*AIA Proposal* si inserirebbe in un contesto caratterizzato dall'asserita, quasi "mitologica"[16], imparzialità algoritmica congiuntamente ad una presunta parvenza di buona amministrazione (se riferito ad enti pubblici) o gestione (se inerente a enti privati) di procedimenti e risultati ottenuti con tali strumenti automatizzati.

In realtà, siffatta condizione di apparente neutralità, che non può essere considerata realistica come dimostrato dalle molteplici decisioni giurisprudenziali in tema, rappresenterebbe una strategia di rassicurazione, se non addirittura di vera e propria propaganda, a giovamento dell'opinione pubblica sulla imparzialità delle scelte adottate con i sistemi automatizzati.

Al contrario, l'*AIA Proposal* vorrebbe rispondere alla necessità di una regolamentazione il più possibile precisa al fine di evitare la manipolazione dell'*humanitas* per mezzo di siffatti sistemi (creati a loro volta da esseri umani) che attuino, facendola accettare socialmente, la deresponsabilizzazione di scelte politiche da parte dei loro fautori[17], siano essi funzionari di Stato ovvero rappresentanti politici eletti democraticamente o addirittura dagli stessi privati.

A questo fine, l'impalcatura concettuale dell'*AIA Proposal* adotta un approccio basato sul rischio del danno che l'IA può causare ai diritti fondamentali degli individui soggetti all'uso degli algoritmi. Secondo tale costruzione logico-

giuridica, pertanto, ad un rischio più elevato sofferto dal pubblico di utenti/consumatori corrisponderanno obblighi più stringenti a carico dei creatori/produttori.

L'uso stesso della locuzione "Intelligenza Artificiale" contiene un equivoco intrinseco, dato che, seguendo le istruzioni loro predisposte attraverso il codice, i programmi automatizzati si limitano ad eseguire calcoli senza consapevolezza di contenuto e significato. Questa circostanza rende possibile la manipolazione tra il senso di questa locuzione e il suo significante, poiché comunemente con esso si intende la capacità attribuita a una macchina di elaborare operazioni complesse riconducibili a quelle effettuate da un cervello umano. Di conseguenza l'intelligenza artificiale è diventata un concetto collegato alla letteratura e alla cinematografia di genere fantascientifico, ove le macchine prendono il sopravvento sul genere umano con effetti catastrofici[18].

In siffatto contesto, l'avvento dei calcolatori in grado di elaborare quantità massive di dati (c.d. *big data*), traendo da essi nuova conoscenza, ha contribuito alla formazione dell'equivoco sulla presunta superiorità (intesa in senso di pericolo, o comunque negativo) dell'intelligenza artificiale rispetto alla mente umana, grazie alla velocità di calcolo nell'elaborazione delle informazioni. Il raggiungimento di tale traguardo tecnologico è stato molto enfatizzato, in particolare, quando i calcolatori elettronici hanno sconfitto i grandi maestri nei giochi di strategia, iconici dell'intelligenza stessa, come gli scacchi[19]. Tuttavia, nonostante le sconfitte, gli esseri umani non hanno smesso di giocare a scacchi per il loro piacimento.

In tempi più recenti questa sensazione si è rafforzata con l'introduzione sul mercato di algoritmi di *machine learning* in grado di interagire con gli utenti non esperti attraverso l'uso del linguaggio colloquiale come nel caso dei programmi di intelligenza artificiale generativa quali gli *LLM* (*Large Language Model*). Tali chatbot sono commercialmente noti con i nomi di *ChatGPT*, *Replika*, *DeeplWrite*, *GitHub*, *Copilot* e sono in grado di elaborare risposte apparentemente chiare, ma espresse con un linguaggio autorevole, pertanto apparentemente più affidabile.

La medesima sensazione di sconcerto e di meraviglia ha accompagnato la diffusione degli algoritmi in grado di creare immagini artistiche "*text to image*" con l'inserimento di specifici "*prompt*" (richieste) quali *MidJourney*, *DALL-E 2*,

Starry AI, *DreamStudio* e così via. Il lancio sul mercato sia dei programmi *LLM* che di quelli *text to image* ha provocato contenziosi tra i rappresentanti di autori e artisti in materia di tutela del *copyright*[20] sui testi e sulle immagini da cui tali sistemi elaborano il loro training, assorbendo anche *bias* discriminatori[21].

Tali programmi utilizzano modelli di calcolo probabilistici elaborati sulle statistiche delle parole maggiormente utilizzate[22] (ovvero delle stringhe di testo[23]) nella materia investigata e pertanto non sono in grado di essere realistici, dato che non comprendono il testo da loro stessi elaborato. A parere di chi scrive la loro peculiarità risiede nella proposta di visioni alternative, scollegate dalla realtà[24]. Si tratta di un tema differente, ma non meno complesso, rispetto a quello socialmente più importante riguardante la diffusione di *fake news* e *hate speech*[25], favorito dalla loro modalità di funzionamento. Siffatta problematica sorge nel momento in cui l'utente utilizza il programma *LLM* per ottenere informazioni, sbagliando perché, allo stato dell'arte, è la caratteristica del modello concettuale che impedisce l'accuratezza necessaria dei risultati[26].

A completamento di ciò, i tentativi di limitare le loro distorsioni intrinseche[27] incontrano ulteriori difficoltà da superare, in particolare nella comprensione del funzionamento di questi sistemi, perché essi sono generalmente protetti dalle tutele legate alla proprietà intellettuale, specie del segreto industriale, e pertanto non disponibili alla pubblica conoscenza e diffusione, al contrario dei sistemi open source[28], tuttavia meno diffusi.

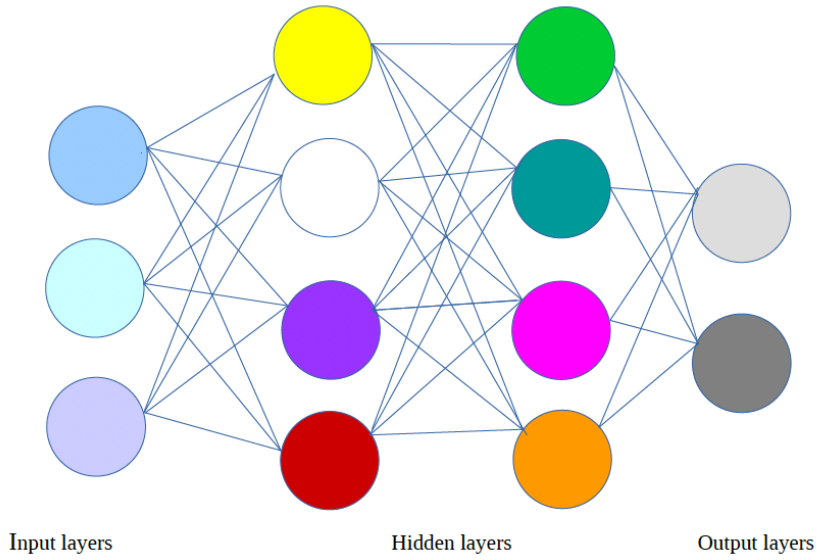
2. L'*Artificial Intellingence Act Proposal* e il *risk-based approach*

L'approccio normativo dell'*AIA Act Proposal*[29] si indirizza verso l'utilizzo di un criterio fondato sulla valutazione preventiva della gravità del rischio[30]. In via generale l'impiego di tale criterio viene giustificato dal risultato positivo ottenuto nel ragionevole bilanciamento tra i vantaggi e le esternalità negative in merito all'autorizzazione dell'uso dello strumento supposto pericoloso, in questo caso i sistemi di intelligenza artificiale. Da un punto di vista normativo, l'utilizzo di siffatti sistemi non comporta né un divieto assoluto né una liceità assoluta, ma sono consentiti nel rispetto di determinate condizioni[31]. Al contrario, seppure il loro utilizzo possa rappresentare una esternalità negativa, il divieto del loro uso

non sarebbe ragionevole in quanto essi apportano vantaggi e miglioramenti rispetto allo *status quo*. Pertanto, bilanciamenti e valutazioni devono essere focalizzati sul risultato, caso per caso.

In questo ambito rientrano i c.d. modelli predittivi, cioè strumenti automatizzati di calcolo probabilistico[32] per la gestione di situazioni ovvero la soluzione di controversie che presentino la caratteristica della serialità, cioè in presenza di circostanze standard ove sussistono quantità massive di dati e condizioni di stabilità socio-culturali[33].

Tali sistemi sono riconducibili nell'alveo degli "algoritmi decisori automatizzati". Con questa espressione ci si riferisce ad una serie successiva di decisioni effettuate attraverso sistemi di calcolo matriciali e interdipendenti tra loro. L'uso di sistemi automatizzati in ambiti seriali è ormai molto diffuso perché vi è la tendenza a sostituire l'elemento decisorio umano[34], nel tentativo (probabilmente illusorio[35], se non del tutto vano) di ottenere risultati prevedibili[36] ed efficienti, ai fini della certezza delle decisioni medesime.



Come è noto[37], tali procedimenti imitano le reti neurali raccogliendo e immagazzinando dati dall'esperienza passata, poi elaborandoli per mezzo di passaggi non visibili: le c.d. "*black box*", dove vengono effettuati i calcoli

matriciali, non ricostruibili nel loro percorso logico. Quanto descritto (e illustrato) ricostruisce in modo molto sommario e semplificato il modello di *machine learning* realizzato con reti neurali artificiali (ANN)[38].

Siffatte ANN imitano ciò che avviene nei neuroni umani attraverso le sinapsi, cioè del procedimento che consente al cervello di imparare. In tali modelli imitativi si riconosce uno strato di ingresso (*input layer*) connesso a sensori che percepiscono le informazioni digitalizzate da elaborare.

Queste possono avere il contenuto più vario, cioè immagini, testi, numeri presenti in diversi strati intermedi nascosti (*hidden layers*) ove le operazioni di calcolo e apprendimento vengono effettuate, e uno strato di uscita (*output layer*) che trasmette i risultati dell'elaborazione.

Ogni neurone di ciascun strato è connesso con tutti gli altri e ciascun collegamento ha una sua valenza, ciò nonostante la ricostruzione del percorso logico seguito dalle unità di apprendimento è confusa e di difficile interpretazione, perciò tali procedimenti vengono definiti ed identificati con l'espressine metaforica di "*black box*".

In tale contesto, i calcoli matriciali che avvengono nelle *black box* non si limitano a ripetere le esperienze già immagazzinate, ma le ricombinano, creando nuova conoscenza; tuttavia, permane presente il rischio di perpetuare i *bias* e i possibili elementi discriminatori in esse contenuti.

Le serialità delle fattispecie a cui siffatti calcoli vengono applicati consente l'elaborazione dei "modelli predittivi", nel senso che si fondano sulla ripetitività delle situazioni al fine di classificare in modo omogeneo i risultati ottenuti.

Facendo riferimento agli Stati Uniti, cioè in una realtà giuridicamente composita nonché tecnologicamente avanzata, è possibile osservare un significativo esempio di come funzionano tali algoritmi predittivi nelle procedure di assegnazione di fondi creditizi o di mutui ipotecari. Siffatto esempio dimostra come sedersi "al tavolo della fratellanza"[39] sia rimasto un sogno non pienamente realizzato e gli stereotipi razziali permeino ancora significativamente la realtà fattuale e quindi quella presente nei database con un evidente *bias* discriminatorio.

In tale fattispecie, la serialità riguarda la situazione dei richiedenti il finanziamento, la valutazione delle loro caratteristiche: sia quelle individuali, quali ad esempio stabilità della posizione lavorativa, della situazione familiare, la propensione al rischio e affidabilità creditizia, ovvero la loro zona di residenza e il

relativo codice postale; sia quelle relative all'ammontare della richiesta di finanziamento e la relativa operazione da concludere, cioè l'apertura di una linea di credito, il prestito a favore di un'azienda, o richiesta di un mutuo per l'acquisto della casa.

Relativamente a quest'ultima circostanza sono state depositate di fronte alle Corti federali americane alcune istanze di "*class action*"[40] da parte dei clienti di una nota banca che hanno ritenuto di essere stati discriminati. Infatti, negli Stati Uniti il tema è sottoposto all'attenzione dell'opinione pubblica da tempo, data la combinazione di effetti insieme sensibili e distorsivi, in relazione alla reale discriminazione su base etnica[41] e le conseguenze ancora presenti sia della crisi dei mutui *subprime* che ha investito il mercato immobiliare americano[42], sia della reazione economica negativa alla Pandemia da *Covid-19*[43], riverberandosi sulla solvibilità ipotecaria tanto collettiva quanto dei singoli.

Ciò consente di proporre una riflessione sull'impatto dei modelli predittivi nella realtà e sulla validità del modello *risk-based approach*. Infatti, l'utilizzo dei modelli predittivi si basa sulla loro efficienza nelle decisioni seriali, pertanto il loro margine di errore, che può presentare importanti conseguenze discriminatorie, viene considerato, malgrado ciò, accettabile. Eppure, siffatto margine è assai più ampio e discriminatorio di quanto possa apparire *prima facie*. L'elemento di rischio, in tale situazione è ambivalente: da un lato è l'uso dell'algoritmo decisorio ad essere aleatorio, nello specifico di violazione dei diritti del soggetto sottoposto alla siffatta decisione predittiva: ed è questa la *ratio* per considerare azzardato l'impiego degli algoritmi in parola.

Dall'altro lato, invece, l'algoritmo predittivo è utilizzato al fine di prevedere se il soggetto sottoposto alla valutazione automatizzata adempierà all'obbligazione, ovvero alla restituzione della somma di denaro per cui ha presentato richiesta, dato che sussiste il rischio dell'inadempimento del debitore e questo va tenuto in conto nel momento del calcolo sulla solvibilità del soggetto richiedente. Infine, sussiste un terzo elemento di rischio, più subdolo perché meno appariscente, e cioè che l'algoritmo elaborerà una decisione potenzialmente discriminatoria combinando i summenzionati fattori.

Alla luce di questa breve analisi è possibile sostenere che l'obiettivo della critica all'uso di algoritmi predittivi è di evitare di cadere in una trappola manipolatoria, cioè confondere la discriminazione (basata per esempio sulle caratteristiche

etniche, di genere, di orientamento politico e così via) mascherandola con gli elementi di rischio.

Il tema concerne la qualità dei dati assunti nel database e come essi vanno trattati nel momento della raccolta e della loro classificazione. Nello specifico, è necessario vagliare come siffatti dati debbano essere (o meno) "purificati" ovvero "puliti" affinché l'algoritmo possa essere programmato esclusivamente sulla solvibilità del richiedente, tralasciando informazioni discriminatorie quali l'origine etnica ovvero la residenza abitativa, dato che si tratta di informazioni che possono essere ricostruite attraverso la combinazione dei dati stessi[44].

