UNIVERSITA' DEGLI STUDI DI ROMA TOR VERGATA

***European Digital Law of the Person, of the Contract and of the Technological Marketplace* - EUDILA**
Cattedra Jean Monnet del Progetto ERASMUS +

IL DIFFICILE EQUILIBRIO TRA AUTOMAZIONE E RESPONSABILITÀ

Diego Maria Napoleoni

0367119

Anno accademico 2024/2025

Indice

Introduzione

Dalla selezione automatica dei contenuti sui social, alla traduzione istantanea di un messaggio, dal riconoscimento facciale per sbloccare uno smartphone, fino alla diagnosi precoce di una malattia rara: l'intelligenza artificiale è ormai ovunque. È capace di guidare veicoli senza conducente, suggerire sentenze nei tribunali, decidere se concedere un prestito o segnalare comportamenti sospetti alle autorità. E non si fermerà qui.

Le sue potenzialità evolvono a un ritmo impressionante: secondo delle stime, nei prossimi anni i sistemi IA saranno in grado di scrivere codici complessi, gestire interi processi decisionali aziendali e persino simulare relazioni empatiche nel supporto psicologico virtuale. In molti casi, senza che l'essere umano intervenga direttamente.

Ma quando queste macchine sbagliano chi paga il prezzo dell'errore? È una domanda che non ha ancora una risposta chiara. Un veicolo a guida autonoma può investire un pedone. Un algoritmo sanitario può consigliare una terapia inefficace. Un sistema di selezione del personale può scartare un candidato per motivi discriminatori. E in ognuno di questi casi, il danno è reale, tangibile.

Il problema è che il nostro sistema giuridico non è nato per queste situazioni. Basato su concetti tradizionali come colpa, imputabilità e nesso causale, fatica a confrontarsi con entità "intelligenti" che agiscono in autonomia, spesso senza che neppure il programmatore sappia prevedere esattamente il loro comportamento. Come ha dichiarato la Commissione Europea nella proposta di Direttiva sulla responsabilità da IA del 2022, le norme esistenti "non sono adatte a gestire le domande di risarcimento per danni causati da prodotti e servizi basati sull'intelligenza artificiale".

E questo significa, in concreto, che chi subisce un danno può trovarsi davanti a una zona grigia legale, dove il responsabile è difficile da identificare, il nesso causale quasi impossibile da dimostrare, e il risarcimento una chimera.

In un'epoca in cui sempre più decisioni vengono delegate agli algoritmi, la sfida del diritto è quella di restare umano.

1. Inquadramento generale: cos'è l'IA

Il termine *Intelligenza Artificiale* non ha un'unica definizione condivisa, ma viene generalmente indicato come la capacità di una macchina di emulare funzioni cognitive umane. John McCarthy, uno dei padri dell'IA, la definì nel 1956 come *"la scienza e l'ingegneria di far macchine intelligenti"*. L'IA si realizza attraverso algoritmi in grado di elaborare grandi quantità di dati e produrre output con vario grado di autonomia.

L'Intelligenza Artificiale, nel contesto normativo europeo, ha recentemente ricevuto una definizione ufficiale con l'adozione del Regolamento (UE) n. 2024/1689, noto come *AI Act*. Tale normativa descrive un "sistema di IA" come un software progettato per operare con un certo grado di autonomia, in grado di trasformare input in output come previsioni, raccomandazioni o decisioni

che possono incidere su ambienti fisici o virtuali. Questa definizione si ispira a quella elaborata dall'OCSE (Organizzazione per la Cooperazione e lo Sviluppo Economico) e adottata a livello europeo, e mette in evidenza il carattere autonomo e adattivo della tecnologia, che mira a replicare funzioni cognitive tipicamente umane attraverso l'uso di algoritmi software.

La dottrina distingue comunemente tra tre principali categorie di Intelligenza Artificiale: IA debole, IA forte e IA generativa.

L'**IA debole** (nota anche come *narrow AI*) comprende quei sistemi progettati per svolgere compiti circoscritti e specifici, come il riconoscimento vocale, la classificazione di immagini o l'elaborazione del linguaggio naturale. È questa la forma di IA attualmente più diffusa e concretamente implementata.

Al polo opposto vi è l'IA forte (o *general AI*), che rappresenta un'intelligenza artificiale teorica capace di ragionare, apprendere e adattarsi con la stessa flessibilità cognitiva di un essere umano. Al momento, l'IA forte resta un obiettivo ancora lontano, confinato più al campo della ricerca teorica che a quello dell'applicazione concreta.

A queste due si affianca l'IA generativa, un sottoinsieme dell'IA debole che si distingue per la capacità di creare nuovi contenuti, come testi, immagini, video o codici, basandosi su modelli addestrati su grandi quantità di dati. Strumenti come i modelli linguistici di grandi dimensioni (Large Language Models) e i generatori di immagini ne sono esempi emblematici. L'IA generativa è oggi al centro dell'interesse pubblico, accademico e normativo, per via delle sue vaste applicazioni e delle complesse questioni etiche e legali che solleva. Tra le forme più recenti e di maggiore impatto si colloca l'IA generativa, ovvero quella categoria di sistemi capaci di creare nuovi contenuti, siano essi testi, immagini, audio o video. Esempi noti al grande pubblico sono ChatGPT o DALL·E, sviluppati da OpenAI. Questi modelli sono addestrati su vasti insiemi di dati e utilizzano tecniche di deep learning per produrre output originali che risultano "il più simili possibile" a quelli presenti nei dati di addestramento. Il risultato è una tecnologia capace di generare contenuti complessi e articolati, come risposte coerenti in linguaggio naturale o immagini realistiche e stilisticamente variate.

