Digeat N.3 - 19 Settembre 2024

Fiducia nelle istituzioni: l'AI Act e i diritti fondamentali, narrazione o realtà?

Di Enrico Pelino

Abstract

L'AI Act costituisce nella vulgata istituzionale la più avanzata disciplina giuridica del momento, anche in termini di garanzie. Viene raccontato come un avamposto di tutela della persona e, al tempo stesso, un sofisticato congegno normativo per dare sostegno e regole alle applicazioni, dilaganti, dell'intelligenza artificiale. Ma è davvero così? L'articolo esplora alcuni evidenti limiti del nuovo regolamento e avanza dubbi sul preteso approccio antropocentrico di cui sarebbe espressione. In realtà, ad avviso di chi scrive, l'AI Act segna addirittura, sotto alcuni profili, un arretramento dei diritti del cittadino digitale e incrina quindi l'aspettativa di protezione che costituisce il presupposto del rapporto di fiducia tra cittadini e istituzioni.

Indice

- Il confronto con il GDPR
- La compressione del diritto alla spiegazione
- Antropocentrico… sì, ma senza anthropos
- Conclusioni

"*Lo **scopo** del presente regolamento è migliorare il funzionamento del mercato interno e **promuovere** la diffusione di un'intelligenza artificiale (IA) **antropocentrica e affidabile**, **garantendo** nel contempo un livello elevato di protezione della salute, della sicurezza e dei **diritti fondamentali** sanciti dalla Carta dei diritti fondamentali dell'Unione europea…*": esordiscono in questo modo sia il primo considerando sia il primo articolo del nuovo regolamento dell'Unione europea 2024/1689, cd. "AI Act" o "AIA".

Lo scopo è segnato con rigore: **l'uomo al centro, l'enfasi sui diritti fondamentali**. Ne segue l'attesa legittima – se le parole hanno un senso – di un atto normativo ad approccio "costituzionale", costruito, cioè, non solo pensando alle esigenze di mercato interno e alla sua facilitazione, ma **introducendo anche uno statuto dei diritti della persona**, ossia un'architettura di contrappesi, in chiave antropocentrica per l'appunto, che bilanci l'incidenza enorme del potere tecnologico.

> Si potrebbe obiettare: perché c'è bisogno di un arricchimento dei diritti o di tutele addirittura completamente nuove, non basta un richiamo all'esistente? La risposta è semplice: se le sfide sono inedite, e fino a qualche anno fa neppure immaginabili, devono esserlo anche le garanzie.

Si pensi, per citare soltanto alcune novità, ai sistemi di **inferenza di emozioni e intenzioni**, implementati su larga scala e pervasivamente, attraverso *machine learning* o ai sistemi di

categorizzazione biometrica o di ***social scoring***. È chiaro che si attingono qui territori giuridici inesplorati, rispetto ai quali il diritto alla libertà e alla dignità umana costituiscono riferimenti opportuni, ma insieme sfocati, se non vengono declinati in modo più specifico.

Ora dobbiamo verificare se l'AI Act contenga davvero **l'auspicato decalogo di diritti avanzati che illuminano la nuova era del rapporto tra uomo e macchina**, esercitando una spinta uguale e contraria all'incidenza dei sistemi di intelligenza artificiale sulla persona, sulla sua socialità, sul suo pensiero, sulle sue fragilità oppure quelle di esordio siano soltanto "*parole, parole, vuote parole; niente dal cuore*"[1]?

Il confronto con il GDPR

Quando va alla ricerca di una legge dal taglio "costituzionale", il giurista digitale non ha dubbi, guarda al GDPR. Del resto, il regolamento generale sui dati personali, nei limiti del suo oggetto, **costituisce un riferimento normativo imprescindibile anche in tema di intelligenza artificiale**; peraltro, ha fornito la matrice per il testo di alcune disposizioni dell'AI Act.