Come risolvere siffatto problema? Innanzitutto, è necessario non cadere nella tentazione di modificare la realtà con la quale gli algoritmi si trovano ad interagire. In questo caso, una modifica della rappresentazione di tale realtà nel linguaggio di programmazione consisterebbe in una sua manipolazione, ovvero una sua ricostruzione fittizia. Una simile operazione snaturerebbe il contesto, ma non muterebbe le condizioni dei soggetti sottoposti a rischio discriminatorio e perciò andrebbe evitata.

Per sfuggire al rischio manipolatorio ci si deve focalizzare sul trattamento dei dati. Già dal momento della raccolta deve essere dedicata una specifica attenzione alla natura e alle caratteristiche dei dati applicando rigorosamente le disposizioni del GDPR. Al contrario, se l'attenzione alla selezione dei dati fosse focalizzata *ex post*, il *database* conterrebbe quelle informazioni inquinate da *bias* che porterebbero a risultati inevitabilmente discriminatori. Ne consegue che l'ipotetica coesistenza nel momento applicativo delle discipline del GDPR, in materia di protezione dei dati personali, e dell'*AIA Proposal* sarebbe solo apparente, in quanto le disposizioni dell'*AIA Proposal* dovrebbero essere applicate nelle successive fasi di elaborazione dei dati (dai quali i *bias* dovrebbero già essere stati espunti) da parte del *machine learning* nella *black box*.

In altri termini, per quel che concerne i risultati relativi all'esempio descritto in precedenza, cioè la solvibilità di un soggetto richiedente un prestito di denaro, possono essere considerati dati essenziali ai fini del loro utilizzo algoritmico: la stabilità della posizione lavorativa, la propensione al rischio del soggetto richiedente nonché la sua situazione familiare. Siffatti dati sarebbero quindi inseribili nel *database* da sottoporre ai procedimenti di *machine learning*, ma non la zona di residenza, con il relativo codice postale, che invece riflettono *bias*

riconducibili alle origini etniche e alle condizioni economiche relative ad una classe o a un gruppo sociale, invece che individuali.

Discutibile, invece, se siano essenziali alla definizione della solvibilità l'età del richiedente ovvero il genere perché si tratta di dati sensibili che possono avere una rilevanza non oggettiva. Quale esemplificazione ci si può riferire agli anni di carriera ovvero di attività professionale del richiedente e quindi risalire all'ammontare della retribuzione per i lavoratori dipendenti, che godono dell'erogazione di uno stipendio, ma non dei lavoratori autonomi, i quali ricavano il loro reddito dall'attività professionale o imprenditoriale. Questa distinzione può assumere rilevanza oggettivamente discriminatoria per quanto riguarda il genere del richiedente il finanziamento, dato che il "*gender gap*" retributivo è una condizione ancora sofferta dal genere femminile rispetto a quello maschile[45].

Da questo semplice esempio riguardante una situazione assai diffusa emerge come la redazione dell'*AIA Proposal* debba tenere conto di circostanze fattuali diversificate che rischiano di frammentare il testo normativo, venendo così meno ad uno dei principi generali tradizionali della norma giuridica, cioè quello di essere "generale e astratta", mentre la norma si sta trasformando in "particolare e concreta", con effetti ancora più intrusivi rispetto a quanto già delineato dalla dottrina che in passato si era occupata della "frantumazione" dell'ordinamento giuridico[46]. Ciò in particolare sia per la categorizzazione sociale con cui occorre confrontarsi quando ci si avvale dei software predittivi, sia per l'intervento delle c.d. "*sandbox*", nella gerarchia delle fonti, delle quali ci occuperemo successivamente.

Come è noto, l'*iter legis* affrontato dai Regolamenti dell'Unione Europea è lungo e laborioso[47] e gli interventi discussi e gli emendamenti apportati al testo originario dell'A*IA Proposal* sono stati indirizzati al riconoscimento dell'AI come una «*intelligenza artificiale antropocentrica e affidabile e garantire un livello elevato di protezione della salute, della sicurezza, dei diritti fondamentali, della democrazia e dello Stato di diritto, nonché dell'ambiente, dagli effetti nocivi dei sistemi di intelligenza artificiale nell'Unione, sostenendo nel contempo l'innovazione e migliorando il funzionamento del mercato interno*»[48].

L'*AIA Proposal* considera i sistemi decisori automatizzati/standardizzati ad alto rischio in quanto «*il termine "automatizzato" si riferisce al fatto che il*

funzionamento dei sistemi di IA prevede l'uso di macchine»[49], che possono non essere soggette al completo controllo umano durante le operazioni di *machine learning.*

Prendendo a prestito le parole di Walt Whitman[50] tale vasta definizione contiene più contraddizioni che moltitudini perché riguarda sia i timori che solleva l'intelligenza artificiale, che devono essere placati con "elevate garanzie di protezione", sia le aspirazioni sul sostegno all'innovazione e il miglioramento del mercato. Timori e garanzie sono però poste nell'area di influenza delle "macchine". Utilizzando tale termine e collegandolo ad un rischio, l'*AIA Proposal* sembra catalogare i sistemi automatizzati in qualcosa di alieno rispetto all'*humanitas*, invece che un prodotto pensato, creato e realizzato dalla stessa mente umana.

Nel modello basato sul rischio l'oggetto della disciplina è il rischio del danno[51], non il danno in sé. Si tende quindi ad anticipare la valutazione delle conseguenze possibili analizzando la probabilità della verifica del danno stesso[52]. Il vantaggio di questo approccio risiederebbe nel tentativo di protezione preventiva che riduca la probabilità stessa delle infrazioni, rispetto alla organizzazione e standardizzazione di una serie di rimedi successivi alla violazione delle norme[53].

In tale prospettiva, l'obbligazione principale per chi produce e immette sul mercato sistemi di intelligenza artificiale è di *(auto)*analizzare la propria attività e quindi di *(auto)*valutare i rischi connessi all'adozione delle tecnologie automatizzate. Di conseguenza siffatto soggetto (o ente) è tenuto ad adottare schemi organizzativi (validati da organismi indipendenti all'uopo preposti) sia verso l'interno, cioè nei propri procedimenti produttivi di beni o servizi, sia all'esterno cioè nei confronti degli utilizzatori, consumatori ovvero utenti dei propri sistemi[54].

Sembrerebbe quindi che l'obiettivo dell'*AIA Proposal* sia di organizzare "il controllo del rischio"[55]. Nonostante il concetto, lo scopo e, di conseguenza, la stessa classificazione individuate dall'*AIA Proposal* siano ancora dibattuti nel trilogo[56], pare di interesse analizzare come essi siano stati elaborati, in quanto tali elementi sono stati oggetto di un intenso confronto scientifico pubblico da quando il testo dall'*AIA Proposal* è stato reso diffuso[57]. A questo fine, ed in via generale, la proposta di Regolamento individua quattro livelli di rischio, che sono raggruppati e analizzati nei seguenti sottoparagrafi:

2.1. I sistemi vietati perché inaccettabilmente rischiosi

Il primo livello raggruppa i sistemi di IA che presentino un rischio valutato come inaccettabile; pertanto, è vietato l'uso di siffatti sistemi quando costituiscono un pericolo eccessivo e sbilanciato a discapito della protezione dei diritti fondamentali dei soggetti che possono subire il trattamento dei loro dati personali da parte di siffatti sistemi.

In questa categoria rientra tutto ciò che rappresenta una «*chiara minaccia per la sicurezza, i mezzi di sussistenza e i diritti delle persone*». Tra questi sono annoverati gli algoritmi manipolativi, di sfruttamento, ovvero di controllo sociale[58] e di *social scoring*[59]. Ugualmente sono proibiti i sistemi automatizzati che utilizzano tecniche subliminali[60], manipolative[61] o ingannevoli per distorcere il comportamento[62], quelli che sfruttano le vulnerabilità di individui o di gruppi specifici, che creano o espandono *database* di riconoscimento facciale attraverso *web scraping*[63] non mirato e i sistemi di polizia predittiva[64]. Parimenti sono vietati i sistemi di intelligenza artificiale mirati a elaborare le emozioni[65] nella gestione delle frontiere, sul posto di lavoro e nell'istruzione e quelli per la valutazione del rischio che prevedono reati penali o amministrativi, nonché l'utilizzo di programmi di riconoscimento biometrico in tempo reale.

Su questi ultimi, il Legislatore europeo si è manifestato molto attento precisando la definizione rispetto a quella contenuta nel GDPR e quindi distinguendo i dati biometrici[66] dai dati di identificazione biometrica[67]. I primi riguarderebbero i connotati relativi ad un soggetto che consentirebbero la sua precisa identificazione, mentre i secondi riguarderebbero una sfera più ampia di caratteristiche personali dell'individuo riconducibili a tratti identificativi più generali che ne permetterebbero la sua individuazione in mezzo a un gruppo ovvero a una folla. Tale distinzione comporta l'introduzione di un aspetto aggiuntivo in materia di biometria[68], cioè il diverso trattamento giuridico tra l'identificazione in tempo reale (in via di principio sempre proibita, tranne in specifici casi previsti dalla legge) da quella remota.

A questo proposito, rispetto al testo originario dell'*AIA Proposal*, il Parlamento europeo introduce un ulteriore elemento, cioè la «*categorizzazione biometrica, la quale prevede l'assegnazione di persone fisiche a categorie specifiche o la deduzione delle loro caratteristiche o dei loro attributi, utili a inserire le singole caratteristiche*

individuali dei soggetti da identificare affinché tali dati siano inseribili in specifici cluster delle banche dati»[69].

Per quel che concerne la distinzione tra sistemi di riconoscimento biometrico in tempo reale e quelli a posteriori, oltre alla distinzione tecnica[70], si evidenzia come l'*AIA Proposal* tratti la questione del consenso della persona sottoposta al trattamento biometrico dei suoi dati senza distinguere con quale delle due modalità i dati in questione siano stati acquisiti[71].

Siffatti sistemi sono vietati dal *Proposal* perché i rischi conseguenti il loro utilizzo sono inaccettabili in quanto in contrasto con i valori della Carta dei diritti fondamentali dell'Unione Europea, specificamente in merito alla protezione della dignità (art. 1), dell'integrità fisica e psichica (art. 7) e della riservatezza dei dati personali (art 8). Tuttavia, il *Proposal* ammette alcune deroghe in caso di situazioni eccezionali che ne giustifichino l'utilizzo in particolare per ragioni di sicurezza pubblica, pericolo imminente per la vita o l'incolumità fisica delle persone fisiche, ricerca di minori scomparsi e rischio di attacchi terroristici, dietro autorizzazione giudiziaria.

Ciò facendo, il *Proposal* concede un ampio *(e vago)* margine di discrezionalità a favore dell'ordinamento statale. Nonostante l'apparente contraddizione relativa alla *(impossibile?)* coesistenza tra divieto totale e sue eccezioni non chiaramente specificate, il Legislatore europeo si è trovato di fronte alla difficoltà di prevedere e bilanciare lo sviluppo di un settore foriero di interessi, anche divergenti, e soggetto ad una velocissima[72] evoluzione tecnologica e tecnica[73].

Per esempio, in materia di algoritmi predittivi utilizzati a scopo di polizia, soccorre la giurisprudenza costituzionale tedesca, che ha fissato rigidi limiti all'utilizzo di software predittivi di polizia. In particolare, Il *Bundesverfassungsgericht*[74] ha stabilito che il ricorso all'analisi automatizzata dei *big data* è costituzionalmente legittimo solo ove serva alla prevenzione di determinati reati definiti in modo preciso dal legislatore affinché le limitazioni del diritto all'autodeterminazione informativa, espressione del diritto alla personalità, sia giustificato. In particolare, la messa in pericolo di beni giuridici rilevanti deve essere concreta e sottoposta ad uno stretto scrutinio di proporzionalità che giustifichi l'intrusione effettuata dai software predittivi. A questo proposito, il grado di incisività di tale scrutinio deve essere concretamente determinato dalle disposizioni di legge relative ai presupposti e modalità d'uso relativi al ricorso ai

suddetti programmi.
Autorevole dottrina[75] osserva che la Commissione Europea ha tentato di istituire un confine fondato su concetti giuridici necessariamente indeterminati, ma non per questo meno concettualmente solidi, in quanto occorre lasciare uno spazio normativo sufficiente da ricomprendere evoluzioni tecnologiche dell'AI ancora non presenti, né immaginabili, senza però cadere nel rischio, ancora più grave, dell'arbitrio[76] e stabilire argini normativi a garanzia delle libertà fondamentali e della presunzione di innocenza.

2.2. I sistemi di IA ad alto rischio accettabili, ma con riserva

Sono classificati "ad alto rischio" i sistemi di IA che presentano un impatto nocivo significativo "sulla salute, la sicurezza e i diritti fondamentali delle persone nell'Unione" (art. 6 del *Proposal*). La disciplina di tali sistemi divide questi sistemi in due categorie:

1. quelli destinati ad essere utilizzati come componenti di sicurezza di prodotti soggetti a valutazione di conformità da parte di terzi;
2. sistemi specificamente elencati nell'Allegato III, che presentano implicazioni principalmente in relazione ai diritti fondamentali[77].

L'uso di simili modelli può essere consentito, ma essi devono essere soggetti ad un controllo preventivo in cui sia accertata la presenza di precisi requisiti di tutela della dignità umana e del rispetto dei diritti fondamentali. L'individuazione di tale categoria si basa sia sulla funzione specifica attribuita all'algoritmo sia sul suo scopo complessivo, il quale deve tuttavia essere comprensivo anche del riferimento ai suoi obiettivi specifici. Ad esempio, gli algoritmi utilizzati dalle piattaforme di *food delivery* presentano la funzione complessiva dell'organizzazione della gestione delle consegne delle pietanze in un certo luogo e per un certo tempo, ma hanno lo scopo specifico di attribuire ciascun *crowdworker* la consegna della pietanza richiesta all'indirizzo del *customer*. Gli algoritmi di valutazione della solidità finanziaria di un cliente bancario hanno lo scopo di gestire il capitale finanziario destinato al soddisfacimento delle richieste di finanziamento, dall'altro si occupano di valutare nello specifico la solvibilità dei clienti e così via.

In ciascuno di questi passaggi possono realizzarsi output discriminatori, sia sotto il profilo categoriale (per esempio, in relazione ad una classe specifica di clienti bancari, i quali sono residenti in una certa zona considerata di basso pregio immobiliare) sia per la presenza di specifiche caratteristiche individuali (per esempio, il cliente non ha conseguito un determinato titolo di studio). Ai fini di tale disamina è necessaria una ulteriore valutazione di conformità sia alla normativa di settore sia alla tutela dei diritti fondamentali[78].