Tuttavia, uno degli aspetti più problematici dei sistemi di IA, soprattutto quelli basati su tecniche di *deep learning*, è la loro natura. Molti di questi sistemi operano come delle vere e proprie "scatole nere" (*black box*), in cui il processo decisionale interno non è chiaramente osservabile né spiegabile,

nemmeno da parte dei programmatori stessi. Secondo parte della dottrina, i moderni sistemi di IA sono caratterizzati da una combinazione di complessità, incompletezza, imprevedibilità. In altre parole, sebbene l'output ottenuto possa essere tecnicamente corretto o funzionale, risulta estremamente difficile, se non impossibile, ricostruire l'iter logico seguito dal sistema. Questi algoritmi sono progettati per apprendere da grandi volumi di dati e per raggiungere un determinato risultato, ma "il risultato non è conoscibile dall'uomo… il 'come' non riesce a essere individuato". In sostanza, l'IA può giungere a decisioni che appaiono ragionevoli o persino brillanti, senza che sia possibile spiegarne con precisione il funzionamento interno.

Tra i settori critici in cui l'IA viene impiegata con maggiore intensità vi sono la sanità, i trasporti, la giustizia e la finanza.

Nel settore sanitario, gli algoritmi vengono utilizzati per la diagnosi attraverso l'analisi di immagini mediche – come radiografie, TAC e istologie – nonché per il supporto alle decisioni cliniche. Inoltre, esistono sistemi predittivi che, tramite l'analisi di grandi quantità di dati biologici, sono in grado di individuare la probabilità di insorgenza di determinate patologie.

Nel campo dei trasporti, sono in fase di sperimentazione veicoli a guida autonoma dotati di sistemi di assistenza avanzata, applicati non solo alle automobili, ma anche a droni e robot logistici. Tuttavia, non esistono ancora veicoli completamente autonomi autorizzati alla circolazione sulle strade pubbliche. Questo scenario rende particolarmente complessa la questione della responsabilità in caso di incidenti causati da questi sistemi.

Nel settore della giustizia, l'IA viene impiegata in strumenti di risk assessment, come il noto sistema COMPAS utilizzato negli Stati Uniti, in grado di prevedere la recidiva o la pericolosità di un soggetto. Tali valutazioni influenzano direttamente le decisioni dei giudici, ad esempio in materia di rilascio condizionale o di determinazione della pena.

Anche nel comparto finanziario, l'intelligenza artificiale è ormai largamente utilizzata. Si ricorre a sistemi di trading algoritmico automatico, a strumenti di valutazione del merito creditizio per l'erogazione di prestiti o mutui, nonché a soluzioni di consulenza finanziaria digitale e di analisi del rischio.

Questi esempi mostrano come l'IA stia rapidamente infiltrandosi in decisioni con grande impatto sulla vita delle persone. In ogni caso, le potenzialità positive (ad es. diagnosi precoce, maggiore

efficienza) coesistono con rischi specifici (errori diagnostici, bias discriminatori, incidenti stradali) che rendono urgente una riflessione giuridica approfondita.

2. La responsabilità civile nel diritto europeo e italiano

2.1 Fondamenti generali

La responsabilità civile per fatto illecito è disciplinata dall'art. 2043 c.c.: *"Qualunque fatto doloso o colposo che cagioni ad altri un danno ingiusto, obbliga colui che ha commesso il fatto a risarcire il danno"*. Gli elementi costitutivi della responsabilità sono quindi:

- **Fatto illecito** imputabile a una persona (dolo o colpa).
- **Danno ingiusto** subito da un terzo.
- **Nesso di causalità** tra fatto illecito e danno.

Il danneggiato, nelle azioni di risarcimento per colpa, deve provare tutti questi elementi: la condotta colpevole dell'autore, il danno concreto e il nesso causale. La giurisprudenza italiana richiede chiarezza sulla catena causale tra fatto e pregiudizio. Su tale assetto si innestano, in alcuni casi, forme di responsabilità **oggettiva** (senza colpa). Ad esempio, l'art. 2050 c.c. prevede che *"chiunque cagiona danno ad altri nello svolgimento di un'attività pericolosa… è tenuto al risarcimento, se non prova di aver adottato tutte le misure idonee a evitare il danno"*. In base a tale norma (attività pericolose) la responsabilità ricade sul titolare dell'attività, senza bisogno di provare colpa. Analogamente l'art. 2051 c.c. stabilisce una responsabilità oggettiva per danni causati da cose in custodia.