Bene, anche **il GDPR esordisce con una premessa dal sapore costituzionale** "*the processing of personal data should be designed to serve mankind*" (cons. 4), che è anche una promessa mantenuta, se si considera che la tessitura giuridica della legge sviluppa senza alcun dubbio un solido guscio protettivo attorno all'individuo.

Si pensi al ricco elenco di diritti agli **articoli dal 12 al 22 e ancora all'art. 34**, alle tre disposizioni sui mezzi di ricorso, **artt. 77-79**, alla norma speciale sul risarcimento del danno, **l'art. 8**2, giacché i diritti sono vuoti senza mezzi di ricorso e senza ristoro del danno. Una serie di altre tutele essenziali della persona è inoltre desumibile dall'intero impianto normativo del regolamento del 2016, innanzitutto dagli **artt. 5 e 25**, dalla delimitazione di basi tassative agli **artt. 6 e 9**, dall'introduzione di presìdi come il **DPO**, giusto per scorrere rapidamente e a macchia di leopardo alcune delle norme di spicco.

Tuttavia, quando ci si sposta sull'AI Act per tracciare un confronto, si resta oltremodo perplessi.

Per prima cosa, **è irragionevolmente sparito** dall'articolato finale l'emendamento del Parlamento UE (art. 4 a) recante il fondamentale (anche per la vicenda istituzionale da cui origina) elenco dei principi generali applicabili a tutti i sistemi di intelligenza artificiale, che troviamo retrocesso a elenco di principi etici non vincolanti, cfr. considerando 27 AIA, i quali fungeranno – almeno questo – da base per l'elaborazione di codici di condotta. È vero che alcuni di essi riaffiorano isolatamente in altre disposizioni, ma **resta la constatazione che abbiamo un atto privo di un articolo dedicato a raccogliere unitariamente i principi vincolanti**, **e non appare buon segno, almeno non in prospettiva antropocentrica**. La scelta ha del resto una sua logica: i principi pongono altrettanti limiti alla libertà di azione e di impresa e offrono spiacevoli spunti in sede contenziosa, come insegna la ricca giurisprudenza della Corte di giustizia sull'art. 5 GDPR.

Procedendo nella disamina, scopriamo che sono riconosciuti alla persona **due soli diritti**, quello di reclamo all'autorità di vigilanza del mercato e quello alla spiegazione della decisione automatizzata. Entrambi sfortunatamente ricevono una disciplina molto circoscritta.

La disposizione sul reclamo, **art. 85 AIA,** pur utilizzando nella parte iniziale l'ottimo calco dell'**art. 77 GDPR**, se ne distanzia poi in maniera alquanto deludente. Il reclamo è trattato, significativamente, come un reclamo sulla conformità dei prodotti, ai sensi del Reg. (UE) 2019/1020, non è assistito dalla fissazione, a livello unionale, di termini minimi di riscontro, come i tre mesi contemplati nel GDPR, da informazioni sull'esito e sull'impugnazione o da regole di riparto di giurisdizione, né figura una disciplina *ad hoc* relativa a reclami che abbiano natura transfrontaliera. **Manca la precisazione del diritto di essere ascoltati in contraddittorio** (questo invero anche nel GDPR). Soprattutto, è a **ssente una chiara indicazione sulla reclamabilità** da parte dei gruppi o relativa alla lesione di interessi di gruppo, nonostante il riferimento al "gruppo" come formazione sociale a rischio di

discriminazione ricorra nel regolamento sull'intelligenza artificiale.

> Certo, molti di questi elementi possono essere recuperati da altri istituti del sistema giuridico, attraverso opera di paziente cucitura, ma una norma autosufficiente avrebbe notevolmente giovato, anche in termini di uniformità nell'Unione e di facilitazione interpretativa. Chi scrive tende ad avvertire "antropocentrico" in questo modo.