Si tratta di una amplissima gamma di strumenti (utilizzati nelle aree specificate dall'*Annex III*[79]) e riguardano i modelli utilizzati nel reclutamento lavorativo ovvero nei dispositivi medici diagnostici, i modelli di identificazione biometrica da remoto delle persone fisiche, di gestione delle infrastrutture (sia quelli impiegati nelle c.d. "*smart cities*", come i semafori intelligenti, sia quelli utilizzati nella gestione delle forniture di servizi quali erogazione dell'acqua, del gas, dell'elettricità), ancora per scopi di istruzione o formazione del personale, di gestione della migrazione, dell'asilo e del controllo delle frontiere, dell'amministrazione della giustizia e dei processi democratici e così via. Il rinvio all'*Annex III* per la definizione di una regola o un divieto riguarda una tecnica normativa funzionale alla velocità di aggiornamento dello schema regolatorio. In caso contrario, la normativa andrebbe incontro ad una tanto rapida quanto fisiologica obsolescenza[80].

Al fine di monitorare il grado di senescenza della normativa, gli artt. 61[81] e 63[82] prevedono che venga attuato un sistema di monitoraggio e meccanismi di vigilanza sia preventivi all'immissione sul mercato dei prodotti attraverso una valutazione di conformità, sia successivi per mezzo di specifiche indagini.

Tali sistemi sono orientati ad attuare una forma di proceduralizzazione del rischio funzionale alla limitazione delle violazioni dei diritti umani mediante requisiti, certificazioni e controlli affinché il rischio venga ridotto ad un livello "ritenuto accettabile"[83].

In cosa consiste siffatta accettabilità del rischio? Il Legislatore europeo sembra essere consapevole della «*portata dell'impatto negativo del sistema di IA sui diritti fondamentali protetti dalla Carta dei diritti fondamentali dell'UE e quest'ultima è il parametro rilevante ai fini della classificazione di un sistema di IA tra quelli ad alto rischio, in particolare di quello discriminatorio*»[84].

Ulteriormente, «*le garanzie necessarie affinché il rischio sia ricondotto*

all'accettabilità comprendono il diritto alla dignità umana, il rispetto della vita privata e della vita familiare, la protezione dei dati personali, la libertà di espressione e di informazione, la libertà di riunione e di associazione e la non discriminazione, il diritto all'istruzione, la protezione dei consumatori, i diritti dei lavoratori, i diritti delle persone con disabilità, l'uguaglianza di genere, i diritti di proprietà intellettuale, il diritto a un ricorso effettivo e a un giudice imparziale, i diritti della difesa e la presunzione di innocenza e il diritto a una buona amministrazione»[85].

Oltre a tali diritti, è importante sottolineare che i minori godono di diritti specifici sanciti dall'articolo 24 della Carta dell'UE e dalla Convenzione delle Nazioni Unite sui diritti del fanciullo (ulteriormente elaborati nell'osservazione generale n. 25 della Convenzione delle Nazioni Unite sui diritti del fanciullo per quanto riguarda l'ambiente digitale), che prevedono la necessità di tenere conto delle loro vulnerabilità e di fornire la protezione e l'assistenza necessarie al loro benessere[86].

Tra gli articoli emendati ed approvati dal Parlamento europeo vi è una norma che metaforicamente sembrerebbe voler rappresentare un'ancora di salvezza rispetto all'effettiva realtà dello sviluppo algoritmico e riguarda la previsione del controllo umano sui sistemi di AI ad alto rischio[87]. A questo proposito, gli operatori sono tenuti ad attuare la sorveglianza umana da parte di soggetti competenti, adeguatamente qualificati e formati e dispongano delle risorse necessarie per assicurare l'efficace supervisione del sistema di IA a norma dell'articolo 14 del *Proposal* medesimo[88].

Infatti, secondo l'intenzione del Legislatore Europeo, la previsione di tale sorveglianza umana «*mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza, i diritti fondamentali o l'ambiente che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista*»[89].

Tuttavia, il *Proposal* sembra mettere sullo stesso piano differenti tipi di algoritmo con tre approcci di apprendimento automatico, cioè:

1. il *machine learning* supervisionato, ove il sistema algoritmico apprende attraverso l'insegnamento, cioè un addestramento condotto sulla base di un set di dati iniziali e di risposte già validate;

2. il *machine learning* "di rinforzo", in cui il sistema apprende dai risultati delle azioni proprie o altrui distinguendo tra successi e fallimenti e conseguentemente assimilando il procedimento di apprendimento al fine di apprendere dai risultati corretti scansando gli errori;
3. il *machine learning* non supervisionato, cioè quando il sistema è autonomo dall'intervento dei suoi creatori o di terzi ed è in grado di individuare autonomamente associando dati e relazioni provenienti dai dati che gli sono stati forniti[90].

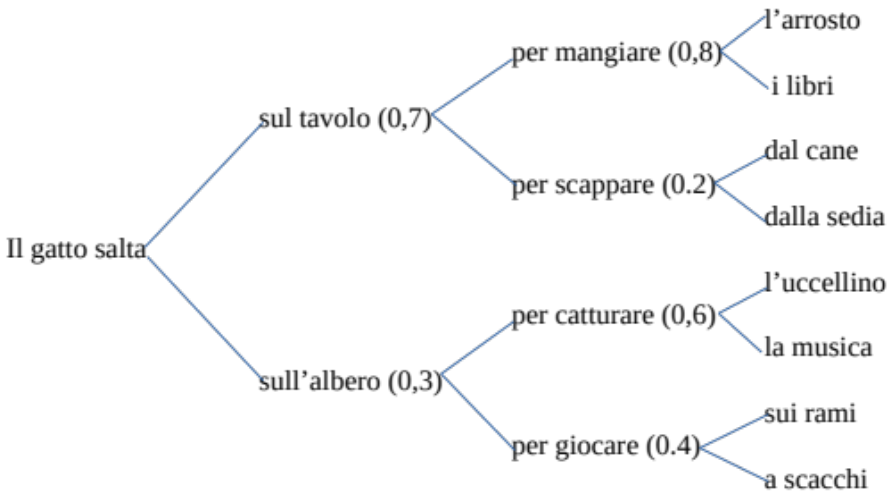
Su questo punto, ci si può chiedere se il ruolo del supervisore umano previsto dal *Proposal* possa essere rivestito da una persona che appartiene allo stesso team aziendale di sviluppo degli algoritmi, supervisionati o meno, o se sia necessario rivolgersi a un soggetto terzo. Il testo dell'art. 14 non è esplicito perché si esprime in termini di "persone fisiche incaricate", privilegiando la loro caratteristica umana, rispetto alla necessaria e contestuale assenza di conflitti di interesse.
Tra i compiti di siffatto supervisore umano vi è quello di essere consapevole delle capacità e dei limiti pertinenti del sistema di IA ad alto rischio e comprenderli a sufficienza, nonché «*essere in grado di monitorarne debitamente il funzionamento, in modo che i segnali di anomalie, disfunzioni e prestazioni inattese possano essere individuati e affrontati quanto prima*»[91]. Inoltre, il supervisore umano deve «*essere in grado di intervenire sul funzionamento del sistema di IA ad alto rischio o di interrompere il sistema mediante un pulsante di "arresto" o una procedura analoga, che consenta di arrestare il sistema in condizioni di sicurezza, tranne se l'interferenza umana aumenta i rischi o è suscettibile di incidere negativamente sulle prestazioni in considerazione dello stato dell'arte generalmente riconosciuto*»[92].
La persona fisica deputata al ruolo di responsabile della sorveglianza deve garantire che le misure pertinenti e adeguate in materia di robustezza e cybersicurezza siano periodicamente monitorate, adeguate ed aggiornate per verificarne l'efficacia[93].
Infine, il Parlamento Europeo ha introdotto nel testo del *Proposal* l'obbligo (dal quale sono esentate le PMI) di predisporre una valutazione d'impatto degli effetti del sistema ad alto rischio sui diritti fondamentali prima di metterli in servizio.
Infine, «*la valutazione d'impatto dovrebbe essere corredata di un piano*

dettagliato che descriva le misure o gli strumenti che contribuiranno ad attenuare i rischi per i diritti fondamentali individuati al più tardi a partire dal momento della loro messa in servizio. Se tale piano non può essere individuato, l'operatore dovrebbe astenersi dal mettere in servizio il sistema»[94].

2.3. I sistemi a rischio minimo o nullo

I modelli di IA con rischio minimo o nullo sono individuati per sottrazione rispetto alle categorie precedenti. L'utilizzo di tali sistemi è consentito, ma nel rispetto di obblighi di informazione e trasparenza sulla loro natura di sistema algoritmico. Tra di essi, nel testo originario dell'*AIA Proposal* sono stati inseriti i chatbot. Come è noto, i chatbot sono stati considerati, specie dagli enti normativi giuridici, programmi algoritmici meramente serventi, cioè in grado di assistere la clientela nell'acquisto di beni e servizi online ovvero aiutare l'utenza nell'uso di personal computer, auto a guida assistita, applicazioni software e così via. L'inserimento di tali software nella categoria di basso rischio da parte del Legislatore europeo ha evidenziato la sottovalutazione sia della varietà di tali programmi per l'ente estensore della prima bozza, cioè la Commissione Europea, non aveva ancora consapevolezza dello stato dell'arte rispetto allo sviluppo tecnologico in materia nonostante il dibattito sul tema fosse presente nella pubblicistica internazionale e nelle conferenze di settore da diverso tempo[95].
In questo contesto occorre distinguere i chatbot rispetto alle loro funzionalità poiché quelli come gli *LLM* possono assurgere allo svolgimento di un ruolo di primo piano nella diffusione di disinformazione e di *fake new*s e manifestano significative questioni per quel che concerne l'utilizzo di tali strumenti algoritmici, in quanto il controllo sulla provenienza dei materiali informativi che li alimenta si indebolisce vieppiù, nonostante i tentativi dei suoi programmatori di evitare le interferenze mistificatrici[96].
Come è ormai noto, siffatti modelli linguistici sono modelli probabilistici istruiti a mappare la probabilità che si verifichi una sequenza di parole (una frase, un'espressione e così via). Gli *LLM* vengono addestrati su massive quantità di testi e da questi ricavano le distribuzioni di probabilità. Più ampia è la quantità di testi di allenamento (*training*), maggiore è la probabilità statistica della frequenza di termini inerenti alla richiesta e quindi la possibile accuratezza del risultato[97]. Le

differenze principali tra gli *LLM* e i modelli linguistici comuni sono che gli *LLM* vengono addestrati sui *big data* con un numero di calcoli esponenzialmente maggiore e usando sistemi molto più sofisticati perché applicano le reti neurali (già analizzate in precedenza). Per visualizzare il loro funzionamento in modo semplice e lineare si propone l'esempio che segue:



Nei compiti che comportano la generazione di testo (ad esempio, l'elaborazione di riassunti, la formulazione di risposte a domande, il completamento di domande), gli *LLM* utilizzano la probabilità condizionale, cioè quando decidono la parola successiva da scegliere, gli *LLM* considerano le parole precedenti in una sequenza e, sulla base di queste, selezionano la parola più probabile[98]. Metaforicamente è come se si volesse andare in avanti ma orientandosi a ritroso in direzione del punto di partenza invece che rispetto al punto di arrivo.

Nel corso dello svolgimento del trilogo, i co-legislatori hanno mutato il loro approccio sui programmi *LLM*[99]. L'elemento innovativo concerne l'introduzione della loro definizione quali "*foundation model*" per distinguerli rispetto ai programmi di IA non generativa.

Si tratterebbe di «*AI model (...) capable to competently perform a wide range of distinctive task*»[100]. Tra questi vanno ulteriormente evidenziati i "*very capable foundation model*", le cui capacità vanno «*beyond the current state-of-the-art and may not be fully understood*». Il riferimento a *ChatGPT* è chiaro, anche se implicito, e comprensibile in relazione al criterio adottato per l'identificazione e

cioè l'enorme capacità di calcolo utilizzata per il suo addestramento[101].

Nel contesto dell'adozione del *risk-based approach*, sembrerebbe che i co-legislatori abbiano surrettiziamente introdotto una quinta categoria di sistemi rischiosi, combinando la valutazione del rischio più elevato che definisce la loro natura, in confronto ai chatbot più evoluti, con le caratteristiche del loro funzionamento[102].

Ci si può chiedere in cosa tali modelli si possano distinguere dagli algoritmi predittivi. Seppure entrambi siano in grado di elaborare risultati discriminatori, i "*foundation model"* possono avere un impiego più generale e vasto, mentre gli algoritmi predittivi limitano la loro utilizzabilità in ambiti più specifici, anche se non meno intrusivi.

Ne conseguirebbe che potenzialmente i "*foundation model*" intrinsecamente abbracciano un maggiore rischio di inquinamento della sfera informativa dalla quale gli algoritmi di diverso genere, predittivi compresi, attingono le loro fonti.

Dalla prospettiva delle modalità di training e apprendimento, la bozza emendata intende regolare l'AI generativa vietando di utilizzare qualsiasi materiale protetto da *copyright* nell'addestramento dei modelli linguistici[103], ma obbligando di esplicitare l'indicazione sull'utilizzo di contenuti protetti da copyright utilizzati per alimentare modelli di altra natura, oltre all'obbligo di dichiarare che il contenuto è stato generato da un'intelligenza artificiale[104].

Infine, i modelli *if/then*, i servizi di traduzione automatica, i filtri antispam sono considerati di rischio nullo. Infatti, il Considerando 6, nella nuova versione modificata dell'Emendamento 18, afferma che le principali caratteristiche dell'intelligenza artificiale riguardano le sue capacità di apprendimento, ragionamento o modellizzazione che la distinguano da sistemi software o approcci di programmazione più semplici.

A tale categoria residuale non si applicherebbero le classificazioni di rischio predisposte dall'*AIA Proposal* e questi algoritmi possono essere utilizzati senza particolari oneri, ferma restando l'opportunità dell'adozione volontaria di codici di condotta (*ex* art. 69)[105]. Tale affermazione sembrerebbe estrapolare i modelli *if/then* (per esempio i sistemi esperti in quanto sistemi a rischio basso o nullo) dal novero dell'AI, lasciando spazio soltanto a quelli che aderiscono al modello relativo al *machine learning*, poiché si tratta di sistemi di IA progettati per funzionare con livelli di autonomia variabili, quindi con un certo grado di

autonomia di azione rispetto ai controlli umani e di capacità di funzionare senza l'intervento umano. Ulteriore ambito di interesse per il Legislatore europeo riguarda tecniche decisorie più semplici come alberi decisori o stime "bayesiane" nel momento in cui vengono inserite in combinazione alle *black box* in sistemi ibridi[106].