2.2 Responsabilità oggettiva o per colpa

In Italia, la responsabilità civile si distingue tra responsabilità per colpa (art. 2043 c.c.) e responsabilità oggettiva (es. art. 2050 c.c.). Nel primo caso, il danneggiato deve dimostrare la colpa (o il dolo) del soggetto responsabile, nonché il nesso causale tra il comportamento e il danno subito. Nel secondo caso, invece, ad esempio nell'ambito delle attività pericolose, è il titolare dell'attività a rispondere del danno, salvo che provi di aver adottato tutte le misure idonee a evitarlo. Un caso particolare di responsabilità oggettiva è rappresentato dalla responsabilità per prodotti difettosi, introdotta a livello comunitario con la Direttiva 85/374/CEE, la quale impone al produttore il risarcimento dei danni causati da un bene difettoso, senza necessità di provare la colpa.

La Direttiva 85/374/CEE, recepita nell'ordinamento italiano, disciplina infatti la responsabilità oggettiva del produttore per i danni provocati da prodotti difettosi: il danneggiato ha diritto al risarcimento se dimostra il difetto del prodotto e il nesso causale con il danno, senza dover provare l'elemento soggettivo della colpa. Con l'avvento dell'intelligenza artificiale, si è aperto un dibattito sulla possibilità di includere sistemi e algoritmi nella nozione di "prodotto". La Commissione Europea ha chiarito che i sistemi di IA e i beni basati su IA rientrano nell'ambito della direttiva sui prodotti difettosi. Ciò implica che, in linea di principio, se un prodotto dotato di IA (ad esempio un veicolo autonomo con software di guida automatica) risulta difettoso e provoca un danno, si applica la responsabilità oggettiva del produttore.

Inoltre, la proposta di revisione della direttiva (COM(2022) 495) prevede un aggiornamento del concetto di "prodotto", includendo non solo l'hardware ma anche il software e i servizi digitali che influiscono sul funzionamento del bene. La revisione estende l'ambito della direttiva ai beni dell'economia digitale, come dispositivi intelligenti e veicoli autonomi, e chiarisce che anche gli aggiornamenti software successivi all'immissione sul mercato, ad esempio quelli derivanti da processi di machine learning, possono mantenere o introdurre un difetto nel prodotto. In sostanza, la nuova disciplina mira a includere esplicitamente i sistemi di intelligenza artificiale tra i prodotti per i quali si prevede una responsabilità oggettiva in capo al "fornitore", inteso come produttore di hardware, software o di servizi digitali, per i danni derivanti da un funzionamento difettoso dell'IA.

2.3 L'art. 2043 e l'art. 2050 c.c. nel contesto dell'IA

Il regime generale di cui agli artt. 2043 e 2050 c.c. rimane comunque applicabile ai danni da IA. Se un danno è imputabile a un comportamento colposo (o doloso) di un soggetto umano (es. un tecnico che configura erroneamente un sistema), si potrà agire con l'art. 2043 c.c. come per qualsiasi torto. Se invece il danno deriva da un'attività intrinsecamente pericolosa che impiega IA (es. test bellici di un drone autonomo), potrebbe trovare applicazione l'art. 2050 c.c. con responsabilità oggettiva. In ogni caso, le sfide poste dall'opacità e dall'autonomia degli algoritmi complicano la prova della colpa e del nesso.

3. Le sfide della responsabilità nell'era dell'IA

3.1 Opacità delle decisioni algoritmiche

Gli algoritmi di IA, specie quelli di apprendimento profondo, generano decisioni *"imperscrutabili"* ("black box") anche ai loro creatori. Ciò ostacola la spiegazione del perché un dato risultato sia stato prodotto. Come rileva la dottrina, i sistemi IA autonomi elaborano dati in modo "non conoscibile dall'uomo" e quindi il "come" del processo decisionale non è individuabile. Questa mancanza di spiegabilità mette in crisi la responsabilità: senza poter comprendere i criteri interni, diventa arduo individuare errori o colpe specifici. Per queste ragioni si parla oggi di *Explainable AI (XAI)*, ovvero IA "spiegabile" che cerchi di rendere interpretabili i risultati del modello. L'idea è garantire che chi è coinvolto in un processo decisionale algortimico (sviluppatore, utilizzatore, vittima) possa comprendere almeno qualitativamente le ragioni di una decisione sospetta. Resta comunque un problema tecnico difficile: rendere trasparenti e controllabili i meccanismi di reti neurali molto complesse è soggetto di ricerca attiva. Dal punto di vista giuridico, la mancanza di spiegazioni automatiche rende più gravoso per il danneggiato provare la colpa e il nesso. Come osserva la Commissione, le norme nazionali vigenti non sono adatte a gestire le domande di risarcimento relative ai sistemi basati su intelligenza artificiale, in quanto questi possono essere complessi e autonomi, rendendo eccessivamente difficile per il danneggiato soddisfare l'onere della prova.