Manca altresì il **riconoscimento di un diritto alla tutela giurisdizionale**, **clamorosa assenza**, come **manca una disposizione dedicata all'impugnazione** delle decisioni dell'autorità di vigilanza del mercato. Tale rimedio, che pure era stato introdotto dal Parlamento UE nel pacchetto di emendamenti del 2023, è poi sparito dal testo finale, senza clamori nella pur copiosa e alquanto ripetitiva (mi sia permesso) fioritura dottrinale dell'ultima stagione.

Anche in questo caso, si supplisce attingendo da altre fonti, come **l'art. 19 TUE e l'art. 47 CDFUE**, tuttavia, l'omissione, ancorché rimediabile attraverso norme di sistema, o talora attraverso discipline di settore quali la **dir. UE 2020/1828**, appare – proprio per questo – **eloquente dell'attenzione realmente riservata alla persona fisica e ai gruppi nella nuova legge sull'intelligenza artificiale**. Siamo davvero ben distanti dalla logica, e dall'utilità concreta, degli articoli 78 e 79 GDPR. Peraltro, la presenza di disposizioni omologhe nell'AI Act avrebbe facilitato lo sviluppo di pronunce della Corte di giustizia, come è avvenuto in materia di dati personali.

Ugualmente, **non si trova alcun riferimento al risarcimento del dann**o, specialmente il danno da discriminazione, insidiosissimo e di difficile dimostrazione, che avrebbe richiesto meccanismi di inversione dell'onere della prova e altre facilitazioni. Non a caso, per trovare disposizioni mirate al contesto in esame occorre ricorrere ad altri strumenti giuridici, il cui principale, la **AILD (AI Liability Directive)** si trova allo stato di mera proposta, non è in ogni caso un regolamento, e appare, al momento in cui si scrive, in fase di evidente stallo nell'iter legislativo. Allo stato della normativa applicabile, **il danno è ovviamente risarcibile** secondo le ordinarie regole civilistiche oppure, ove applicabile, utilizzando il regime di favore dell'**art. 82 GDPR**, scontentando tuttavia chi auspicava che la peculiare invasività delle applicazioni di intelligenza artificiale dovesse trovare espressione, già nel contesto del nuovo regolamento dedicato, in una disciplina risarcitoria altrettanto peculiare per persone e gruppi.

La compressione del diritto alla spiegazione

Quanto al **diritto alla spiegazione**, presidio imprescindibile in tema di intelligenza artificiale, cominciamo con l'osservare che è riconosciuto (ma soltanto a seguito di un emendamento del Parlamento UE), limitatamente ai soli sistemi ad alto rischio di cui all'Allegato III.

> Ne risultano perciò non solo irragionevolmente esclusi quelli, di pari rischio, individuati in base ai criteri di cui al primo paragrafo dell'art. 6, ma anche, a monte, le pratiche proibite elencate all'art. 5, come se non sussistesse anche in tali casi la legittima pretesa di comprenderne la logica, non foss'altro in occasione di contenzioso. Sono altresì esclusi, a valle, i modelli per finalità generale (General Purpose AI). È una limitazione di cui sfugge la *ratio*: perché non assicurare un diritto generale alla spiegazione?

Giova qui osservare che il diritto alla spiegazione è una valvola vitale del sistema, è cioè **fondamentale** per superare il primo e più ovvio ostacolo dinanzi ai risultati prodotti dai sistemi di intelligenza artificiale, **l'opacità**. La spiegazione costituisce altresì un precursore per l'esercizio di altri diritti, tra cui quello di difesa e quello di contestare i risultati automatici, che, diversamente, assumerebbero un significato oracolare. Come infatti rilevato dalla Corte di giustizia nell'arresto del 21 giugno 2022, Ligue des droits humains, C?817/19, punto 195: "*Tenuto conto dell'opacità che*

caratterizza il funzionamento delle tecnologie di intelligenza artificiale, può risultare impossibile comprendere la ragione per la quale un dato programma sia arrivato ad un riscontro positivo. In tali circostanze, l'uso di siffatte tecnologie potrebbe privare gli interessati anche del loro diritto a un ricorso giurisdizionale effettivo sancito dall'articolo 47 della Carta". Va altresì ricordata la pregevole giurisprudenza sviluppata dal Consiglio di Stato, a partire da clausole generali, in tema di spiegabilità in ambito pubblico delle decisioni automatizzate, es. *ex multis* sez. VI, sent. 4 febbraio 2020, n. 881.