3. Gli algoritmi predittivi, l'impatto dei diritti fondamentali sui sistemi ad alto rischio e la discussione sull'*AIA Proposal*

Durante la procedura del trilogo[107], successiva all'approvazione degli emendamenti da parte del Parlamento Europeo, le delegazioni degli Stati Membri appartenenti al Consiglio Europeo hanno focalizzato l'attenzione su due specifici aspetti, entrambi inerenti alla disciplina degli algoritmi predittivi, cioè le misure di supporto all'innovazione, contenute negli articoli 53-55a, dedicati alle applicazioni concrete degli spazi di sperimentazione normativa, e la valutazione dell'impatto della disciplina dei diritti fondamentali sui sistemi di intelligenza artificiale ad alto rischio.

Entrambi gli argomenti sono rilevanti per l'analisi della disciplina sugli algoritmi predittivi e i loro effetti discriminatori.

Per quanto riguarda la parte relativa all'innovazione, le obiezioni del Consiglio europeo riguardano le disposizioni relative alle *sandbox*[108], mentre per quel che concerne l'impatto sui diritti fondamentali con l'uso degli *high-risk AI system* si osserva che la versione approvata dal Parlamento europeo ha introdotto un nuovo obbligo di sorveglianza umana sugli effetti dell'uso degli algoritmi (specie se predittivi) a carico degli operatori; mentre, al contrario, tale obbligo è stato sconfessato dal Consiglio europeo. Ciò nonostante, ai fini del raggiungimento di un compromesso politico per l'approvazione del *Proposal* entro la scadenza della Legislatura europea nel giugno 2024, la Presidenza del Consiglio europeo si è resa disponibile ad acconsentirne l'inclusione solo a seguito a specifiche modifiche.

Di cosa si tratta? Innanzitutto, il "*Fundamental Rights Impact Assessment*" dovrà essere adottato esclusivamente laddove nel settore pubblico verranno utilizzati sistemi ad alto rischio. Ciò rappresenta una questione rilevante perché se il settore pubblico è tenuto a rispettare i limiti che ciascun ordinamento costituzionale

degli Stati Membri stabilisce a tutela dei cittadini, anche in caso di utilizzo di software predittivi (come si è già espressa la giurisprudenza[109], anche costituzionale[110], sul punto, in particolare sul principio di presunzione di innocenza), nel settore privato i rischi discriminatori o di risultati ad impatto negativo sui destinatari non sono minori. Infatti, è sufficiente richiamare alla mente le conseguenze del loro utilizzo nei settori creditizio e lavoristico, che peraltro sono sottostimati ovvero ignorati.

Tale valutazione sulla potenziale violazione dei diritti fondamentali sembrerebbe essere doverosa solo per gli aspetti non coperti da altri obblighi legali, come per esempio la valutazione dell'impatto sulla protezione dei dati ai sensi del GDPR e dovrebbe essere allineata proceduralmente con i processi esistenti, evitando sovrapposizioni e oneri aggiuntivi.

Inoltre, in caso di adempimento agli obblighi già stabiliti per il fornitore di *high-risk AI systems*, vi sarebbe esclusivamente un obbligo di verifica sui rischi residui che non sarebbero stati attenuati dal fornitore.

Ai sensi degli artt. 9, 10 e 13 dell'*AIA Proposal* ci si potrebbe chiedere se tale obbligo sia sostanziale ovvero meramente formale. A questo proposito ci si potrebbe ulteriormente domandare, in assenza di ulteriori elementi, quali potrebbero essere questi rischi "residui", e rispetto all'identificazione della parte di rischio non residua, in quanto essa non sia stata attenuata dal fornitore e permanga presente nel funzionamento dell'algoritmo medesimo. Al contrario, ci si potrebbe chiedere perché non coordinare siffatto *Fundamental Rights Impact Assessment* con l'utilizzo dello spazio di sperimentazione normativa[111].

Il punto è molto delicato, infatti nel trilogo si dovrà stabilire se includere un nuovo obbligo di condurre una valutazione d'impatto sui diritti fondamentali, come descritto nei punti precedenti. A questo proposito c'è un'ampia richiesta in senso positivo da parte della c.d. società civile[112], alla quale però si oppongono i portatori di interessi delle grandi piattaforme, come quelle statunitensi[113] attraverso i relativi "*position papers*".

I primi comprensibilmente chiedono un rafforzamento della protezione dei diritti umani, in particolare al mantenimento delle disposizioni relative all'art. 29 sul *Fundamental Rights Impact Assesment*, prima dell'utilizzo di software ad alto rischio, coinvolgendo rappresentanti della società civile e dei consumatori, in quanto persone coinvolte in questo processo. Inoltre, viene manifestamente

richiesto a tutti operatori degli *high-risk AI system* di registrarsi nel database pubblico dedicato a siffatti programmi.
L'esperienza giurisprudenziale illustra che, assorbendo i *bias* presenti nella società, gli algoritmi predittivi tendono a svantaggiare le categorie vulnerabili per origine etnica, come accaduto in un caso olandese in merito ad una ingiusta richiesta di restituzione di assegni familiari a carico di famiglie formate da persone di origine straniera, in particolare provenienti dalle ex colonie[114] mantenendo nei fatti una discriminazione che, sotto un profilo giuridico-costituzionale, i servizi del settore pubblico sarebbero tenuti ad evitare tutelando invece il principio di imparzialità, dignità ed uguaglianza.
Ulteriormente, i fornitori di IA con sede al di fuori dell'UE, i cui sistemi hanno un impatto sulle persone al di fuori dell'UE, devono essere soggetti agli stessi requisiti di quelli all'interno dell'UE, come accade già per il GDPR, dal quale questa disciplina pare essere mutuata.

4. Il rimedio specifico per le decisioni automatizzate, predittive e non

La conoscibilità del procedimento delle reti neurali è di difficile interpretazione per la modalità stessa del loro funzionamento. Come accennato in precedenza[115], una rete neurale gestisce i sentieri logici dei dati nella formazione del risultato in maniera non trasparente poiché tale percorso è composto da *input, hidden layers* e *output*, in tali casi si parla di *black box*[116]. Accanto a questi, in specifici ambiti empirici sono applicati modelli di c.d. "*grey box*", la cui interpretabilità è parziale poiché sussiste la combinazione tra passaggi noti e altri rimasti ignoti[117].
Ciò nonostante, anche in siffatti casi può sussistere una modalità di verifica trasparente della rete neurale, cioè attraverso la tracciabilità dell'informazione e del dato con cui l'algoritmo viene nutrito dalla mano umana, cioè dal suo programmatore che decide, e quindi sceglie, a quali fonti informative consentirgli l'accesso. In questo quadro si inserisce il tema della tutela del segreto industriale del proprietario dell'algoritmo. Seppure la proprietà intellettuale sia considerata un diritto fondamentale ai sensi dell'art. 17.2 della Carta dei diritti fondamentali dell'Unione Europea[118], nel bilanciamento tra diritti essa recede se contrapposta all'art. 41 della medesima Carta, relativa al riconoscimento del diritto ad una

buona amministrazione. Infatti, per la pubblica amministrazione sussiste l'obbligo, e quindi il dovere, di motivare le proprie decisioni[119], vieppiù in presenza di un algoritmo che in quelle deliberazioni svolge una funzione decisiva a tal punto da poter essere ritenuto parte integrante del procedimento amministrativo stesso[120].

Parimenti, il diritto all'accesso all'algoritmo e alla spiegazione della decisione da questo prodotto è riconosciuto in diverse statuizioni del Regolamento UE 679/2016, specificamente *in primis* dall'art. 22, il quale riconosce alla parte coinvolta in siffatto processo automatizzato il diritto a che la decisione non sia esclusivamente automatizzata, ma che l'amministrazione, con un esplicito contributo umano, controlli i passaggi decisionali, non convalidandoli in caso di violazione di diritti fondamentali.

Sul punto il *Proposal* interviene in senso aggiuntivo, innovativo ed estensivo nella tutela dei soggetti alla decisione algoritmica rispetto all'art. 22 GDPR. Infatti, l'art. 68 *quater*[121], rubricato "diritto alla spiegazione dei singoli processi decisionali", attribuisce maggiori tutele alla persona soggetta a una decisione automatizzata sulla base dei risultati di un sistema di intelligenza artificiale ad alto rischio, implementato da un operatore[122].

Secondo il disposto della norma emendata dal Parlamento europeo, le persone soggette a una decisione presa dall'operatore sulla base dell'output di un sistema di IA ad alto rischio hanno il diritto di chiedere al suddetto operatore spiegazioni chiare e significative, ai sensi dell'articolo 13, paragrafo 1, sul ruolo del sistema di IA nella procedura decisionale, sui principali parametri della decisione presa e sui relativi dati di input. Siffatta spiegazione del funzionamento dell'algoritmo utilizzato deve essere sufficientemente trasparente da consentire agli utenti di interpretare l'output del sistema e utilizzarlo adeguatamente, sul ruolo dell'algoritmo nella procedura decisionale, sui parametri principali della decisione presa e sui relativi dati utilizzati.

Per poter ottenere la doverosa spiegazione è necessario che la decisione automatizzata produca effetti giuridici o incida in modo analogo e significativo sulle persone stesse in un modo che esse ritengono abbia un impatto negativo sulla salute, la sicurezza, i diritti fondamentali, il benessere socioeconomico o qualsiasi altro dei loro diritti derivanti dagli obblighi stabiliti[123].

Tuttavia, vi è l'eccezione nel caso in cui l'uso di tali sistemi possa derivare dagli

obblighi stabiliti dalla legislazione dell'Unione, ovvero nazionale, purché tali eccezioni o restrizioni rispettino l'essenza dei diritti e delle libertà fondamentali e costituiscano una misura necessaria e proporzionata in una società democratica.
Il testo dell'art. 68 *quater* dell'*AIA proposal* appare innovativo, ma, ciò nonostante, le tutele promosse da tale disposizione restano insufficienti. Invero, l'individuazione dell'operatore quale referente per la summenzionata richiesta di una spiegazione "chiara e significativa" della decisione automatizzata non dovrebbe escludere dal medesimo dovere di risposta il produttore dell'algoritmo utilizzato, cioè il reale creatore della metodologia di raccolta e trattamento dei dati secondo il procedimento logico-decisorio utilizzato dal software.
Nel caso la risposta ottenuta dall'operatore non soddisfacesse, a chi si può rivolgere la persona soggetta richiedente la spiegazione? A parere di chi scrive le possibilità di ricorso possono essere tre:

1. nei confronti dell'Autorità nazionale di controllo, predisposta dall'art. 68 *bis* dell'*AIA Proposal*, secondo cui «f*atto salvo qualsiasi altro ricorso amministrativo o giurisdizionale, ogni persona fisica o gruppo di persone fisiche ha il diritto di presentare un reclamo a un'autorità nazionale di controllo, in particolare nello Stato membro in cui risiede abitualmente, in cui lavora o in cui si verifica la presunta violazione, se ritiene che il sistema di IA che lo riguarda violi il presente regolamento*». Il *petitum* di tale ricorso all'autorità nazionale di controllo sarebbe appunto l'asserita violazione del diritto alla spiegazione non soddisfatta *ex* art. 68 quater;
2. se nella decisione automatizzata criticata, la parte ricorrente constati un caso di trattamento di dati non adeguato ai sensi del GDPR, la parte richiedente può ricorrere al Garante della Privacy;
3. la giurisdizione ordinaria in quanto l'illecito trattamento dei dati personali consiste in una violazione di diritti fondamentali.

Ci si può chiedere se tale previsione si ponga in competizione con il rimedio *ex* art. 22 GDPR dato che quest'ultimo fa espresso riferimento ai trattamenti decisionali, nel senso che questi producono e applicano una decisione avente effetti diretti o indiretti sulla persona interessata, tuttavia apparentemente la risposta sembrerebbe essere negativa, dato che è il solo art. 68 *quater* a riconoscere

esplicitamente che la spiegazione debba essere chiara e significativa, mentre l'attuale testo dell'art. 22 GDPR stabilisce «*almeno il diritto di ottenere l'intervento umano da parte del titolare del trattamento, di esprimere la propria opinione e di contestare la decisione*» da parte della persona sottoposta alla decisione automatizzata.

Apparentemente, quindi, si potrebbero considerare riconosciuti due diritti distinti: l'art. 22 GDPR che riconosce il diritto all'intervento umano sul trattamento dei dati e l'art. 68 quater che (nel momento in cui entrerà in vigore) assorbirà il diritto di richiedere spiegazioni chiare e significative sul procedimento decisionale sia che si tratti di *black box* ovvero di *grey box* per le parti di queste ultime non linearmente esplicabili.

In conclusione, il diritto alla spiegazione consiste in un rimedio predisposto a favore del soggetto che subisce gli effetti della decisione automatizzata prodotta da un algoritmo già operativo e presente sul mercato. L'*AIA Proposal* predispone anche uno strumento utilizzabile in via preventiva, cioè prima che il meccanismo algoritmico possa interagire con il pubblico o venga messo in commercio, cioè le *sandbox*.

5. L'introduzione delle *sandbox* e l'*AIA Proposal*

Un aspetto innovativo e di grande interesse sotto il profilo normativo riguarda l'introduzione delle c.d. "*sandbox*" e se questa possa rappresentare una soluzione appropriata per equilibrare la velocità dello sviluppo tecnologico con le esigenze di tutela dei diritti umani di chi è sottoposto al vaglio decisorio algoritmico.

Le *sandbox*, termine tradotto in lingua italiana «*come spazio di sperimentazione normativa*»[124], sono definite quali «*quadri concreti che, fornendo un contesto strutturato per la sperimentazione, consentono, se del caso in un ambiente reale, di testare tecnologie, prodotti, servizi o approcci innovativi – al momento soprattutto nel contesto della digitalizzazione – per un periodo di tempo limitato e in una parte limitata di un settore o di un ambito soggetto a vigilanza regolamentare, garantendo la messa in atto di opportune misure di salvaguardia*»[125].

Questo tipo di strumento di produzione legislativa è già stato introdotto nell'ordinamento italiano con il d.l. n. 76/2020, conv., con modificazioni, dalla l, n. 120/2020, recante "Misure urgenti per la semplificazione l'innovazione

digitale", attraverso l'art. 36 (rubricato "Misure di semplificazione amministrativa per l'innovazione") agevola la sperimentazione di progetti che non obbedirebbero alla normativa vigente[126]; in particolare, è possibile domandare l'autorizzazione alla sperimentazione di iniziative attinenti all'innovazione tecnologica in deroga temporanea ad una o più norme impeditive dello Stato[127].

