CAMMINO DIRITTO
Rivista di informazione giuridica
https://rivista.camminodiritto.it

Articolo Divulgativo

CHATBOT E RESPONSABILITÀ CIVILE: NUOVE FRONTIERE TRA DIFETTO INFORMATIVO E TUTELA DELL'AFFIDAMENTO

Il saggio esamina le sfide che i chatbot fondati su Large Language Models pongono alla responsabilità civile, configurandosi come un banco di prova decisivo per il diritto privato. Centrale è il difetto informativo, inteso non come guasto tecnico, ma come insufficienza strutturale di trasparenza, avvertenze e supervisione. L'opacità dell'architettura algoritmica genera nuove forme di affidamento e vulnerabilità cognitiva, difficilmente riconducibili a colpa, causalità e danno secondo i modelli tradizionali. Ne discende l'esigenza di una riconfigurazione antropocentrica della responsabilità, che valorizzi trasparenza sostanziale, tracciabilità e controllo umano, così da coniugare innovazione tecnologica e tutela effettiva dei diritti.

di **Matteo Pertosa**
IUS/01 - DIRITTO PRIVATO
Articolo divulgativo - ISSN 2421-7123

Direttore responsabile
Alessio Giaquinto

Pubblicato, Giovedì 21 Agosto 2025

CamminoDiritto.it
ISSN 2421-7123

Abstract ENG

This essay examines the challenges that chatbots based on Large Language Models pose to civil liability, emerging as a decisive test for private law. At its core lies the informational defect, conceived not as a technical fault but as a structural deficiency of transparency, warnings, and supervision. The opacity of algorithmic design gives rise to novel forms of reliance and cognitive vulnerability, scarcely encompassed by traditional notions of fault, causation, and damage. From this stems the need for an anthropocentric reconfiguration of liability, enhancing substantive transparency, traceability, and human oversight, thereby reconciling technological innovation with the effective protection of rights.

---

1. Profili introduttivi e quadro tecnico-funzionale.

I chatbot di ultima generazione (basati su Large Language Models) costituiscono oggi un banco di prova cruciale per il diritto privato, segnando una trasformazione profonda che, per portata innovativa, richiama quella introdotta dall'automobile agli inizi del Novecento [1]. A detta di chi scrive, dinnanzi a queste nuove tecnologie occorre una verifica sistematica degli istituti della responsabilità e della tutela, valutando come colpa, nesso causale, difetto di prodotto e doveri informativi interagiscono con l'intelligenza artificiale generativa quando questa trova applicazione concreta nei settori della sanità, del credito, del lavoro e dell'istruzione. Sul piano tecnico, i Large Language Models si fondano sull'architettura Transformer e operano mediante meccanismi di self-attention e generazione auto-regressiva, raggiungendo prestazioni elevate ma senza una reale comprensione semantica nella sequenza di produzione[2].

La scala parametrica e il pre-addestramento su corpora eterogenei, pur incrementando l'efficacia, amplificano l'opacità del funzionamento, generando una variabilità dell'output che rende complessa e tutt'altro che lineare la ricostruzione del nesso eziologico tra input e output[3]. Traslando tali osservazioni sul piano giuridico, l'errore non si identifica con il guasto di un singolo componente, ma è il prodotto di una catena causale socio-tecnica che coinvolge dati di addestramento, architettura del modello, prompt, fasi di fine-tuning e utente finale.

CamminoDiritto.it
ISSN 2421-7123

Ne consegue che il difetto rilevante si configura non già come mera rottura funzionale del sistema, bensì come carenza strutturale delle informazioni e della presentazione del prodotto o servizio, espressa in avvertenze e istruzioni inadeguate, nella mancata esplicitazione dei limiti d'uso e nell'assenza di canali di supervisione effettivi[4]. A ciò si aggiunge la colpa organizzativa del fornitore del prodotto, che si traduce in doveri di assetto, controllo e tracciabilità lungo l'intero ciclo di vita del servizio, imponendo una supervisione umana significativa ogniqualvolta l'automazione incida su posizioni giuridiche individuali[5].