3.2 Autonomia parziale o totale delle macchine

L'automazione parziale o totale introduce incertezza su chi 'guida' il processo decisionale. Ad esempio, le auto attualmente autorizzate in strada operano con guida assistita di livello 2-3: il conducente deve rimanere vigile e può intervenire. Finché gli esseri umani mantengono la sorveglianza, la responsabilità resta in capo al conducente o al proprietario. Tuttavia, in prospettiva di una guida completamente autonoma (SAE livello 5) la situazione cambia: in teoria la responsabilità civile potrebbe spostarsi sul produttore dell'auto o del software. Nei casi reali noti, come l'incidente mortale Tesla in Texas (2021), l'assenza fisica del guidatore legittima invece l'attribuzione della colpa all'utente sopravvissuto. Le incertezze derivano dalla coesistenza di livelli di autonomia diversi e dalla transizione di controllo tra umano e macchina. Ad oggi, come osserva un esperto, *"non sono ancora su strada auto a guida completamente autonoma, ma è già acceso il dibattito sulle eventuali responsabilità"*. In altri settori (droni, robotica militare, sistemi decisionali) si hanno analoghi gradi di automazione: ogni livello intermedio pone problemi di attribuzione di responsabilità tra il sistema e l'operatore umano.

3.3 Assenza di soggettività giuridica dell'IA

Un altro tema è che l'IA non è soggetto di diritto. Macchine e algoritmi non possiedono capacità giuridica né possono essere *persone* in senso legale. Di conseguenza, non possono essere direttamente imputati di responsabilità civile come farebbe un individuo. Ciò implica che, in ogni caso, la responsabilità ricade sulle persone fisiche o giuridiche collegate all'IA. Le linee guida etiche europee 2019 ribadiscono che dev'esserci sempre *"un regime di responsabilità a carico delle persone (fisiche e giuridiche) che sviluppano, implementano o gestiscono sistemi di IA*". In altri termini, se il sistema provoca un danno, si ricercherà chi (imprese, professionisti, gestori, fornitori) ha causato o facilitato quell'evento. Anche il Regolamento UE riconosce esplicitamente il concetto di diversi attori: al "fornitore" (che progetta o commercializza il sistema IA) e al "deployer" (chi lo utilizza sotto la propria responsabilità) può essere imputato l'obbligo di diligenza. La sfida è capire come distribuire equamente tali responsabilità lungo la catena della filiera tecnologica.

3.4 Chi è responsabile? Programmatore, utilizzatore, fornitore, piattaforma…

Nel contesto IA si dibatte su quali figure debbano rispondere dei danni:

- **Sviluppatore/Fornitore**: chi progetta e immette sul mercato il sistema (hardware o software) potrebbe essere ritenuto responsabile se il danno deriva da un difetto di progettazione o addestramento errato del modello.
- **Utilizzatore/Deployer**: colui che mette in funzione l'IA (ad es. un ente, ospedale, azienda) potrebbe essere chiamato a rispondere se la violazione è dovuta a un uso inadeguato o all'assenza di vigilanza (per esempio se non rispetta le avvertenze d'uso).
- **Piattaforma intermedia**: in alcuni casi si ipotizza la responsabilità di piattaforme online che ospitano sistemi IA (es. social network con algoritmi di moderazione automatica) o gestori di marketplace digitali.
- **Responsabilità distribuita**: la tendenza normativa è di non attribuire tutti i rischi a un unico soggetto, ma di creare una **responsabilità condivisa**. Il nuovo quadro UE, infatti, delinea un sistema in cui *"ogni soggetto (fornitore, sviluppatore, utilizzatore)"* può essere ritenuto responsabile per i pregiudizi causati da un sistema IA sotto il suo controllo. In pratica, l'attribuzione della responsabilità terrà conto del ruolo concreto svolto da ciascun attore nella catena di produzione e utilizzo del sistema IA.

4. Casi reali ed emblematici

4.1 Caso Tesla (USA): incidente con pilota automatico

Un caso celebre è quello delle auto Tesla dotate di pilota automatico. Ad esempio, nell'incidente mortale in Texas del 2021, il pilota automatico non era in funzione (Tesla specificò che era il modello base senza Autopilot avanzato). Tuttavia, l'incidente ha riacceso il dibattito: secondo gli esperti, se in futuro esistessero auto completamente autonome (livello SAE 5), in caso di sinistro la responsabilità verrebbe probabilmente attribuita al produttore del veicolo o del software.

Nel caso concreto, invece, è stato ritenuto responsabile il conducente, in virtù del fatto che attualmente in strada si circola solo con sistemi di automazione parziale. Come nota la dottrina, il massimo livello di automazione oggi consentito è quello di livello 2; ciò significa che la responsabilità resta a carico della persona che stringe il volante. Questo esempio mette in luce come, finché il controllo è umano, costui sarà chiamato a rispondere, ma mostra anche l'inquietudine legale su futuri scenari di automazione completa.

4.2 Caso IBM Watson (USA): consigli medici errati

Nel settore sanitario, il sistema *Watson for Oncology* di IBM fu promosso come supporto avanzato per la terapia del cancro. Tuttavia, indagini giornalistiche hanno rivelato problemi significativi: documenti interni IBM mostrano che Watson spesso produceva consigli terapeutici pericolosi o sbagliati. In pratica, Watson era stato addestrato su casi sintetici e linee guida parziali anziché su dati reali, per cui le sue raccomandazioni si rivelavano spesso inaccurate e sollevavano gravi interrogativi sul processo di costruzione dei contenuti e sulla tecnologia sottostante.