L'avvento dell'evoluzione scientifica e tecnologica ha mutato il panorama giuridico e ordinamentale tradizionale probabilmente in maniera irreversibile. In questo contesto, invece che la tradizionale piramide gerarchica delle fonti del diritto ci si ritrova a ragionare di una rete di fonti di provenienza diversa, extra statuale che deve adeguarsi caso per caso alle circostanze fattuali, che se da un lato si riferiscono a realtà settoriali, dall'altro lato provengono da "un arcipelago globale" ultroneo rispetto allo Stato.

Nonostante nella teoria tradizionale sulla gerarchia delle fonti siano presenti aporie e disordini[128], l'introduzione dello "spazio di sperimentazione normativa" rappresenta una ulteriore interferenza "orizzontale" in uno schema che per sua natura è sempre stato verticale. Inoltre, gli effetti della complessità ordinamentale multilivello, insieme alle conseguenze dell'impatto tecnologico sul sistema giuridico, provocano l'affievolimento (se non proprio la perdita) della sanzione quale elemento caratteristico tradizionale della norma di legge[129].

Ulteriormente, l'introduzione di tali spazi di sperimentazione normativa è agevolata dalla lentezza dei procedimenti legislativi parlamentari, rispetto alla velocità dell'evoluzione tanto tecnologica quanto, di conseguenza, sociale perché la tecnica ha ormai un imprescindibile impatto modificativo sulla società e sulla relativa disciplina. Pertanto, apparentemente, sembrerebbe inevitabile che la legislazione non riesca più a regolare le fattispecie future in senso generale ed astratto, ma faccia addirittura fatica a organizzare il presente[130], a causa della sua particolarità e concretezza[131].

Da ciò deriva un impatto sia sui tempi legislativi, sia (e forse soprattutto) una influenza inevitabile, e forse nefasta, sull'organizzazione dello Stato e sull'equilibrio, tradizionalmente ricercato fin dagli albori del costituzionalismo occidentale, della tripartizione dei poteri di ispirazione "montesquieiana", ove la produzione legislativa era nelle mani di un preciso potere pubblico che trova(va) la sua legittimazione nella rappresentanza parlamentare.

Si tratterebbe quindi dell'introduzione di un passaggio da una piramide

verticistica a una rete di fonti orizzontali, nel caso specifico «*mediante la creazione di un ambiente controllato di sperimentazione e prova nella fase di sviluppo e pre-commercializzazione al fine di garantire la conformità dei sistemi di IA innovativi al presente regolamento e ad altre normative pertinenti dell'Unione e degli Stati membri, e il rafforzamento della certezza del diritto per gli innovatori e della sorveglianza e della comprensione da parte delle autorità competenti delle opportunità, dei rischi emergenti e degli impatti dell'uso dell'IA*»[132].

In dottrina, tale passaggio è stato definito quale "diritto proteiforme", che raccoglie nel suo alveo un insieme di fonti di *soft law* inerenti a raccomandazioni, programmi di azioni, codici di condotta, prassi internazionali, libri bianchi e verdi, contrattistica internazionale e così via[133].

In questo contesto va letto l'inserimento dei *sandbox* nel titolo V dell'*AIA Proposal*, secondo cui tali spazi di sperimentazione normativa devono essere istituiti dagli Stati membri, prevedendo una durata limitata di tempo il controllo delle autorità nazionali; in presenza di questi presupposti, diventa legittimamente possibile sperimentare e testare sistemi di intelligenza artificiale innovativi in prospettiva di una loro immissione sul mercato[134].

Tuttavia, va segnalato che durante la fase di trilogo dell'iter di approvazione dell'*AIA Proposal*, il Consiglio Europeo ha manifestato divergenze rispetto alla posizione degli altri due co-legislatori, non trovandosi d'accordo con l'obbligo di stabilire almeno una *sandbox* a favore dell'IA in ciascun Stato Membro, mentre l'istituzione di tale strumento dovrebbe essere semplicemente opzionale.

Questo approccio sembrerebbe essere paradossalmente in contrasto con la forza cogente del Regolamento, una volta ipoteticamente entrato in vigore, poiché l'opzionalità della *sandbox* non perseguirebbe con successo l'obiettivo di rendere omogenea la regolamentazione dell'intelligenza artificiale su un punto rilevante a discapito dell'armonizzazione del mercato interno, dei diritti fondamentali e della protezione dei consumatori.

Il testo del Parlamento europeo prevede che i sistemi di IA che superano positivamente il vaglio di una *sandbox* regolamentare di IA beneficino della presunzione di conformità a requisiti specifici. L'intenzione di questa misura, che non è inclusa nell'approccio generale del Consiglio, è quella di incentivare i fornitori a partecipare alle *sandbox* regolamentari di IA.

Nonostante questa breccia nell'impianto del Regolamento, i co-legislatori parrebbero contraddirsi nel ritenere che la relazione contenente il risultato della *sandbox* di un sistema ad alto rischio debba essere obbligatoriamente contenuta nella dichiarazione di conformità. Qualora l'orientamento sulla non obbligatorietà della *sandbox* ottenesse l'approvazione finale si verificherebbe un deciso vulnus nella coerenza normativa, perché un sistema orientato a prevenire discriminazioni perderebbe la sua cogenza per assumere una valenza di mero marketing[135].

6. Sommarie riflessioni conclusive

L'*Artificial Intelligence Act Proposal* rappresenta il tentativo promosso dalla Commissione europea di conciliare due aspetti che, se lasciati alla disponibilità degli operatori economici, contrastano fortemente tra loro, cioè la tutela dei diritti fondamentali degli utenti e dei consumatori in contrapposizione con le esigenze dell'innovazione, tra le quali il ritorno economico degli enormi investimenti in ricerca e sviluppo e lo sfruttamento di nuovi mercati.

Il *Proposal* predispone una classificazione di rischio degli algoritmi e gradua le cautele nell'uso e nella diffusione sul mercato di tali prodotti sulla base dell'accettabilità o meno di tale rischio in merito alla paventata violazione dei diritti umani. Al contempo esso è il cardine della politica digitale dell'Unione Europea in quanto intende soddisfare «*le aspettative degli imprenditori dell'UE fornendo certezza giuridica all'interno del mercato unico e incoraggiando le istituzioni del settore pubblico e privato a implementare servizi di IA in modo rapido, su larga scala e senza compromettere la sicurezza*»[136].

Ciò è dimostrato dalla circostanza che ampia parte della disciplina, messa in discussione nel trilogo, è dedicata alle *sandbox* e agli spazi normativi in cui gli operatori possono sperimentare anticipatamente e in sicurezza se un sistema algoritmico rispetta tutti i requisiti previsti dal *Proposal* stesso.

In realtà, la predisposizione di *sandbox* e la tutela dei diritti fondamentali sono elementi distintivi e caratterizzanti che costituiscono la forza innovativa del *Proposal*, nonostante il medesimo sia anche didascalico e classificatorio.

Il *Proposal* sembra realizzare l'invocazione faustiana all'attimo di fermarsi[137], sia per godere della sublime bellezza del momento in cui la realizzazione di una

grande opera, come l'IA, è stata compiuta, sia per fissare il bilanciamento tra i diritti, i benefici e vantaggi garantiti dall'uso dell'IA con i doveri, i costi e gli svantaggi legati allo sviluppo delle tecnologie automatizzate.

In tal senso e concentrando l'attenzione sui suoi aspetti positivi, qualora il *Proposal* superasse il vaglio delle negoziazioni del trilogo, esso rappresenterebbe la normativa cogente contenuta nel quadro del diritto dell'Unione Europea. Insieme a ciò esso assumerebbe anche una forza persuasiva quale modello per altri ordinamenti perché fissa quale principio irrinunciabile il rispetto dei diritti fondamentali dell'individuo che interagisce (per qualsiasi motivo) con tali programmi.

In merito allo specifico caso dei programmi predittivi, essi sono considerati efficienti, e incentivati nell'uso nei casi in cui le circostanze fattuali siano seriali, come osservato per i software in materia di calcolo della solvibilità finanziaria. Ciò nonostante, occorre prudenza, perché i software predittivi si caratterizzano per varianza di struttura ovvero di scopi e certuni vengono utilizzati per prevenire reati o crimini, con il rischio che si tenda a prevedere il comportamento umano a detrimento di ogni tutela nei confronti del diritto all'autodeterminazione, del libero arbitrio e della presunzione di innocenza. A questo proposito l'approccio dell'*Artificial Intelligence Act Proposal* orientato alla prudenza e alla cautela nell'autorizzare l'uso dei software predittivi, specie in ambito investigativo o giudiziario, non può che essere condivisibile, perché nessuno può e deve essere condizionato irragionevolmente, e senza adeguata motivazione, dal proprio passato.

1. P.J. Bowler, *A History of the Future: Prophets of Progress from H. G. Wells to Isaac Asimov*, Cambridge University Press, Cambridge, 2017; S. Leslie-McCarthy, *Asimov's Posthuman Pharisees: The Letter of the Law Versus the Spirit of the Law in Isaac Asimov's Robot Novels,* in *Law, Culture and the Humanities*, 3(3), 2007, pp. 398-415.
2. Le Tre leggi della Robotica recitano «*1.Un robot non può recar danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno. 2. Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge. 3. Un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge. Manuale di Robotica, 56ª Edizione – 2058 d.C.*» (così I. Asimov, *Io, Robot*, trad. it. R. Rambelli, Bompiani, Milano, 1963; R.R. Murphy, D.D. Woods, *Beyond Asimov: The three laws of responsible robotics*, in W. Wallach, P. Asaro, *Machine Ethics and Robot Ethics,* Routledge,

London, 2020, pp. 405-411).

3. Rechtbank Deen Haag 5 febbraio 2020, C-09-550982-HA 18-388 (al link https://rechtspraak.nl). Come accaduto nei Paesi Bassi attraverso l'algoritmo "SyRI": si trattava una controversia inerente l'uso di algoritmi nel controllo delle domande di sussidi pubblici da parte di persone che versavano in stato di bisogno. In dottrina, A. Rachovitsa, N. Johann, *The human rights implications of the use of AI in the digital welfare state: Lessons learned from the Dutch SyRI case,* in *Human Rights Law Review*, 22(2), 2022, ngac010; G. Avanzini, *Intelligenza artificiale e nuovi modelli di vigilanza pubblica in Francia e Olanda*, in *Giornale di diritto amministrativo*, 3, 2022, pp. 316-325.
4. Trib. Bologna, 2 gennaio 2021, in *Giur. It.*, 2021, p. 1158; Trib. Palermo, 24 novembre 2020, n. 3570, in *Lavoro nella Giur.*, 2021, 3, p. 318; Trib. Roma, 17 maggio 2023, in *Banca Dati One Legale*. In senso contrario, però Trib. Firenze, 9 febbraio 2021 in *Lavoro nella Giur*., 2021, 6, 664. In dottrina, A. Biagiotti, *Algoritmo discriminatorio - distorsioni e (ab)usi delle piattaforme digitali*, in *Giur. It.*, 5, 2021, pp. 1158 ss.; *S. Renzi, Decisioni automatizzate, analisi predittive e tutela della privacy dei lavoratori* in *Lavoro e diritto*, 3, 2022, pp. 583-603; A. Aloisi, V. De Stefano, *Your Boss in an Algorithm*, Hart, Oxford, 2022; A. Aloisi, V. De Stefano, *Between risk mitigation and labour rights enforcement: Assessing the transatlantic race to govern AI-driven decision-making through a comparative lens,* in *European Labour Law Journal,* 14(2), 2012, pp. 283-307; C. Clifford, J. Goldenfein, A. Jimenez, M. Richardson, *A Right of Social Dialogue on Automated Decision-Making: From Workers' Right to Autonomous Right,* in *Technology and Regulation,* 2023, pp. 1-9.
5. La controversia è sorta con l'approvazione della legge sulla c.d. "Buona Scuola" (l. n. 107/2015), la quale prevedeva un piano straordinario di assuzioni degli insegnanti inseriti nelle "graduatorie ad esaurimento" su base nazionale. La disciplina stabiliva che la modalità di presentazione della domanda fosse esclusivamente telematica, attraverso l'utilizzo di una piattaforma predisposta dal MIUR. La raccolta dei dati, la loro elaborazione e l'attribuzione di una proposta di assunzione a ciascun docente avente diritto era predisposta da un algoritmo relativo ai trasferimenti interprovinciali del personale docente scolastico previsto dal C.C.N.I. sulla mobilità 2016 di cui alla l. n. 107/2015 (D.U. Galetta, G. Pinotti, A*utomation and Algorithmic Decision-Making Systems in the Italian Public Administration*, in *CERIDAP,* 1, 2023, pp. 13-23; M. Sciacca, *Algocrazia e sistema democratico. Alla ricerca di una mite soluzione antropocentrica*, in *Contratto e Impresa*, 2022, pp. 1173 ss; S. Clara, *Prospettive di regolazione della decisione amministrativa algoritmica: un'analisi comparata*, in *Rivista Italiana di Diritto Pubblico Comunitario*, 2022, pp. 265 ss.; E. Carloni, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo,* in *Diritto Amministrativo*, 2020, pp. 271 ss.).
6. T.A.R. Napoli (sezione III), 14 novembre 2022, n. 7003, in *Diritto dell'Informazione e dell'Informatica*, 2023, 1, pp. 91 ss.
7. M.B. Armiento, *Prove di regolazione dell'intelligenza artificiale: il Regolamento della*

Banca d'Italia sulla gestione degli esposti, in *Giornale di diritto amministrativo*, 2023, pp. 105 ss.