In questa prospettiva, si consideri un istituto di credito che, tramite il chatbot integrato nell'home banking, indichi all'utente un termine di recesso o un costo essenziale non conformi alla disciplina vigente e non metta a disposizione un contatto umano immediato. L'interfaccia fa parte della comunicazione ufficiale, l'affidamento che ne deriva è certamente prevedibile e l'assenza di cautele è sicuramente verificabile. Pertanto, in una situazione del genere la qualificazione in termini di difetto informativo e di colpa organizzativa risulta giustificata, poiché risultano disattesi i doveri di chiarezza, correttezza e vigilanza sull'uso del sistema da parte dell'utente. Tuttavia, non ogni imprecisione comporta responsabilità.

Se il servizio si limita a fornire "informazioni generali", espone i limiti ed esclude risposte operative garantendo un contatto umano tracciabile, la soglia del difetto può non risultare superata. In questa direzione si muove anche la giurisprudenza della Corte di giustizia nel caso SCHUFA. La decisione ha qualificato lo scoring creditizio come decisione automatizzata idonea a produrre effetti giuridici e ha richiesto trasparenza sostanziale, accesso alle informazioni rilevanti e un intervento umano effettivo e capace di incidere sull'esito**,** confermando che, quando l'automazione concorre alla formazione di scelte che toccano diritti individuali, l'ordinamento pretende intelligibilità e possibilità di correzione da parte di un supervisor umano posto a monte[6].

2. Il difetto informativo nelle intelligenze artificiali generative: fulcro della responsabilità efficiente.

A detta di chi scrive, nei sistemi di IA generativa il difetto informativo assume un ruolo centrale nella definizione della responsabilità civile. Non si tratta semplicemente di un errore tecnico, ma della mancanza di una comunicazione adeguata sull'affidabilità, sui limiti e sulla natura stessa del sistema, che può indurre l'utente a un affidamento ingiustificato[7].

Tale fenomeno affonda le sue radici nel disordine intrinseco alle tecnologie auto-addestranti, le quali producono contenuti plausibili e coerenti nella forma, ma non

CamminoDiritto.it
ISSN 2421-7123

necessariamente corretti nella sostanza. L'alterazione non deriva soltanto dall'eterogeneità dei dati di addestramento, ma dal modo in cui l'algoritmo li rielabora, generando un flusso comunicativo che, pur apparendo attendibile, può tradursi in informazione fuorviante o ingannevole[8]. In questo senso, l'IA amplifica le insidie tipiche della disinformazione tradizionale, rendendole più pervasive e difficili da decifrare[9]. Come detto in precedenza, i Large Language Models (LLM) non possiedono una vera capacità di comprensione semantica, ma operano su base probabilistica attraverso l'architettura Transformer.

Proprio questa natura opaca trasforma l'interazione con il chatbot in una relazione più che in un semplice uso strumentale. Tale relazione trova origine nell'affidamento che l'utente ripone nell'output prodotto da un sistema non correttamente introdotto. In tale prospettiva, il difetto informativo si configura come un vero e proprio "vizio di introduzione", ravvisabile allorché difettino avvertenze chiare, istruzioni adeguate o concreti canali di supervisione umana[10]. Secondo tale impostazione, viene a delinearsi un vero e proprio responsibility gap, ossia un divario strutturale tra l'assetto normativo della responsabilità civile e le concrete modalità di imputazione del fatto illecito. Difatti, nei sistemi ad alta complessità, l'opacità e la distribuzione delle fasi di sviluppo e impiego impediscono di individuare con immediatezza il soggetto responsabile. In termini giuridici, tale scarto comporta il rischio di dissolvere il nesso tra condotta ed evento dannoso, creando zone di irresponsabilità sistemica.