Questo caso dimostra il rischio sanitario reale: un medico si può fidare ciecamente di un output errato generato da IA, finendo per nuocere al paziente. Sul piano della responsabilità, sorgono questioni delicate: è il produttore del software (IBM) a dover rispondere per negligenza nel training del sistema, o il medico che ha deciso di affidarsi senza sufficiente controllo umano?

A livello internazionale, la vicenda ha sottolineato la necessità di vigilanza rigorosa sui sistemi clinici basati su IA. In ogni caso, è un chiaro esempio di come errori algoritmici possano tradursi in danni gravi alla salute.

4.3 Caso COMPAS (USA): bias razziale negli algoritmi di giustizia

COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) è un algoritmo statunitense utilizzato per valutare il rischio di recidiva di imputati e detenuti. Un'inchiesta di ProPublica ha documentato che il programma produceva risultati assolutamente inaccurati, con un evidente bias razziale: ad esempio una donna nera (Brisha Borden) venne etichettata a *rischio alto* mentre un uomo bianco (Vernon Prater) a basso rischio, benché in realtà dopo due anni Borden non avesse più commesso reati mentre Prater fu arrestato di nuovo. Questa anomalia dimostra come gli algoritmi di IA possano incorporare e amplificare pregiudizi nei dati di training, discriminando per razza o classe sociale.

Il caso ha avuto rilevanza giuridica nel caso *State v. Loomis* del Wisconsin (2016): la Suprema Corte locale, pur ammettendo l'uso del punteggio COMPAS nella decisione della pena, ha criticato la sua segretezza e ha affermato che i giudici non dovrebbero farvi affidamento totale. In Italia simili sistemi non sono ancora in uso ufficiale, ma il dibattito sul loro impatto sui diritti è molto acceso. Questo esempio evidenzia la necessità di tutela dei diritti fondamentali: gli algoritmi predittivi di giustizia sollevano questioni di discriminazione e di equità, alimentando la domanda se e come la legge possa correggere comportamenti algoritmici ingiusti.

4.4 Altri casi (Italia/Europa): chatbot e assistenza medica

Non mancano episodi simili anche in Europa. Ad esempio, sono emerse notizie su chatbot o assistenti virtuali bancari che forniscono consulenze errate agli utenti (spingendo acquisti non richiesti o informazioni fuorvianti). Anche in campo medico, a livello europeo si registrano segnalazioni di errori in diagnosi assistite da IA, che hanno portato a richieste di risarcimento da parte di pazienti. Un caso particolarmente recente negli Stati Uniti riguarda ChatGPT: nel maggio 2025 un tribunale della Georgia ha stabilito che il gestore della piattaforma non poteva essere ritenuto responsabile per le *"allucinazioni"* (informazioni inventate) del chatbot. Il giudice ha infatti riconosciuto che i risultati forniti da un chatbot non vanno necessariamente presi sul serio, evidenziando i limiti d'affidabilità di questi sistemi. Sebbene non si tratti di una vicenda europea, il principio è rilevante: anche se un'informazione generata da IA inganna l'utente, attribuire la responsabilità è complesso. Nel complesso, questi casi sottolineano l'importanza di capire chi tutelare e come in scenari in cui l'errore arriva da un'entità 'non umana'.

5. Il quadro normativo europeo

5.1 Il Regolamento AI Act (approvato 2024)

Il Regolamento UE 2024/1689 (AI Act), approvato nel 2024, crea un quadro armonizzato per l'uso dell'IA in Europa. Si basa su un approccio basato sul rischio: gli usi dell'IA vengono classificati in base al livello di potenziale danno. In particolare, il regolamento prevede:

- **Usi vietati ("rischio inaccettabile")**: pratiche considerate pericolose per valori fondamentali (es. sistemi di *social scoring* governativo, manipolazione subliminale massiva, sorveglianza biometrica indiscriminata).
- **Usi ad alto rischio**: sistemi IA impiegati in settori critici (es. sanità, trasporti, infrastrutture critiche, educazione, giustizia, gestione di funzioni pubbliche, settore finanziario e banche). Questi devono rispettare requisiti stringenti: gestione del rischio, qualità dei dati, documentazione tecnica, sorveglianza umana, robustezza e sicurezza. Verranno sottoposti a valutazioni di conformità pre-market.
- **Usi a rischio limitato**: sistemi che richiedono trasparenza (ad es. chatbot devono informare l'utente che si tratta di IA; filtri di contenuti generati) ma non elevati livelli di controllo.
- **Usi a rischio minimo**: applicazioni generiche come modelli di linguaggio o videogiochi, per le quali il regolamento non impone obblighi particolari. Ad esempio, i modelli generativi come ChatGPT sono esenti da obblighi specifici se utilizzati in modo non critico.