8. *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts {sec(2021) 167 final} - {swd(2021) 84 final} - {swd(2021) 85 final}.*
9. Commissione per il mercato interno e la protezione dei consumatori, Commissione per le libertà civili, la giustizia e gli affari interni. PE731.563. Emendamenti del Parlamento europeo, approvati il 14 giugno 2023, alla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).
10. Il documento cui si fa riferimento è il "*Proposal for a Regulation of the European Parliament and of the Councillaying down harmonised rules on artificial intelligence (Artificial IntelligenceAct) and amending certain Union legislative acts- Preparation for the trilogue*".
11. Ulteriori sessioni dedicate al trilogo si sono svolte nell'ottobre 2023 e sono previste nel mese di dicembre 2023 (Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial IntelligenceAct) and amending certain Union legislative acts- Preparation for the trilogue*, 17.10.2023, pp. 2-4).
12. F. Konopczyński, *One Act to Rule Them All: What Is At Stake In the AI Act Trilogue?,* in *VerfBlog*, 2023/8/18, reperibile al link https://verfassungsblog.de/one-act-to-rule-them-all/, DOI: 10.17176/20230818-062853-0.
13. B. Perrigo, *Exclusive: OpenAI Lobbied the E.U. to Water Down AI Regulation*, Time Magazine, 20.6.2023, al link https://time.com/6288245/openai-eu-lobbying-ai-act/. Nello stesso senso si segnala il *OpenAI White Paper on the European Union's Artificial Intelligence Act*, disponibile su https://s3.documentcloud.org/documents/23850240/ares20226851313-openai_aia_white-paper.pdf. A questo proposito si ricorda la controversia sorta tra il Garante della *Privacy* e la nota azienda americana *OpenAI*, titolare di *ChatGPT*. In quell'occasione il Garante dispose, con effetto immediato, la limitazione provvisoria del trattamento dei dati degli utenti italiani nei confronti di *OpenAI* sia per quel che concerneva la raccolta illecita di dati personali, sia per la mancata verifica dell'età minima degli utenti (Garante per la protezione dei dati personali, 30 marzo 2023, 9870832). A seguito di ciò l'azienda bloccò l'accesso ai suoi prodotti agli utenti italiani. La questione si risolse l'adattamento del servizio a disposizione degli utenti rispetto ad una serie di informazioni aggiuntive quali la predisposizione dell'informativa sulla *privacy*, con il riconoscimento del diritto all'opposizione al trattamento dei propri dati personali, ai sensi GDPR, estendendo tali garanzie anche agli utenti extraeuropei (Garante per la protezione dei dati personali, 28

aprile 2023, al link https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9881490). Alla luce di siffatti adempimenti *ChatGPT* è stato reso nuovamente accessibile agli utenti italiani. In dottrina, L. Megale, *Il Garante della privacy contro ChatGPT: quale ruolo per le autorità pubbliche nel bilanciare sostegno all'innovazione e tutela dei diritti?*, in *Giornale di diritto amministrativo*, 2023, pp. 403 ss.

14. Siffatto diritto venne riconosciuto dalla giurisprudenza costituzionale tedesca, la quale stabilì che il potere dell'individuo di decidere autonomamente in merito alla divulgazione e all'utilizzo dei propri dati personali è riconosciuto come diritto fondamentale. Le restrizioni a tale diritto all'"autodeterminazione informativa" sono consentite solo nell'interesse pubblico prevalente (BVerfG, 15 dicembre 1983 - 1 BvR 209/83, 1 BvR 484/83, 1 BvR 440/83, 1 BvR 420/83, 1 BvR 362/83, 1 BvR 269/83). Nella dottrina italiana, si veda S. Rodotà, *Il mondo nella rete. Quali i diritti, quali i vincoli,* Laterza, Bari-Roma, 2014, pp. 31 ss.
15. S. Feldstein, *Evaluating Europe's push to enact AI regulations: how will this influence global norms?*, in *Democratization*, 2023, pp. 1-18; E. Dickhaut, A. Janson, M. Söllner, J.M. Leimeister, *Lawfulness by design–development and evaluation of lawful design patterns to consider legal requirements*, in *European Journal of Information Systems*, 2023, pp. 1-28.
16. E. Campo, A. Martella, L. Ciccarese, *Gli algoritmi come costruzione sociale. Neutralità, potere e opacità*, in E. Campo, A. Martella, L. Ciccarese, (a cura di), *Gli algoritmi come costruzione sociale*, in *LQ The Lab's Quaterly*, 2018, 4, pp. 11 ss.
17. G. Pinotti, *Amministrazione digitale algoritmica e garanzie procedimentali*, in *Labour and Law Issues*, 7, 2021, pp. 80 ss.
18. A.D., Hudson, E. Finn, R. Wylie, *What can science fiction tell us about the future of artificial intelligence policy?,* in *AI & Society*, 2021, pp. 197-211.
19. E. Hearst, *Man and machine: Chess achievements and chess thinking*, in F.W. Frey, *Chess skill in man and machine*, Springer Verlag, Berlin-Heildelberg, 1977, pp. 167-200; F.H. Hsu, *Behind Deep Blue: Building the computer that defeated the world chess champion,* Princeton University Press, 2002.
20. Tremblay et al v. OPENAI, INC. et al., US District Court for the Northern District of California, case number 3:2023cv03223; M. Sag, *Copyright Safety for Generative AI*, in *Houston Law Review*, Vol. 61, No. 2, 2023; A. Strowel, *ChatGPT and Generative AI Tools: Theft of Intellectual Labor?,* in II*C-International Review of Intellectual Property and Competition Law*, 54(4), 2023, 491-494; C. M. Hayes, Hayes, *Generative Artificial Intelligence and Copyright: Both Sides of the Black Box*, 2023, SSRN: https://ssrn.com/abstract=4517799.
21. L. Aničin, M. Stojmenović, *Bias Analysis in Stable Diffusion and MidJourney Models*, in S. Nandan Mohanty, V. Garcia Diaz, G.A.E. Satish Kumar (a cura di) *International Conference on Intelligent Systems and Machine Learning*, 2022, Springer Nature, Cham, pp. 378-388; N. Kenig, J.M Echeverria, A.M. Jives, *Human Beauty according to Artificial Intelligence*, in *Plastic and Reconstructive Surgery Global Open*, 11(7), 2023; A. Jo, *The*

promise and peril of generative AI, in *Nature*, 614(1), 2023, pp. 214-216.

22. F.F. Xu, U. Alon, G. Neubig, V.J. Hellendoorn, *A systematic evaluation of large language models of code,* in *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, 2022,* pp. 1-10, arXiv:2202.13169; N. Carlini et al., *Extracting training data from large language models,* in *30th USENIX Security Symposium (USENIX Security 21)*, 2021, pp. 2633-2650.
23. C. Wei, Y.C. Wang, B. Wang, C.C.J. Kuo, *An overview on language models: Recent developments and outlook,* in *arXiv preprint arXiv:2303.05759,* 2023.
24. R.L. Logan IV, N.F. Liu, M.E. Peters, M. Gardner, S. Singh, *Barack's wife hillary: Using knowledge-graphs for fact-aware language modeling*, in *arXiv:1906.07241*, 2019; E.Volokh, *Large Libel Models? Liability for AI Output*, in *J. FREE SPEECH L.*, 3, 2023, pp. 489-494.
25. P.P. Ray, *ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope*, in *Internet of Things and Cyber-Physical Systems*, 2023; K.T. Gradonm, *Electric Sheep on the Pastures of Disinformation and Targeted Phishing Campaigns: The Security Implications of ChatGPT,* in A. J. Menezes, D. Stebila (a cura di), *IEEE Security & Privacy, 21(3)*, 2023, pp. 58-61.
26. A. Azaria, T. Mitchell, *The Internal State of an LLM Knows When Its Lying*, 2023, arXiv:2304.13734; L. Lian, B. Li, A. Yala, T. Darrell, *LLM-grounded Diffusion: Enhancing Prompt Understanding of Text-to-Image Diffusion Models with Large Language Models*, 2023, arXiv:2305.13655.
27. L.P. Argyle, E.C. Busby, N. Fulda, J.R. Gubler, C. Rytting, D. Wingate, *Out of one, many: Using language models to simulate human samples*, in *Political Analysis*, 31(3), 2023, pp. 337-351.
28. F.F. Xu, U. Alon, G. Neubig, V.J. Hellendoorn, *A systematic evaluation of large language models of code,* cit.
29. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence, cit.
30. C. Colapietro, *La Proposta di Artificial Intelligence Act: quali prospettive per l'Amministrazione digitale?, Public Administration facing the challenges of digitalisation*, in *CERIDAP*, Fasc. Spec. n. 1, 2022, p. 6; L. Arnaudo, R. Pardolesi, *Ecce robot. Sulla responsabilità dei sistemi adulti di intelligenza artificiale*, in *Danno e Responsabilità,* 2023, pp. 409 ss.
31. A. Simoncini, *Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo*, in *BioLaw Journal – Rivista di BioDiritto,* 2, 2021, p. 413.
32. A. Garapon, J. Lassègue, *La giustizia digitale. Determinismo tecnologico e libertà*, trad. it. F. Morini, Il Mulino, Bologna, 2021, *passim*; F. Scamardella, M. Vestoso, *Modelli predittivi a supporto della decisione giudiziaria. Alcuni spunti di riflessione,* in *Rivista di filosofia del diritto*, 12(1), 2023, pp. 135-156; E. Rulli, *Giustizia predittiva, intelligenza artificiale e modelli probabilistici. Chi ha paura degli algoritmi?* in *Analisi giuridica dell'economia,* 17(2), 2018, pp. 533-546.

33. Sotto quest'ultimo profilo, la stabilità socio-culturale è almeno apparente dato che sono i soggetti più vulnerabili a subire gli effetti discriminatori delle decisioni standardizzate (J. Saccomani, *L'impatto della giustizia algoritmica sul diritto all'equo processo*, in *Cass. pen*., 2023, *passim*).
34. Sul punto si rinvia a E. Falletti, *Discriminazione algoritmic*a, Giappichelli, Torino, 2022, *passim*.
35. G.M. Flick, *La sorte degli ultimi. Dalla Bibbia alla Costituzione attraverso la Pandemia,* in *Cass. pen.*, 2022, *passim.*
36. E. Gabellini, *Algoritmi decisionali e processo civile: limiti e prospettive*, in *Rivista trimestrale di diritto e procedura civile,* 2022, pp. 59 ss.
37. R.C. Fong, W.J. Scheirer, D.D. Cox, *Using human brain activity to guide machine learning*, in *Scientific reports,* 8(1), 2018, p. 5397; J. Tang, A. LeBel, S. Jain, et al. S*emantic reconstruction of continuous language from non-invasive brain recordings*, in *Nat Neurosci* 26, 2023, pp. 858–866, reperibile al link https://doi.org/10.1038/s41593-023-01304-9.
38. M. A. Boden, *L'intelligenza artificiale*, Il Mulino, Bologna, 2016, pp. 79 ss.
39. M.L. King, *I Have a Dream*, in T. Golway (a cura di), *American Political Speeches*, Penguin Books, London, 2012; C.B. Jones, S. Connelly, *Behind the Dream: The Making of the Speech that Transformed a Nation*, Palgrave McMillan, London, 2011, *passim.*
40. Aaron Braxton et al., v. Wells Fargo Bank, N.A., et al. Case No. 4:22-cv-01748-KAW, in the U.S. District Court Northern District of California; A. Pope v. Well Fargo Bank. Case No. .4:22-CV-01793-KAWJ in the U.S. District Court Northern District of California; L. Pope v. Wells Fargo Bank N A et al. Case No., 2:2023cv00086, in the US District Court for the District of Utah. In dottrina, W. Arnold, *What Lenders Should Know About AI and Algorithmic Bias, Bloomberg Law Analysis*, in https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-what-lenders-should-know-about-ai-and-algorithmic-bias, 25 aprile 2023; S. Donnan, A. Choi, A. Levitt, C. Cannon, *Wells Fargo Rejected Half Its Black Applicants in Mortgage Refinancing Boom*, Id., 11 marzo 2020, in https://www.bloomberg.com/graphics/2022-wells-fargo-black-home-loan-refinancing/?sref=2XhWEs2V#xj4y7vzkg; J. Edwards, *Wells Fargo Class Action Claims Company Discriminates Against Black Home Mortgage Applican*ts, in https://topclassactions.com/lawsuit-settlements/employment-labor/discrimination/wells-fargo-rejected-more-than-half-of-black-applicants-in-2020/.
41. E. Martinez, L. Kirchner, *Denied. The Secret Bias Hidden in Mortgage-Approval Algorithms*, in *The Markp,* August, 25 2021, in https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms; Id., *How We Investigated Racial Disparities in Federal Mortgage Data*, *ibidem*, su https://themarkup.org/show-your-work/2021/08/25/how-we-investigated-racial-disparities-in-federal-mortgage-data.
42. K. Scanlon, J. Lunde, C. Whitehead, *Responding to the housing and financial crises:*

mortgage lending, mortgage products and government policies, in *International Journal of Housing Policy*, 11(1), 2011, pp. 23-49.

43. C.S. Spatt, *A tale of two crises: The 2008 mortgage meltdown and the 2020 COVID-19 crisis*, in *The Review of Asset Pricing Studies,* 10(4), 2020, pp. 759-790; D. Furceri, P. Loungani, J. D. Ostry, P. Pizzuto, *Will COVID-19 have long-lasting effects on inequality? Evidence from past pandemics*, in *The Journal of Economic Inequality,* 20(4), 2022, pp.811-839.
44. X. van Bruxvoort, M. van Keulen, *Framework for assessing ethical aspects of algorithms and their encompassing socio-technical system,* in *Applied Sciences*, 11(23), 2021 p. 11187.
45. E. Grabham, *Decertifying gender: The challenge of equal pay,* in *Feminist Legal Studie*s, 31(1), 2023, pp. 67-93.
46. N. Irti, *L'età della decodificazione*, Giuffrè, Milano, 1999; Id. *Nichilismo Giuridico*, Laterza, Roma-Bari, 2004; F. Galgano, *Il diritto e le altre arti. Una sfida alla divisione fra le culture,* Compositori edizioni, Bologna, 2009.
47. F. Battaglia, *La trasparenza del procedimento legislativo europeo all'esame del giudice dell'Unione nel caso "De Capitani"*, in *federalismi.it*, 15, 2018, p. 5.
48. Emendamento n. 3 – Considerando 1 (Commissione per il mercato interno e la protezione dei consumatori, Commissione per le libertà civili, la giustizia e gli affari interni. PE731.563. Emendamenti del Parlamento europeo, approvati il 14 giugno 2023, alla proposta di regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).
49. Emendamento 18 – Considerando 16.
50. W. Whitman, *Song of Myself*, in *Leaves of Grass*, Brooklyn, New York, 1855.
51. R.M. Gellert, *The role of the risk-based approach in the General data protection Regulation and in the European Commission's proposed Artificial Intelligence Act. Business as usual?,* in *Journal of Ethics and Legal Technologies*, 2021; J. Schuett, *Risk management in the artificial intelligence act*, in *European Journal of Risk Regulation*, 2023, pp. 1-19.
52. R. Baldwin, J. Black, *Really responsive regulation*, in *The Modern Law Review*, 1, 2008, 65 ss.; A. Simoncini, V*erso la regolamentazione*, cit.
53. A. Simoncini, *op. cit*.
54. A. Simoncini, *ult. op. loc. cit*.
55. I. P. Di Ciommo, *La prospettiva del controllo nell'era dell'Intelligenza Artificiale: alcune osservazioni sul modello Human In The Loop*, in *federalismi.it*, 9, 2023, 79 ss.; D. Messina, *La proposta di regolamento europeo in materiadi Intelligenza Artificiale: verso una "discutibile" tutela individuale di tipo consumer centric nella società dominata dal "pensiero artificiale"*, in *Medialaws - Rivista di diritto dei media*, 2022, pp. 196 ss.
56. Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*, *cit.* p. 2.
57. F. Fedorczyk, *AI legislation in flux: tracking evolving modifications of the AI Act,* in *Diritti Comparati-Blog*, 2023.