La dottrina più attenta[11] ha evidenziato come il superamento di tale vuoto non implichi l'introduzione di automatismi sanzionatori, bensì la ridefinizione dei doveri di organizzazione, trasparenza e vigilanza lungo l'intero ciclo di vita del sistema, in modo che la responsabilità rimanga saldamente ancorata a parametri di prevedibilità e controllabilità dell'output. Pertanto, appare evidente che il problema del difetto informativo non si riduce a una mera carenza comunicativa, ma incide sulla stessa architettura della relazione fra uomo e macchina. In definitiva, il giurista è chiamato a valorizzare il design comunicativo dei sistemi di IA, rendendo espliciti i limiti, garantendo la possibilità di intervento umano e assicurando una trasparenza sostanziale mediante la quale è possibile ricostruire ogni step che ha generato un determinato output. Solo così la responsabilità civile può adeguarsi alla trasformazione tecnologica senza rinunciare alla sua funzione primaria di tutela effettiva della persona.

3. Dall'affidamento al danno nell'ecosistema algoritmico delle IA

Se in precedenza l'attenzione si è concentrata sulla natura strutturale del difetto informativo, occorre ora rivolgere lo sguardo al danno e alla relativa prova, specie quando l'output dei chatbot non si traduce in un pregiudizio materiale immediato, bensì in

CamminoDiritto.it
ISSN 2421-7123

conseguenze di natura immateriale, cognitiva o reputazionale.

Come si è visto in precedenza, l'uso crescente di strumenti di IA generativa mostra come l'errore non si manifesti più unicamente in termini patrimoniali tradizionali, ma possa assumere forme diverse, quali, ad esempio, la perdita di chance decisionali o la compromissione della libertà informativa dell'utente. Il nodo problematico risiede nel fatto che la variabilità degli output generati dai modelli linguistici di grandi dimensioni rende difficile distinguere tra l'imprecisione tollerabile e l'errore giuridicamente rilevante.

In tali condizioni, la tradizionale prova del nesso eziologico rischia di trasformarsi in un onere sproporzionato per la vittima, soprattutto quando la disinformazione algoritmica opera in maniera sottile e progressiva, erodendo la capacità di autodeterminazione senza tradursi in un evento dannoso tangibile e immediato. Si pensi, ad esempio, a un sistema di IA generativa che, sottoposto a ripetute sollecitazioni da parte dell'utente, modifichi le proprie risposte nel tempo sino a fornire un output apparentemente favorevole alle aspettative soggettive di un utente, ma in realtà erroneo e potenzialmente fuorviante per un altro[12]. In questa prospettiva, l'affidamento dell'utente assume un rilievo centrale.

Come si è evidenziato in precedenza, l'interazione costante con interfacce conversazionali, che simulano un linguaggio naturale e forniscono risposte convincenti, genera aspettative di affidabilità difficilmente scindibili dalla percezione di veridicità[13]. Il diritto civile, a detta di chi scrive, non può continuare a ridurre tale danno alla mera perdita patrimoniale. La sua natura richiede un ampliamento concettuale che tenga conto delle nuove forme di vulnerabilità cognitiva, ponendo al centro il parametro dell'utente ragionevole e dell'affidamento legittimo da questi maturato. In tal senso, il danno da chatbot non si identifica soltanto con l'errore tecnico dell'algoritmo, ma con la lesione dell'equilibrio informativo necessario all'esercizio della libertà di scelta.

La sfida consiste dunque nel ripensare la categoria del danno risarcibile alla luce di tali trasformazioni, elaborando criteri idonei a riconoscere e compensare le nuove forme di pregiudizio digitale che non si lasciano facilmente ricondurre agli schemi tradizionali. Solo attraverso questo ampliamento, la responsabilità civile potrà continuare a svolgere la sua funzione di tutela effettiva in un ecosistema algoritmico che, sempre più, condiziona le decisioni individuali e collettive.