Il regolamento introduce altresì un sistema di governance: ogni Stato membro istituirà autorità competenti che monitoreranno l'applicazione della norma. L'obiettivo è garantire che l'innovazione proceda in sicurezza: come sottolinea la Commissione, norme efficaci sulla responsabilità e sulla sicurezza investono nella fiducia e nella diffusione dell'IA" nell'UE. Questo provvedimento non riscrive la responsabilità civile, ma stabilisce condizioni per l'immissione e l'uso sicuro dei sistemi IA, coadiuvando gli operatori a prevenire i danni.

5.2 Proposta di Direttiva sulla responsabilità da IA (COM(2022) 496)

Oltre al regolamento di settore, la Commissione Europea ha proposto, nel novembre 2022, una Direttiva specifica finalizzata ad aggiornare le norme sulla responsabilità civile in presenza di sistemi di intelligenza artificiale. L'obiettivo principale della proposta è quello di facilitare il risarcimento alle

vittime di danni causati dall'utilizzo di tecnologie basate su IA, superando in particolare le difficoltà legate all'onere della prova che spesso ostacolano l'ottenimento di un giusto indennizzo.

La direttiva introduce alcune importanti novità. Anzitutto, viene alleggerito l'onere probatorio a carico del danneggiato, prevedendo una presunzione di causalità nei casi in cui la vittima riesca a dimostrare che un soggetto ha violato obblighi specifici relativi all'uso dell'IA e che vi sia una ragionevole probabilità che tale violazione abbia causato il danno. In simili circostanze, il giudice potrà presumere che il comportamento illecito sia all'origine dell'evento dannoso, salvo che la parte chiamata in causa riesca a fornire una prova contraria.

In secondo luogo, viene esteso il regime di responsabilità per colpa anche ai casi in cui il danno sia causato da un algoritmo autonomo. Ciò significa che sarà possibile intentare un'azione risarcitoria anche in assenza di un intervento umano diretto, purché sia possibile individuare almeno un soggetto umano o giuridico cui imputare una colpa o un'omissione nella gestione, nello sviluppo o nell'utilizzo del sistema IA.

Infine, la proposta prevede l'introduzione di un meccanismo di accesso semplificato alle informazioni, volto a garantire trasparenza e tracciabilità. I fornitori di sistemi di intelligenza artificiale e gli altri attori della filiera saranno tenuti a conservare documentazione tecnica e dati rilevanti per la ricostruzione di eventuali incidenti, e a condividerli con le vittime e le autorità giudiziarie qualora ciò sia necessario per far valere un diritto o accertare una responsabilità.

La proposta di Direttiva (attualmente ancora in fase di negoziazione) ha l'obiettivo di **adattare le regole attuali sulla responsabilità civile** — sia contrattuale che extracontrattuale — al nuovo contesto tecnologico dell'intelligenza artificiale. L'idea di fondo è garantire che chi subisce un danno causato da un sistema IA abbia una reale possibilità di ottenere un risarcimento, soprattutto quando il danno deriva da un errore o da una negligenza da parte di chi ha sviluppato, fornito o utilizzato quel sistema.

È importante chiarire che la direttiva non introduce un nuovo tipo di responsabilità oggettiva, cioè non prevede un risarcimento automatico solo perché c'è stato un danno. Al contrario, resta valido l'impianto giuridico già previsto dalle leggi nazionali, come gli articoli 2043 e 2050 del codice civile italiano (che regolano la responsabilità per colpa e per attività pericolose).

La novità è che, quando il danno è causato da un sistema di intelligenza artificiale, la proposta semplifica alcuni passaggi della prova a carico della vittima. In questo modo, si vuole rendere più

accessibile e concreto l'ottenimento del risarcimento, pur rimanendo all'interno delle regole giuridiche già esistenti.

5.3 Revisione della Direttiva Prodotti Difettosi (COM(2022) 495)

La proposta COM(2022) 495, che rivede la Direttiva 85/374 sulla responsabilità per danno da prodotti difettosi, nasce dall'esigenza di aggiornare un impianto normativo ormai superato rispetto alla complessità dei prodotti moderni, in particolare quelli che integrano software o sistemi di intelligenza artificiale. La proposta amplia infatti il concetto stesso di "prodotto", includendo espressamente beni con componenti digitali, come dispositivi intelligenti o veicoli autonomi, considerando parte del prodotto anche i sistemi di IA che ne costituiscono una funzionalità.

In questo contesto, la responsabilità del produttore viene estesa anche ai casi in cui il difetto emerga dopo la vendita, ad esempio a seguito di un aggiornamento software o di modifiche dovute all'apprendimento automatico: in tali situazioni, l'autore dell'aggiornamento potrà essere considerato a tutti gli effetti responsabile, come se fosse il produttore originario.

Un altro elemento rilevante è la ridefinizione del concetto di danno, che non si limita più ai soli danni fisici o materiali tradizionali, ma include anche i cosiddetti "danni digitali", come la perdita o l'alterazione di dati e contenuti immateriali collegati al prodotto. Questo consente, ad esempio, di ottenere un risarcimento qualora un sistema IA difettoso provochi la cancellazione di dati importanti. La proposta si occupa inoltre della cosiddetta economia circolare, stabilendo che anche i prodotti rigenerati o ricondizionati devono garantire lo stesso livello di tutela previsto per quelli nuovi, assicurando alle vittime la possibilità di ottenere un risarcimento analogo.