58. Emendamento 37 - Considerando 15.
59. Il tema del "social scoring" è molto sentito dal Legislatore europeo, che pertanto ha inteso stigmatizzare «(*I) sistemi di IA che classificano le persone fisiche assegnandole a categorie specifiche, in base a caratteristiche sensibili o protette note o dedotte, sono particolarmente intrusivi, violano la dignità umana e presentano un elevato rischio di discriminazione. Tali caratteristiche includono il genere e l'identità di genere, la razza, l'origine etnica, lo status di cittadinanza o migrazione, l'orientamento politico, l'orientamento sessuale, la religione, la disabilità o qualsiasi altro motivo in base al quale la discriminazione è vietata a norma dell'articolo 21 della Carta dei diritti fondamentali dell'Unione europea, nonché a norma dell'articolo 9 del regolamento (UE) 2016/769. È pertanto opportuno vietare tali sistemi*» (Emendamento 39 – Considerando 16 bis). In Italia, il Garante della *Privacy* ha aperto tre istruttorie sui "meccanismi di scoring che premiano i cittadini "virtuosi" (al link https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9778361, 8 giugno 2022. In dottrina, G. Cerina Feroni, *Intelligenza artificiale e sistemi di "scoring" sociale. Tra distopia e realtà*, in *Il Diritto dell'informazione e dell'informatic*a, 1, 2023, pp. 1-24; A. Vigorito, *Sul crinale tra "data altruism" e "social scoring": esperienze applicative della sequenza dati-algoritmi nel nuovo contesto regolatorio europeo,* in *MediaLaws*, 1, 2023, E. di Carpegna Brivio, *Il "Reputation scoring" e la quantificazione del valore sociale,* in *federalismi.it*, 18, 2022, pp. 119-147.
60. J. Bermudez, R. Nyrup, S. Deterding, C. Mougenot, L. Moradbakhti, F. You, R. Calvo, *What is a subliminal technique?*, in *Ieee Ethics-2023 Conference Proceedings,* 2023.
61. R.J. Neuwirth, *Prohibited artificial intelligence practices in the proposed EU artificial intelligence act (AIA)*, in *Computer Law & Security Review*, 48, 2023, 105798.
62. R.J. Neuwirth, *Law, artificial intelligence, and synaesthesia*, in *AI & SOCIETY*, 2022, pp. 1-12.
63. M.A. Khder, *Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application*, in *International Journal of Advances in Soft Computing & Its Applications,* 13(3), 2021.
64. A. Shapiro, *Reform predictive policing,* in *Nature*, 541(7638), 2017, pp. 458-460.
65. A. Saxena, A. Khanna, D. Gupta, *Emotion recognition and detection methods: A comprehensive survey* in *Journal of Artificial Intelligence and Systems*, 2(1), 2020, pp. 53-79.
66. Questi sono dati aggiuntivi «*ottenuti da un trattamento tecnico specifico in relazione ai segnali fisici, fisiologici o comportamentali di una persona fisica, quali ad esempio le espressioni facciali, i movimenti, la frequenza cardiaca, la voce, la pressione esercitata sui tasti o l'andatura, che possono o non possono consentire o confermare l'identificazione univoca di una persona fisica*» (Emendamento 21 – Considerando 7).
67. La nozione di "identificazione biometrica" utilizzata nel *Proposal* «*dovrebbe essere definita come il riconoscimento automatico di caratteristiche fisiche, fisiologiche, comportamentali e psicologiche di una persona, quali il volto, il movimento degli occhi, le espressioni facciali, la forma del corpo, la voce, il linguaggio, l'andatura, la postura, la frequenza cardiaca, la*

pressione sanguigna, l'odore, la pressione esercitata sui tasti, le reazioni psicologiche (rabbia, angoscia, dolore, ecc.), allo scopo di determinare l'identità di una persona confrontando i suoi dati biometrici con quelli di altri individui memorizzati in una banca dati (identificazione "uno a molti"), indipendentemente dal fatto che la persona abbia fornito il proprio consenso» (Emendamento 22 – Considerando 7*bis*).

68. G. Simonini, *La responsabilità del fabbricante nei prodotti con sistemi di intelligenza artificiale*, in *Danno e Resp.*, 4, 2023, pp. 435 ss.
69. Le caratteristiche relative alle "categorizzazioni" «*sono il genere, il sesso, l'età, il colore dei capelli, il colore degli occhi, i tatuaggi, l'origine etnica o sociale, la salute, l'abilità mentale o fisica, i tratti comportamentali o di personalità, la lingua, la religione o l'appartenenza a una minoranza nazionale o l'orientamento sessuale o politico, sulla base dei loro dati biometrici o dei dati basati su elementi biometrici o che possono essere dedotti da tali dati*» (Emendamento 23 - Considerando 7*ter*).
70. L'Emendamento 24, relativo alla riformulazione del Considerando 8, afferma che «*nel caso dei sistemi "in tempo reale", il rilevamento dei dati biometrici, il loro confronto e l'identificazione del soggetto avvengono tutti istantaneamente o in ogni caso senza ritardi significativi. A tale riguardo, non dovrebbe essere possibile eludere le regole dell'AIA Proposal, seppur con ritardi minimi, dato che tali sistemi comportano l'uso di materiale "dal vivo" generato da una telecamera o da un altro dispositivo con funzionalità analoghe. Al contrario, nel caso dei sistemi di identificazione "a posteriori", invece, i dati biometrici sono già stati rilevati e il confronto e l'identificazione avvengono solo con un ritardo significativo*» (Emendamento 24 – Considerando 8).
71. Si tratta di materiale, come immagini o filmati generati da telecamere a circuito chiuso o da dispositivi privati, che è stato generato «*prima che il sistema fosse usato in relazione alle persone fisiche interessate". Poiché la nozione di identificazione biometrica è indipendente dal consenso della persona, tale definizione si applica anche quando le avvertenze sono collocate nel luogo soggetto alla vigilanza del sistema di identificazione biometrica remota, e non è di fatto annullata dal pre-inserimento*» (Emendamento 24, cit.).
72. C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale,* in *Biolaw Journal – Rivista di Biodiritto*, 3, 2021, 417 ss.
73. Durante la discussione avvenuta nello svolgimento del trilogo i co-legislatori sembrano aver raggiunto un compromesso restringendo le eccezioni all'uso dei programmi di riconoscimento biometrico "real time" riservandoli ai casi di ricerca delle vittime di specifici reati (rapimenti, traffico di esseri umani e sfruttamento sessuale di donne e bambini), prevenzione di imminenti minacce alla vita e integrità fisica delle persone fisiche o di attacchi terroristici e al perseguimento di reati collegati alla criminalità organizzata, terrorismo, traffico di esseri umani, sfruttamento sessuale di donne e bambini, pornografia infantile, traffico di droga e armi, omicidio, traffico di organi, rapimenti, presa d'ostaggi, razzismo e xenofobia, rapina armata, traffico illecito di materiali radioattivi, violenze sessuali). Al contempo, sono state previste ulteriori garanzie nell'uso di tali sistemi di

riconoscimento biometrico in real time, quali la necessità di un'autorizzazione giudiziaria preventiva, la redazione di un rapporto annuale sul loro uso da parte dell'autorità nazionale competente, il monitoraggio da parte della Commissione europea stessa su tali attività (Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence,* cit., pp. 24 ss.).

74. BVerfG, 16.02.2023 - 1 BvR 1547/19 - 1 BvR 2634/20, su http://www.bverfg.de/e/rs20230216_1bvr154719.html. Oggetto dello scrutinio costituzionale era l'uso da parte delle polizie locali dell'Assia e di Amburgo di un software proprietario "Gotham", prodotto dalla *software house* americana "Palantir" che si caratterizza nello sviluppo di modelli capaci di gestire enormi quantità di dati, raccolti in operazioni di sorveglianza massiva ovvero a strascico sulla Rete (M. Meyer, *Daten und Sicherheit*, in *Computer und Recht*, 39(3), 2023, r28-r29).
75. T.E. Frosini, *L'orizzonte giuridico dell'intelligenza artificiale*, in *BioLaw Journal*, 1, 2022, p. 14.
76. T.E. Frosini, *L'orizzonte giuridico*, cit.; I. Di Ciommo, *La prospettiva*, cit.
77. I.P. Di Ciommo, op. cit.
78. C. Casonato, B. Marchetti, *Prime osservazioni,* cit., p. 424.
79. J. Adams-Prassl, *Regulating algorithms at work: Lessons for a 'European approach to artificial intelligence',* in *European Labour Law Journal*, 13(1), 2022, pp. 30-50; F. Sovrano, S. Sapienza, M. Palmirani, F. Vitali, *Metrics, explainability and the European AI act proposal,* in *J,* 5(1), 2022, 126-138; H. Ruschemeier, *AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal, in ERA Forum, Vol. 23, No. 3*, Springer Berlin Heidelberg, 2023, pp. 361-376.
80. I.P. Di Ciommo, op. cit.
81. Rubricato «*Monitoraggio successivo all'immissione sul mercato effettuato dai fornitori e piano di monitoraggio successivo all'immissione sul mercato per i sistemi di IA ad alto rischio*».
82. Rubricato «*Vigilanza del mercato e controllo dei sistemi di IA nel mercato dell'Unione*».
83. A. Adinolfi, *L'intelligenza artificiale tra rischi di violazione dei dirittifondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione,* in A. Pajno, F. Donati, A. Perrucci (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Vol. I, Il Mulino, Bologna, 2022; I. P. Di Ciommo, op. cit.
84. Emendamento 56 – Considerando 28*bis*.
85. Emendamento 56 – Considerando 28*bis*. T. A. Madiega, *Artificial intelligence act. European Parliament,* in *European Parliamentary Research Service,* 2021; M. Van Bekkum, F.Z. Borgesius, *Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?*, in *Computer Law & Security Review*, 48, 2023, p. 105770.
86. A questo proposito si richiama lo specifico intervento dell'Autorità Garante della *Privacy* nel confronto con la società OpenAI per l'accesso al chatbot ChatGPT (P. G. Chiara,

Italian DPA v. OpenAI's ChatGPT: The Reasons Behind the Investigation and the Temporary Limitation to Processing, in *European Data Protection Law Review*, 9(1), 2023, pp. 68-72; F. Panagopoulou, C. Parpoula, K. Karpouzis, *Legal and ethical considerations regarding the use of ChatGPT in education*, in *arXiv:2306.10037,* 2023).

87. Articolo 14, Emendamenti 314-320.
88. Tale articolo tabilisce che i sistemi ad alto rischio devono essere progettati e sviluppati con strumenti di interfaccia uomo-macchina che siano in grado di essere supervisionati da persone fisiche con un grado "sufficiente grado di alfabetizzazione" in materia di intelligenza artificiale che consenta a tale addetto di essere autorevole ed in grado di effettuare indagini, si presume tecniche, approfondite a seguito di un incidente.
89. Ovvero in cui si tratta di condizioni di uso improprio ragionevolmente prevedibile, in particolare quando tali rischi persistono nonostante l'applicazione di altri requisiti di cui al presente capo e qualora le decisioni basate unicamente su un trattamento automatizzato da parte di sistemi di IA producano effetti giuridici o altrimenti significativi sulle persone o sui gruppi di persone sui quali il sistema deve essere utilizzato (Art. 14 *AIA Proposal*).
90. M. Faccioli, *Intelligenza artificiale e responsabilità sanitaria*, in *NGCC*, 2023, 732 ss.; R. Cavallo Perin, *Ragionando come se la digitalizzazione fosse data,* in *Diritto Amministrativo*, 2020, 305 ss.; M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, A.J. Aljaaf, *A systematic review on supervised and unsupervised machine learning algorithms for data science*, in M.W. Berry, A. Mohamed, B. Waa Yap (a cura di), *Supervised and unsupervised learning for data science*, Springer Verlag, Cham, 2020, pp. 3-21.
91. Art. 14 *AIA Proposal*, cit.
92. Per i sistemi di IA ad alto rischio di cui all'*Annex III*, punto 1, lettera a), le misure di cui al paragrafo 3, relative alla identificazione biometrica in tempo reale o a posteriori sono tali da garantire che, inoltre, l'utente non compia azioni o adotti decisioni sulla base dell'identificazione risultante dal sistema, a meno che essa non sia stata verificata e confermata da almeno due persone fisiche che dispongono delle competenze, della formazione e dell'autorità necessarie (Art. 14 *AIA Proposal*, cit.).
93. art. 29 1 *bis AIA Proposal*, cit.
94. Emdendamento 92 – Considerando 58 *bis*.
95. La bibliografia sul tema è molto ampia, tra gli articoli ricostruttivi più interessanti si segnalano: T. Klüwer, *From chatbots to dialog systems*, in D. Perez-Marin, I. Pascual-Nieto, C*onversational agents and natural language interaction: Techniques and Effective Practices,* 2011, IGI Global, Hershey, pp. 1-22; T. Mikolov, G. Zweig, *Context dependent recurrent neural network language model*, in *2012 IEEE Spoken Language Technology Workshop (SLT)*, IEEE, 2012, pp. 234-239; L. Deng, G. Hinton, B. Kingsbury, *New types of deep neural network learning for speech recognition and related applications: An overview*, in 2*013 IEEE international conference on acoustics, speech and signal processing*, pp. 8599-8603; J. Hirschberg, C.D. Manning, *Advances in natural language processing*, in *Science,* 349(6245), 2015 pp. 261-266.