4. Considerazioni conclusive

La progressiva diffusione dei chatbot basati su modelli linguistici di grandi dimensioni mostra come l'ordinamento civile sia chiamato a confrontarsi con un fenomeno che trascende le categorie tradizionali. Il danno non si esaurisce più nella dimensione

CamminoDiritto.it
ISSN 2421-7123

patrimoniale, ma si radica nella sfera cognitiva e relazionale dell'individuo, minandone la capacità di autodeterminazione attraverso meccanismi di disinformazione sottile, opacità funzionale e affidamento ingannevole. In tale scenario, il concetto di difetto informativo non è un mero corollario tecnico, ma diviene la chiave di volta per comprendere la natura giuridica del pregiudizio. A detta di chi scrive la responsabilità civile deve essere riconsiderata in senso antropocentrico, riaffermando la centralità della persona come parametro ordinante[14].

Non basta accertare se l'algoritmo abbia prodotto un output inesatto; occorre piuttosto domandarsi se il sistema sia stato progettato, presentato e supervisionato in modo da evitare la creazione di un affidamento irragionevole nell'utente. L'obbligo di trasparenza e il presidio del controllo umano assumono così il ruolo di strumenti giuridici imprescindibili, volti a preservare l'equilibrio tra innovazione tecnologica e tutela effettiva dei diritti, ad esempio attraverso la progettazione dell'interfaccia utente in modo tale da mostrare un chiaro avvertimento quando l'output derivi da dati non confermati da una validazione umana. In definitiva la disciplina civilistica deve evolvere verso una concezione del danno digitale capace di valorizzare l'esperienza dell'utente e la sua integrità cognitiva.

Il futuro della responsabilità non può ridursi a un semplice adattamento linguistico degli istituti esistenti, ma richiede un'elaborazione critica in grado di riconoscere le nuove forme di pregiudizio e di predisporre regole di imputazione coerenti con le relazioni tecnologiche contemporanee. Solo in tal modo sarà possibile garantire che lo sviluppo dell'intelligenza artificiale avvenga nel rispetto dei principi fondamentali dell'ordinamento, senza trasformarsi in una nuova area di irresponsabilità.

Diversamente, il rischio è che, confidando in una autoregolamentazione insufficiente o in meri ritocchi normativi, si lasci proliferare un fenomeno destinato a divenire ingestibile. Con l'avanzare delle nuove tecnologie, il diritto potrebbe allora trovarsi costretto non a regolarle, ma a bloccarne l'uso per prevenire catastrofi sistemiche, con un esito regressivo che mortificherebbe sia l'innovazione sia la tutela dei diritti fondamentali.

CamminoDiritto.it
ISSN 2421-7123

Note e riferimenti bibliografici

[1] Cfr. U. Ruffolo, Responsabilità da A.I. e da algoritmo, in ID. (a cura di), Intelligenza artificiale. Il diritto, i diritti, l'etica, Milano, 2020, pp. 65-66.

[2] Cfr. T. Lin, Y. Wang, X. Liu, X. Qiu, A survey of transformers, in Ai Open, 3, 2022, pp. 111 ss.

[3] Cfr. C. Agata, Intelligenza artificiale, big data e nuovi diritti, in Rivista Italiana di informatica e diritto, 1, 2022, pp. 94-97 e Cfr. A. Engel, M. Mauer, Regulating ChatGPT and other Large Generative AI Models, in FAccT '23: Proceeding of the 2023 ACM Conference, in Fairness, Accountability, and Transparency, 6, 2023, pp. 1112-1123.

[4] Cfr. D. Campilongo, La responsabilità civile del produttore, in il diritto privato nella giurisprudenza, (a cura di) P. Cendon, La responsabilità civile, XI, Torino, 1998, pp. 453-454 ss.

[5] Sul punto Cfr. G. Finocchiaro, Responsabilità da algoritmi e nuove forme del danno nella società digitale, in Rivista di diritto privato europeo, 2, 2024, pp. 80 ss.