> L'obiezione che molti sistemi non siano spiegabili è un dato di fatto, non di diritto. Inserire l'obbligo di una spiegazione incide sui criteri di progettazione.

Va poi notato che nell'AI Act **l'istituto in parola è limitato ai soli risultati qualificabili come "decisioni"**, termine non definito e dunque dai lembi concettuali incerti. A parere di chi scrive, è particolarmente scivoloso **il confine tra "decisione" e "raccomandazione"**, parola presente anch'essa nell'AI Act e altrettanto non definita, e perplessità simili possono essere estese anche alla "previsione", in considerazione del fenomeno ben noto dell'*automation bias*, vale a dire **il condizionamento subito dal decisore umano rispetto al risultato della macchina**, anche quando la stessa si limita a raccomandare o perfino a restituire un contenuto. In tal caso potrebbe risultare utile, *mutatis mutandis*, un riferimento allo storico della Corte di giustizia, penso alla sentenza 7 dicembre 2023, Schufa Holding, C?634/21.Infine, il diritto alla spiegazione può essere esercitato solo nei confronti del *deployer*. Vero che si tratta dell'anello più prossimo all'interessato, ma non si comprende il perché, in definitiva, di una limitazione a risalire la filiera.

Ora, dobbiamo chiederci **se un diritto alla spiegazione così circoscritto aggiunga davvero qualcosa alle garanzie già sussistenti**, ossia all'analogo diritto già previsto dal GDPR oppure a quanto desumibile, con ampiezza ancora maggiore, dai principi generali del diritto.

> Si può dubitarne: il diritto alla spiegazione[2] riconosciuto nel regolamento sui dati personali sembra, a una prima analisi, intercettare un'area di maggiore ampiezza (quantomeno rispetto alle persone fisiche), riguardando tutti processi decisionali unicamente automatizzati, ambito assai maggiore di quello dell'AI. Inoltre, non pone limiti connessi al livello di rischio, include la pretesa di essere preventivamente informati sull'importanza e le conseguenze per l'interessato ed è accompagnato dal riconoscimento, sia pure con alcuni limiti, di un connesso diritto alla contestazione della decisione.

Otto anni più tardi, e dinanzi a fenomeni esponenzialmente più invasivi, si sarebbe dovuti partire da questa base e allargare, anziché restringere, l'area di tutela. Addirittura, l'attuale art. 86 AIA non prevede neppure che siano fornite spiegazioni in relazione **ai dati di input**, come era invece previsto nell'emendamento originario del Parlamento UE.

In definitiva, la disciplina che dovrebbe segnare la nuova era del rapporto tra uomo e macchina fallisce sul più essenziale e più umano dei diritti.

Antropocentrico… sì, ma senza *anthropos*

Ancora più imbarazzante è, a ben vedere, che **il protagonista, l'uomo, non abbia trovato spazio nel pur lunghissimo elenco di definizioni di cui all'art. 3 AIA**. Le definizioni in un atto normativo europeo, e soprattutto in un atto con questo oggetto e con queste ambizioni, **sono i blocchi concettuali** della successiva costruzione. Ebbene, all'art. 3 troviamo definiti tutti gli attori, ci sono il fornitore, il deployer, il distributore, l'importatore, ma, paradossalmente, **manca una voce dedicata alla persona fisica o al gruppo di persone che subisce l'impatto delle applicazioni intelligenza artificiale**. Eppure, approccio antropocentrico vuol dire, si converrà, che la persona e i gruppi sono protagonisti.