96. E. van Dis, J. Bollen, W. Zuidema, R. van Rooij, C. Bockting, *ChatGPT: five priorities for research,* in *Nature*, 614(7947), 2023, pp. 224-226; D. Sobania, M. Briesch, C. Hanna, J. Petke, *An Analysis of the Automatic Bug Fixing Performance of ChatGPT,* in *arXiv preprint arXiv*, 2023, 2301.08653.
97. L. Manzoni, *Generate Everything. How machine learn to produce text, images, sounds*, presentazione svolta all'interno del panel di E. Falletti, C. Gallese (a cura di), *Regulating misinformation and disinformation. Challenges and issues of Generative Models*, presso la *2023 Society of Legal Scholars, Annual Conference,* in corso di pubblicazione.
98. M. Aboufoul, *Despite Their Feats, Large Language Models Still Haven't Contributed to Linguistics. A review of Chomsky's views on linguistics and LLMs*, in *Towards Data Science*, 5 dicembre 2022.
99. C. Zhou et al., *A Comprehensive Survey on Pretrained Foundation Models: A History from Bert to ChatGPT,* in *arXiv:2302.09419*, 2023; S. Yang et al., *Foundation Models for Decision Making: Problems, Methods, and Opportunities,* in *arXiv:2303.04129*, 2023.
100. Tra gli obblighi a carico dei fornitori di tali sistemi si evidenziano lo specifico *enforcement* della protezione della proprietà intellettuale sui contenuti utilizzati da tali sistemi e l'introduzione sia di un obbligo di trasparenza sui materiali prodotti dai "*foundation model*" sia dell'istituzione di una struttura di supervisione dell'attività di siffatti sistemi (Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*, cit., pp. 18 ss).
101. In riferimento ai modelli più performanti al momento in cui l'AIA proposal entrerà in vigore (Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council,* cit., p. 19).
102. Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council*, cit. p. 18.
103. Altresì obbligando di indicare la presenza di materiali prodotti con i software generativi linguistici ovvero di intelligenza artificiale.
104. M. Foti, *AI Act: con il voto del Parlamento l'UE traccia il futuro dell'Intelligenza Artificiale*, in *Altalex*, 23 giugno 2023.
105. I.P. Di Ciommo, op. cit.
106. T. Wang, Q. Lin, *Hybrid predictive models: When an interpretable model collaborates with a black-box model*, in *The Journal of Machine Learning Research*, 22(1), 2021, pp. 6085-6122.
107. *Proposal for a Regulation of the European Parliament and of the Councillaying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts- Preparation for the trilogue*.
108. *Infra*, §5.
109. Rechtbank Deen Haag 5 febbraio 2020, C-09-550982, cit.
110. BVerfG, 16 febbraio 2023 - 1 BvR 1547/19 - 1 BvR 2634/20, cit.
111. Infine, la Presidenza ritiene che si potrebbe anche prevedere che non vi sia l'obbligo per gli utenti del settore pubblico di tenere una consultazione di sei settimane con le autorità

competenti e/o i rappresentanti delle persone che potrebbero essere interessate dal sistema di IA ad alto rischio o che tale consultazione sia volontaria (*Proposal for a Regulation of the European Parliament and of the Councillaying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts-Preparation for the trilogue*, cit., 6).

112. *EU Trilogues: The AI Act must protect people's rights. A civil society statement on fundamental rights in theEU Artificial Intelligence Act*, disponibile su https://edri.org/wp-content/uploads/2023/07/Civil-society-AI-Act-trilogues-statement.pdf.
113. *AmChamEU, Our Position, Artificial Intelligence Act, Priorities for trilogue*.
114. Sul punto si fa riferimento all'inchiesta del Parlamento olandese, J.F.C. Freriks, R.M. Leijten, F.M. vanKooten-Arissen, R.J. de Bakker, C.J.L. van Dam, R.R. van Aalst, A.J. van Meeuwen, S.Belhaj, A.H. Kuiken, A.C. Verbruggen-Groot, T.M.T. van der Lee, J. van Wijngaarden, M.C.C. van Haeften, W. Bernard-Kesting Verslag - *Parlementaire ondervragingscommissie Kinderopvangtoeslag. Ongekend onrecht*, Den Haag, 20 dicembre 2020. Oltre allo scandalo olandese e riferimento all'inchiesta parlamentare, si segnala la discriminazione subita dai soggetti di diverse etnie nell'ottenimento di un mutuo negli Stati Uniti, vedi supra, par. 2.
115. *Supra*, §2.
116. F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press, Cambridge, 2015, *passim*.
117. Il modello "grey-box" combina l'uso di modelli "*white-box*" basati su principi e i modelli empirici "*black-box*", offrendo particolari vantaggi quando: (a) è carente la teoria fondamentale per descrivere il sistema o il processo modellato; (b) vi è una scarsità di dati sperimentali adeguati per la convalida o (c) vi è la necessità di ridurre la complessità del modello. L'approccio *grey-box* è stato utilizzato, per esempio, per creare modelli matematici (S. Estrada-Flores, I. Merts, B. De Ketelaere, J. Lammertyn, *Development and validation of "grey-box" models for refrigeration applications: A review of key concepts* in *International Journal of Refrigeration*, 2006, 29(6), pp. 931-946; M. E. Khan, F. Khan, *A comparative study of white box, black box and grey box testing techniques,* in *International Journal of Advanced Computer Science and Applications*, 2012, p. 3(6); K. C. Tan, Y. Li, *Grey-box model identification via evolutionary computing* in *Control Engineering Practice,* 10(7), 2002, pp. 673-684; M. Büchi, W. Weck, *The Greybox Approach: When Blackbox Specification Hide too much*, 1999, Turku Centre for Computer Science.) Un siffatto modello può essere definito come un'astrazione concettuale, fisica o matematica di un oggetto, fenomeno, processo o sistema reale. I modelli matematici facilitano la comprensione di un processo o di un sistema, consentendo all'utente di prevederne il comportamento al variare del valore di una o più variabili influenti. L'obiettivo del modellatore può essere quello di stimare o prevedere il valore di una o più variabili che descrivono il comportamento o la proprietà di interesse (note anche come variabili dipendenti) in funzione di una o più variabili indipendenti che descrivono la natura

intrinseca o le proprietà del sistema, il suo funzionamento o il suo ambiente esterno. Un modello può anche prevedere solo parti selezionate del comportamento di un sistema, aiutando a capire come funziona una sottosezione del sistema (S. Estrada-Flores, I. Merts, B. De Ketelaere, J. Lammertyn, *Development and validation of "grey-box" models*, cit.).

118. C. Geiger, *Intellectual property shall be protected!? Article 17 (2) of the Charter of Fundamental Rights of the European Union: a mysterious provision with an unclear scope,* in *EIPR*, 31(3), 2009, pp. 113-117; Id, *Building an ethical framework for intellectual property in the EU: time to revise the Charter of Fundamental Rights,* in *Reforming Intellectual Property*, Edward Elgar Publishing, London, 2022, pp. 77-91; P. Oliver, C. Stothers, *Intellectual property under the Charter: are the Court's scales properly calibrated?* in *Common Market Law Review,* 2017, p. 54(2).
119. P. Craig, *Article 41–Right to Good Administration,* in *The EU Charter of Fundamental Rights,* Nomos Verlagsgesellschaft mbH & Co, 2014, pp. 1112-1141; I. Rabinovici, *The right to be heard in the charter of fundamental rights of the European Union* in *European public law*, 18(1), 2012; D. U. Galetta, *Il diritto ad una buona amministrazione nei procedimenti amministrativi oggi (anche alla luce delle discussioni sull'ambito di applicazione dell'art. 41 della Carta dei diritti UE)*, in *Rivista Italiana di Diritto Pubblico Comunitario*, 2019, pp. 165 ss.; L. Parona, *L'influenza del diritto europeo sulla disciplina dei procedimenti amministrativi nazionali, ibidem*, 2021, pp. 491 ss.
120. Cons. Stato, Sez. VI, 8 aprile 2019, n. 2270.
121. Procedure 2021/0106/COD COM (2021) 206, cit.
122. Cioè del soggetto che si occupa dell'implementazione dell'algoritmo in un ambiente operativo (J. Adams-Prassl, R. Binns, A. Kelly-Lyth, *Directly discriminatory algorithms* in *The Modern Law Review*, 86(1), 2023, p. 151.
123. Nel testo approvato dal Parlamento Europeo, l'articolo 68 *quater* prevede altri due commi, cioè «*2. Il paragrafo 1 non si applica all'uso di sistemi di IA per i quali il diritto dell'Unione o nazionale preveda eccezioni o limitazioni all'obbligo di cui al paragrafo 1 nella misura in cui tali eccezioni o limitazioni rispettino l'essenza dei diritti e delle libertà fondamentali e siano una misura necessaria e proporzionata in una società democratica. 3. Il presente articolo si applica fatti salvi gli articoli 13, 14, 15, e 22 del regolamento 2016/679*».
124. Consiglio dell'Unione Europea, Conclusioni del Consiglio sugli spazi di sperimentazione normativa e le clausole di sperimentazione come strumenti per un quadro normativo favorevole all'innovazione, adeguato alle esigenze future e resiliente che sia in grado di affrontare le sfide epocali nell'era digitale, 16 novembre 2020, su https://data.consilium.europa.eu/doc/document/ST-13026-2020-INIT/it/pdf. In dottrina, A. Merlino, *Il regulatory sandbox e la teoria delle fonti*, in *Diritto Pubblico Europeo Rassegna Online,* 1, 2022, pp. 116 ss.; M. Trapani, *L'utilizzo delle sandboxes normative: una ricognizione comparata delle principali esperienze di tecniche di produzione normativa sperimentali e il loro impatto sull'ordinamento*, in *osservatorio sulle fonti*, 3, 2022, pp. 217 ss.; M. T. Paracampo, *Il percorso evolutivo ed espansivo delleregulatory*

sandboxes da FinTech ai nuovi lidioperativi del prossimo futuro, in *federalismi*, 18, 2022, pp. 209 ss.

125. Commissione europea, STRUMENTO #21. Ricerca e innovazione, Strumenti per legiferare meglio; Commissione europea; 6783/20 (COM (2020)103); Consiglio dell'Unione Europea, Conclusioni, cit.
126. A questo proposito, l'11 aprile 2023 presso la Camera dei Deputati è stato presentato un disegno di legge volto a regolamentare un'utilizzo più ampio, si tratta dell' AC1084, Centemero ed altri. "Disposizioni concernenti l'adozione di una disciplina temporanea per la sperimentazione dell'impiego di sistemi di intelligenza artificiale" (1084).
127. A. Chiappini, *Regulatory Sandbox italiano in una prospettivaeuropea, costituzionale ed amministrativa*, in *Amministrativamente*, 2021, pp. 885 ss.
128. A. Merlino, *Il regulatory sandbox*, cit.; V. Frosini, *Ordine e disordine nel diritto,* Guida, Napoli, 1979, *passim*; M.G. Losano, *Diritto turbolento. Alla ricerca di nuovi paradigmi nei rapporti fra diritti nazionali e normative sovrastatali*, in *Rivista internazionale di filosofia del diritto*, 2005, *passim*.; G. Zagrebelsky, *Il diritto mite*, Einaudi, Torino, 1992, *passim.*
129. H. Kelsen, *Teoria generale del diritto e dello Stato*, trad. it. S. Cotta, G. Treves, Edizioni di Comunità, Milano, 1952, *passim*.
130. M.R. Ferrarese, *Il diritto al presente. Globalizzazione e tempo delle istituzioni,* Il Mulino Bologna, 2002, *passim*.
131. A. Merlino, *op. cit.*
132. Considerando n. 72, cit.
133. A. Algostino, *Diritto proteiforme e conflitto sul diritto. Studio sulla trasformazione delle fonti del diritto*, Giappichelli, Torino, 2018, pp. 184-185; A. Merlino, op. cit.
134. Sul punto, il Considerando 72 del *Proposal* afferma che: «(G)*li obiettivi degli spazi di sperimentazione normativa dovrebbero essere la promozione dell'innovazione in materia di IA, mediante la creazione di un ambiente controllato di sperimentazione e prova nella fase di sviluppo e pre-commercializzazione al fine di garantire la conformità dei sistemi di IA innovativi al presente regolamento e ad altre normative pertinenti dell'Unione e degli Stati membri, e il rafforzamento della certezza del diritto per gli innovatori e della sorveglianza e della comprensione da parte delle autorità competenti delle opportunità, dei rischi emergenti e degli impatti dell'uso dell'IA, nonché l'accelerazione dell'accesso ai mercati, anche mediante l'eliminazione degli ostacoli per le piccole e medie imprese (PMI) e le start-up. Al fine di garantire un'attuazione uniforme in tutta l'Unione ed economie di scala, è opportuno stabilire regole comuni per l'attuazione degli spazi di sperimentazione normativa e un quadro per la cooperazione tra le autorità competenti coinvolte nel controllo degli spazi di sperimentazione. Il presente regolamento dovrebbe fornire la base giuridica per l'utilizzo dei dati personali raccolti per altre finalità ai fini dello sviluppo di determinati sistemi di IA di interesse pubblico nell'ambito dello spazio di sperimentazione normativa per l'IA, in linea con l'articolo 6, paragrafo 4, del regolamento (UE) 2016/679, e con l'articolo 6 del regolamento (UE) 2018/1725, e fatto salvo l'articolo 4, paragrafo 2, della direttiva (UE)*

2016/680. I partecipanti allo spazio di sperimentazione dovrebbero fornire garanzie adeguate e cooperare con le autorità competenti, anche seguendo i loro orientamenti e agendo rapidamente e in buona fede per attenuare eventuali rischi elevati per la sicurezza e i diritti fondamentali che possono emergere durante lo sviluppo e la sperimentazione nello spazio sopraindicato. È opportuno che le autorità competenti, nel decidere se infliggere una sanzione amministrativa pecuniaria a norma dell'articolo 83, paragrafo 2, del regolamento 2016/679 e dell'articolo 57 della direttiva 2016/680, tengano conto della condotta dei partecipanti allo spazio di sperimentazione».

135. Nel corso del trilogo, i co-legislatori hanno raggiunto una proposta di compromesso sull'utilizzo di testing dei sistemi AI ad alto rischio in condizioni reali, al di fuori delle aree regolamentate dall'AI. Le modifiche proposte includono l'approvazione da parte dell'autorità di vigilanza del mercato per i test in condizioni reali, una registrazione dei test in alcune aree sensibili, una durata massima di sei mesi per i test, il requisito di giustificare i test senza il consenso delle persone coinvolte, il diritto delle persone coinvolte di cancellare i loro dati e il potere delle autorità di vigilanza del mercato di richiedere informazioni e condurre ispezioni. Il punto politico rilevante concerne se gli Stati membri sono d'accordo nell'aggiungere queste misure di salvaguardia ai test dei sistemi AI ad alto rischio in condizioni reali (Council of the European Union, *Proposal for a Regulation of the European Parliament,* cit., pp. 3 ss.).
136. F. Konopczyński, *One Act to Rule Them All,* cit.
137. «*All'attimo direi: sei così bello, fermati!*» («*Zum Augenblicke dürft ich sagen:/Verweile doch, du bist so schön*») (così J.W. von Goethe, *Faust. Urfaust*, trad. a cura di A. Casalegno, vol. II, Garzanti, Milano, 2004, pp. 1040-1041).