[6] Per un analisi più accurata si suggerisce la lettura del caso SCHUFA - CGUE, Sez. I, 7 dicembre 2023, causa C-634/21, OQ c. Land Hessen - che ha coinvolto un sistema automatizzato di credit scoring basato su IA, il cui output ha inciso in modo determinante sulla concessione di un prestito. La Corte ha riconosciuto che una decisione fondata su un calcolo algoritmico di probabilità, qualora priva di un intervento umano significativo e produttiva di effetti giuridici, rientra nell'ambito applicativo dell'art. 22 RGPD, con conseguente necessità di garantire trasparenza, accessibilità e controllo umano. Il caso evidenzia come l'uso dell'IA, se non accompagnato da adeguate garanzie, possa rafforzare pratiche opache e potenzialmente discriminatorie, compromettendo i diritti dell'individuo alla parità di trattamento, alla protezione dei dati personali e a un ricorso effettivo, specialmente quando l'algoritmo si basa su dati storici o correlazioni statistiche che riproducono bias sistemici. La sentenza rafforza pertanto l'esigenza di una governance algoritmica responsabile, in cui il principio di non discriminazione sia tutelato anche nei processi decisionali delegati alla macchina. Sul punto Cfr. V. Caforio, F. Paolucci, Processi decisionali automatizzati, tutela dei dati personali e del segreto commerciale nell'Unione europea: spunti dal caso "SCHUFA" (causa C-634/21), in Il Diritto industriale, 3, 2025, p. 254. Un ulteriore esempio si veda in G. Finocchiaro, Diritto dell'intelligenza artificiale, Bologna, 2024, p. 98, ove si sottolinea come l'impiego di chatbot nei servizi al pubblico imponga la presenza di una costante supervisione umana, al fine di evitare che l'affidamento dell'utente venga compromesso da risposte imprecise o fuorvianti. In questa prospettiva si colloca anche il caso Moffat v. Air Canada, deciso dalla Supreme Court of British Columbia nel 2024, che ha affermato la responsabilità della compagnia aerea per le informazioni errate fornite dal proprio chatbot in merito alle condizioni di rimborso, ribadendo che l'automazione non può mai tradursi in deresponsabilizzazione, ma al contrario rafforza i doveri organizzativi di prevenzione dei danni da disinformazione.

[7] U. Ruffolo, Il disordine informativo e l'Intelligenza Artificiale: tra insidie e possibili strumenti di contrasto, in Rivista di diritto civile, 3, 2023, p. 5 ss.

[8] Op. Cit., p. 12 ss.

[9] Cfr. G. Comandé, Intelligenza artificiale e responsabilità tra liability e accountability, il carattere trasformativo dell'IA e il problema della responsabilità, in Analisi Giuridica dell'Economia, 1, 2019, p. 178.

[10] G. Finocchiaro, Il diritto dell'intelligenza artificiale, Bologna, 2024, p. 152.

[11] Cfr. U. Ruffolo, A. Amidei, Diritto dell'intelligenza artificiale, vol. 1, Roma 2024, p. 29-30 ss.

[12] Cfr. S. Calzolaio, Vulnerabilità della società digitale e ordinamento costituzionale dei dati, in Rivista italiana di informatica e diritto, 2, 2023, p. 21-22 ss.

[13] Cfr. A. Ruffo, Il disordine informativo e l'Intelligenza Artificiale; tra insidie e possibili strumenti di contrasto, in Rivista di diritto dei media, 1, 2024, p. 470-471.

[14] In merito al principio di antropocentricità, si veda G. Finocchiaro, Diritto dell'intelligenza artificiale, Bologna,

CamminoDiritto.it
ISSN 2421-7123

2024, p. 152, ove si rileva che i sistemi di intelligenza artificiale, pur costituendo un ausilio prezioso soprattutto nelle controversie fondate su circostanze chiaramente delimitate, non sono in grado di sostituire l'attività interpretativa propria dell'avvocato e del giudice. Tale osservazione evidenzia l'indispensabilità della supervisione umana e richiama il principio di antropocentricità dell'ordinamento giuridico, che impone di preservare la centralità dell'essere umano nei processi decisionali, specialmente laddove siano richieste valutazioni discrezionali, giudizi di valore e responsabilità ermeneutica.

---

* Il simbolo {https/URL} sostituisce i link visualizzabili sulla pagina:
https://rivista.camminodiritto.it/articolo.asp?id=11301