Non si tratta di un formalismo, così come nel GDPR la nozione di “interessato” non è un formalismo, e non è un caso che il Parlamento UE avesse adottato un emendamento apposito diretto a introdurre la definizione di “*affected person*”, caldeggiato autorevolmente nell’opinione congiunta n. 5/2021 da EDPB ed EDPS, vale a dire dai massimi riferimenti istituzionali in materia di protezione dei dati personali e nuovamente dall’EDPS ancora nell’opinione n. 44/2023. L’espunzione dell’emendamento nel testo finale appare perciò rispondere a un calcolo deliberato.

> In buona sostanza, oggi per definire le *affected person* o i gruppi discriminati occorre ricorrere a perifrasi oppure utilizzare, atecnicamente e con imprecisione, il termine “interessato”, come leggiamo per esempio all’art. 2.1.g), al 50 o all’art. 86 AIA. La realtà è che il focus dell’AI Act è, ed è sempre stato, al di là di ipocrisie di facciata, sul mercato, sulla *safety* e sulla *security*, non sui diritti fondamentali. L’operazione finale approdata in Gazzetta resta in ogni caso, a parere di chi scrive, spiccatamente mercato-centrica.

Anche così, tuttavia, **qualche forma di tutela alla persona indirettamente perviene**. Il riferimento più ovvio è all’art. 5, che elenca le pratiche di AI proibite. Da tali divieti emergono, *a contrario*, altrettanti diritti, ossia pretese giuridicamente fondate di non essere fatti oggetto delle pratiche vietate e di poterlo contestare, attraverso strumenti di ricorso da cercare non di rado altrove nel sistema giuridico.

È però un fatto che **l’elenco delle pratiche vietate – per lunghissimo tempo oggetto di contesa politica nel corso dell’*iter* normativo – risulti assai limitato**, falcidiato da numerose eccezioni e soprattutto permeabile, dato che spesso a formulazioni ampie e astratte si sono preferite formulazioni precise e alquanto strutturate. In definitiva, l’articolo 5, se da un lato introduce opportuni divieti, dall’altro fa filtrare molto, finendo per autorizzare (almeno in prima battuta) ciò che non è puntualmente vietato, incluse per esempio alcune forme di *social scoring*.

Facciamo un esempio per chiarirci, riprendendo un altro dei temi di apertura di questo articolo, quello dell’**inferenza di emozioni e intenzioni**. Il legislatore europeo rivela piena consapevolezza della dubbiosa affidabilità di tali tecniche: “*Tra le principali carenze di tali sistemi figurano la **limitata affidabilità***” (considerando 44), dovuta anche alla circostanza che “*l’espressione delle emozioni varia notevolmente in base alle culture e alle situazioni e persino in relazione **a una stessa persona***”.

Nondimeno, e con una certa incoerenza, limita il divieto di tali sistemi ai soli settori dell’istruzione e dell’occupazione, qualificando le altre applicazioni come trattamento solo ad alto rischio, ma perfettamente lecito. **Ne viene normalizzata con ciò la percezione**. La disciplina dei trattamenti ad alto rischio è, sì, presidiata da una serie di norme di dettaglio, ma non appare, in tutta onestà, eccessivamente stringente. Sono state superate diverse linee rosse, culturali e giuridiche.

> L’obiezione maggiore a tali pratiche non sta tanto, o comunque non solo, nella precisione (magari perfettibile), ma nella sostanza, vale a dire nel fatto che determinano l’invasione di un luogo sacro, quello della pace mentale. In definitiva, infatti, attraverso le espressioni si cerca di catturare, sia pure dall’esterno, le reazioni che avvengono nel chiuso della mente. Non è chi non colga il significato giuridico e culturale profondo del normalizzare queste pratiche. Vuol dire varcare una frontiera, viene sfondata l’ultima parete a protezione dei neurodiritti, un tabù, oltre il quale esistono solo forme odiose di compressione dell’unico luogo rimasto ormai privato, quello chiuso nella scatola cranica. Possiamo facilmente immaginare uno scenario distopico in cui chi vuole tutelarsi dalla lettura delle emozioni è costretto a non mostrarle più.