In sostanza, la revisione della direttiva punta a colmare le lacune emerse negli ultimi decenni, in cui l'evoluzione tecnologica ha reso sempre più difficile per i danneggiati dimostrare la colpa e il nesso causale nei casi complessi. Per questo motivo, la Commissione propone un aggiornamento che consenta di applicare il regime di responsabilità oggettiva anche ai prodotti con IA, alleggerendo l'onere della prova e introducendo procedure più semplici per le vittime. Tutti i sistemi di intelligenza artificiale che funzionano come beni rientreranno così nel perimetro della direttiva, offrendo una tutela più adeguata alle nuove sfide poste dal digitale.

5.4 Iniziative nazionali italiane e posizioni del Garante Privacy

L'Italia si sta muovendo per affrontare le sfide giuridiche poste dall'intelligenza artificiale, cercando di allinearsi al nuovo quadro normativo europeo delineato dall'AI Act. A livello nazionale, è attualmente in discussione un disegno di legge delega, presentato nell'agosto 2023, che ha l'obiettivo di recepire e attuare i principi del Regolamento europeo e di introdurre una legge quadro sull'IA.

In questo contesto, il Garante per la protezione dei dati personali ha svolto un ruolo attivo nel valutare lo schema normativo italiano. Con il provvedimento n. 477/2024, l'Autorità ha espresso un parere sostanzialmente favorevole, riconoscendo l'importanza di dotarsi di un quadro nazionale solido e coerente con le direttive europee. Tuttavia, nel suo intervento il Garante ha evidenziato alcuni "punti di attenzione", soprattutto per quanto riguarda la tutela dei diritti fondamentali. Tra le osservazioni principali, l'Autorità sottolinea la necessità di garantire la trasparenza degli algoritmi, di definire con chiarezza le responsabilità giuridiche dei soggetti coinvolti e di prevedere rigorosi meccanismi di controllo, in particolare nei settori a rischio elevato.

Oltre all'ambito della responsabilità civile, la riflessione normativa si intreccia con la disciplina della protezione dei dati personali. Il Regolamento generale sulla protezione dei dati (GDPR), insieme alla normativa sull'ePrivacy, contiene infatti articoli che impongono obblighi specifici nell'uso dell'IA per il trattamento dei dati. In linea con questi principi, il Garante Privacy italiano ha già pubblicato linee guida su alcune applicazioni critiche, come il riconoscimento facciale e gli assistenti virtuali, sottolineando l'importanza della proporzionalità, del consenso informato e della minimizzazione dei dati.

6. L'evoluzione degli approcci e prospettive

Il tema della responsabilità civile nell'era dell'intelligenza artificiale è stato oggetto di un ampio dibattito, che ha visto emergere diverse proposte e visioni nel tempo. Il confronto iniziale ha riguardato soprattutto la contrapposizione tra un approccio soggettivo, basato sulla colpa, e uno oggettivo, volto a facilitare il risarcimento alle vittime. Questo dibattito ha progressivamente lasciato spazio a una soluzione mista, che oggi si riflette nelle proposte europee: responsabilità oggettiva per i prodotti IA e forme attenuate di responsabilità soggettiva con presunzioni di causalità per gli altri casi.

Nel pieno di questa evoluzione teorica si è anche fatta strada, per un breve periodo, l'idea di attribuire personalità giuridica autonoma ad alcune IA avanzate, trattandole come entità responsabili. Una proposta che, pur suscitando interesse, è stata poi abbandonata per la difficoltà di fondarla giuridicamente e per l'opacità dei sistemi IA stessi, che rende complesso identificare decisioni effettivamente "autonome".

Più concretamente, si sta affermando una logica di responsabilità distribuita, che riconosce come, nei sistemi IA, il danno possa derivare da una molteplicità di fattori lungo l'intera filiera: dallo sviluppo all'utilizzo. In quest'ottica si discute la creazione di fondi di garanzia o meccanismi assicurativi che sollevino la vittima dal peso di individuare un unico responsabile.

Infine, prende sempre più forza il principio della "spiegabilità" dell'IA (Explainable AI): la trasparenza dei sistemi diventa un presupposto essenziale non solo per l'etica e la fiducia, ma anche per l'effettivo esercizio dei diritti. L'AI Act recepisce questo orientamento, rafforzando gli obblighi informativi soprattutto nei sistemi ad alto rischio, e aprendo la strada a un futuro diritto a ricevere spiegazioni sulle decisioni algoritmiche.

7. Considerazioni finali e nodi aperti

Nel corso degli ultimi anni, l'emergere dell'intelligenza artificiale ha sollevato interrogativi sempre più pressanti sul piano della responsabilità giuridica. Se da un lato le istituzioni europee e nazionali stanno cercando di colmare i vuoti normativi attraverso strumenti come l'AI Act e la revisione della direttiva sui prodotti difettosi, dall'altro lato permangono aree grigie e problemi aperti che meritano attenzione.