Ciò non vuol dire che non possano essere ravvisati divieti all'inferenza emotiva, ma in larga misura vanno cercati all'esterno dell'AI Act. E tuttavia, proprio la circostanza che il giurista sia costretto a volgersi altrove rivela il sostanziale fallimento, sul piano umano, della legge che ci occupa.

Conclusioni

Tolto l'elenco delle pratiche vietate per rischio inaccettabile, resta una manciata, assai modesta, di garanzie, soprattutto nel caso di trattamenti ad alto rischio, quali la FRIA (*Fundamental Rights Impact Assessment*), comunque limitata quanto ai soggetti obbligati; il circoscritto diritto alla spiegazione già toccato; e poco altro. Anche gli obblighi di informazione rispetto ai sistemi di AI destinati a interagire con le persone fisiche sono, in tutta franchezza, assai basici, ossia lontani dalla ricchezza informativa desiderabile.

Prescindono dai limiti di questo articolo riflessioni più ampie e più analitiche, tuttavia già dai pochi cenni sviluppati emerge, credo, un quadro fortemente carente rispetto sia alle premesse sia alle legittime attese che la prima legge dell'Unione dedicata all'intelligenza artificiale autorizzava.

Affiorano in definitiva **contraddizioni lampanti** rispetto alla vantata centralità dell'essere umano. L'impressione è rafforzata dalla constatazione che la maggior parte dei diritti e delle tutele va infatti cercata altrove, nel sistema giuridico complessivamente inteso e nelle *leges speciales* fatte salve.

> In definitiva, il regolamento più recente e più atteso, anziché tesaurizzare dal passato ed eccellere in quello che fino a qualche tempo addietro era stato il primato dell'Unione, ossia un'attenzione ai diritti ben calibrata da meccanismi fini di bilanciamento, espone sul piano strettamente umano, a parere di chi scrive, una soluzione di modesto profilo, a ribasso, titubante: non andava scontentata un'industria che si stenta a inseguire.

Certo, **le lacune potranno essere colmate in via interpretativa**, con percorsi incerti e onerosi, o potranno essere riempite da nuovi provvedimenti normativi più puntuali, ma la sensazione è che l'effetto "Bruxelles", ossia la forza trainante che trasforma gli atti normativi dell'Unione in modelli mondiali da imitare, sarà qui alquanto contenuto. L'effetto "Bruxelles" in passato poggiava infatti non solo sulla portata innovatrice, sulla chiarezza degli istituti giuridici, ma altresì sulla forza etica e sulla rigorosa tutela della persona, sul coraggio, in altre parole, di disturbare centri di potere, nella convinzione che una società più equilibrata è più sana.

Non che non siano contenute norme di pregio nell'AI Act, come l'art. 10 sulla *governance* dei dati, l'art. 14 che impone una progettazione dei sistemi ad alto rischio che permetta supervisione umana o l'art. 15, volto a precisare parametri minimi di *cybersecurity*, per portare solo alcuni esempi, tuttavia, **l'impressione è che laddove nel mondo le istanze antropocentriche saranno avvertite con intensità appare assai dubbio che l'esempio a cui guardare sarà rappresentato dal nuovo regolamento sull'intelligenza artificiale**.

Ciò osservato, l'AI Act riesce probabilmente più efficace nello scopo pratico di facilitazione del mercato interno, ancorché si avverta anche in questo caso l'impressione di un'impronta più burocratica che sostanziale in molti dei processi introdotti.

NOTE

[1] Shakespeare, Troilo e Cressida, V.3, trad. C.V. Lodovici, Torino, 1974.

[2] Ritiene lo scrivente che gli istituti richiamati agli artt. 13.2.f), 14.2.g), 15.1.h) e 22.3 GDPR vadano letti alla luce del "diritto… di ottenere una spiegazione della decisione" ricordato al considerando 71 GDPR.