Il primo nodo riguarda la tutela effettiva delle vittime. In un ecosistema tecnologico dove i danni possono derivare da decisioni algoritmiche opache e distribuite lungo molteplici livelli di responsabilità, c'è il rischio concreto che nessuno risponda pienamente. Se il sistema legale non riesce a identificare un responsabile o se la prova diventa tecnicamente irraggiungibile, la vittima può trovarsi priva di rimedi. Il diritto sta cercando di reagire, ad esempio introducendo presunzioni di causalità e ipotizzando fondi di garanzia che consentano di accedere al risarcimento anche in assenza di una colpa diretta chiaramente individuabile. Ma la reale efficacia di queste misure dipenderà, più che dalle formulazioni teoriche, dalla loro attuazione pratica e dal coordinamento tra sistemi giuridici nazionali.

Un altro elemento critico è rappresentato dai limiti strutturali del diritto vigente, pensato in un'epoca pre-algoritmica. Le regole attuali si basano su un nesso causale lineare e su condotte umane identificabili, mentre l'IA opera in modo dinamico e a volte imprevedibile. Inoltre, la continua evoluzione dei sistemi, che possono essere aggiornati da remoto o modificarsi autonomamente, rende complicata la ricostruzione del momento e del soggetto in cui è sorto il difetto.

A ciò si aggiunge il rischio di deresponsabilizzazione: quando l'IA agisce come supporto alle decisioni umane, ma l'errore non è attribuibile né all'algoritmo né all'operatore, può crearsi un vuoto di responsabilità. Se non si chiarisce chi debba effettivamente vigilare, controllare o intervenire, si rischia che nessuno risponda, né a livello civile né a livello assicurativo. In questi casi, il bilanciamento tra innovazione tecnologica e garanzie per le vittime si fa particolarmente delicato.

L'automazione pone anche questioni etiche e culturali, che il diritto non può ignorare. L'utilizzo massiccio di sistemi IA in ambiti sensibili può amplificare bias preesistenti e ridurre la responsabilità etica individuale. Per questo, sempre più si parla del diritto alla "spiegabilità" delle decisioni algoritmiche (Explainable AI): un principio che si affianca a quello della trasparenza e della supervisione umana, nella prospettiva di mantenere un controllo effettivo sull'azione delle macchine. Anche il conflitto tra protezione dei dati personali e necessità di alimentare i sistemi IA con grandi quantità di informazioni pone sfide importanti, richiedendo un equilibrio tra innovazione e rispetto dei diritti fondamentali sanciti dalla Carta dell'Unione Europea.

La sfida che il diritto oggi affronta non è solo tecnica, ma anche politica e culturale: trovare regole che accompagnino lo sviluppo dell'IA senza sacrificare la giustizia, la responsabilità e i diritti delle persone.

Bibliografia

Commissione Europea (2022). Proposta di Direttiva del Parlamento Europeo e del Consiglio relativa all'adattamento delle norme in materia di responsabilità extracontrattuale ai danni causati dall'intelligenza artificiale (Direttiva sulla responsabilità per l'IA), COM(2022) 496 final.

Commissione Europea (2020). Libro bianco sull'intelligenza artificiale – Un approccio europeo all'eccellenza e alla fiducia, COM(2020) 65 final.

Parlamento Europeo e Consiglio dell'Unione Europea (2024). Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce norme armonizzate sull'intelligenza artificiale (AI Act).

McCarthy, J. et al. (1956). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth College.

Garlaschelli, D. (2023). Intelligenza artificiale e responsabilità civile. I nuovi scenari normativi in Europa. In: Rivista di Diritto Civile, vol. 69, n. 2, pp. 215–240.

Giannini, M. (2022). Il problema della "black box" nei sistemi di IA: limiti e prospettive dell'Explainable AI. In: Giurisprudenza Italiana, vol. 174, n. 5, pp. 1023–1040.

Pagallo, U. (2020). Responsabilità civile e tecnologie intelligenti: linee evolutive e scenari futuri. In: Diritto dell'Informazione e dell'Informatica, vol. 36, n. 4, pp. 623–645.

European Commission Expert Group on Liability and New Technologies (2019). Liability for Artificial Intelligence and Other Emerging Digital Technologies. Bruxelles: Commissione Europea.

Vincent, J. (2023). How AI “black boxes” are being opened with explainability methods. In: Nature Machine Intelligence, vol. 5, pp. 9–11.

Perlingieri, P. (2016). Il danno alla persona e la responsabilità civile: nuovi orizzonti. Napoli: Edizioni Scientifiche Italiane.

Sartor, G. (2021). Artificial Intelligence and Legal Responsibility. In: Philosophy & Technology, vol. 34, pp. 239–257.

Wachter, S., Mittelstadt, B., Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. In: International Data Privacy Law, vol. 7, n. 2, pp. 76–99.

Kroll, J.A. et al. (2017). Accountable Algorithms. In: University of Pennsylvania Law Review, vol. 165, n. 3, pp. 633–705.

Bryson, J. J., Diamantis, M. E., Grant, T. D. (2017). Of, for, and by the People: The Legal Lacuna of Synthetic Persons. In: Artificial Intelligence and Law, vol. 25, n. 3, pp. 273–291.

Goodman, B., Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. In: AI Magazine, vol. 38, n. 3, pp. 50–57.