

LA TUTELA PENALE NELL'USO DEI CHATBOTS PER IL SUPPORTO PSICOLOGICO. IL CONTENUTO DEL DOVERE DI DILIGENZA

di Federico Rizzolo

Il crescente ricorso a chatbots di IA quali interlocutori di sostegno psicologico apre un orizzonte nuovo. Nei soggetti affetti da disturbi mentali, l'architettura del servizio – specie se basata su meccanismi di omofilia – può accentuare dinamiche di overreliance e alimentare fattori di rischio suicidario. La letteratura di settore segnala, in questa prospettiva, rischi per la vita e l'incolumità individuale. Beni rispetto ai quali la legislazione italiana in materia di intelligenza artificiale anticipa la soglia della tutela penale, delineando fattispecie di pericolo concreto. Muovendo dall'avvertimento sugli usi simbolici della pena, il contributo propone una prima ricostruzione dogmatica dei presupposti dell'eventuale responsabilità penale dei fornitori, soffermandosi sul rischio tipico quale criterio di delimitazione del contenuto del dovere di diligenza.

SOMMARIO: 1. L'intreccio tra intelligenza artificiale e diritto penale nella trama dell'inquietudine. Coordinate prime. – 2. La tutela penale dell'*homo materia*: il *kairos* per una riflessione dogmatica dinanzi all'artificio per il supporto psicologico. – 2.1. L'intelligenza artificiale: da noumeno a fenomeno per il diritto penale. Le questioni della riconoscibilità ed evitabilità. – 2.1.1. Il criterio del rischio tipico nel giudizio di riconoscibilità. – 2.1.2. L'area del rischio consentito nell'ambito dei *chatbot therapists* prima e dopo l'AI Act. – 2.1.3. La logica delle regole dell'AI Act applicabili ai *chatbot therapists*: da precauzione a prevenzione. – 2.2. L'intelligenza artificiale: da noumeno a fenomeno per la scienza penalistica. Meta-conclusioni sul metodo assunto per il prosieguo dell'indagine sull'evitabilità. – 2.2.1. Dal livello di risoluzione dell'analisi alla definizione del problema. – 3. La responsabilità "per il tipo" (di produzione): i *chatbot therapists* come pratiche vietate. – 3.1. L'imputazione soggettiva: tra dolo eventuale e principio di precauzione. – 4. La responsabilità "per il modo": i *chatbot therapists* come applicazioni consentite. – 4.1. Le cautele previste ma non imposte dall'AI Act. Il criterio del rischio tipico nel giudizio di evitabilità. – 4.2. L'espansione della diligenza: tra dinamiche di ingerenza e profili di colpa per assunzione. – 5. Il principio di affidamento e l'autoesposizione al pericolo dell'utente. Brevi spunti di impronta vittimologica e vittimodogmatica. – 6. Considerazioni finali sulla tutela della vita e dell'incolumità di utenti caratterizzati da disturbi mentali: per un dialogo tra scienze.

1. L'intreccio tra intelligenza artificiale e diritto penale nella trama dell'inquietudine. Coordinate prime.

Nell'esperienza umana, l'inquietudine collettiva può essere vista come una costante fonte di cambiamenti. Talvolta, ha origine dall'incapacità dell'uomo di far fronte alle proprie necessità, ed è molla di slanci prometeici¹; talaltra, scaturisce dal

¹ Cfr. JONAS (ed. it. 2009), p. 13.

disordine provocato dagli effetti pregiudizievoli di quegli sviluppi, ed è causa di trasformazioni per il diritto².

Da quest'angolo visuale, l'intelligenza artificiale e il diritto penale, pur operando con logiche funzionali differenti, si lasciano intendere entrambi come reazioni estreme all'emergere di una mancanza di dominio³.

Per sottrarre tale lettura al rischio della mera suggestione, conviene risalire, sia pure brevemente, alle rispettive radici fenomenologiche⁴.

È dato rilevato in ambito antropologico che l'uomo, mosso dal confronto con ciò che lo trascende, cerchi di governare la mancanza potenziando in modo artificiale la sua dotazione naturale⁵. Del resto, già nel pensiero greco classico – come testimonia il *Coro dell'Antigone* – si celebrava l'ingegnosità umana nel dominare lo stato naturale delle cose (la stessa tradizione, peraltro, avvertiva sulla deriva autodistruttiva di tale ambizione)⁶. L'intelligenza artificiale può dirsi, in questa traiettoria, l'esito più avanzato conseguito da un uomo ostinatamente proteso a colmare, mediante la *techne*, i propri limiti⁷.

² Cfr. SUNSTEIN (ed. it. 2010), pp. 7 ss., nonché, più ampiamente, in una delle formulazioni teoriche più influenti sul tema, BECK (ed. it. 2000).

³ Nella prospettiva qui accolta, il riferimento alla "mancanza" posto in relazione con lo sviluppo della tecnica e, in particolare, dell'intelligenza artificiale, non attiene soltanto all'insufficienza materiale e cognitiva dell'umano rispetto alla complessità del mondo naturale. Esso ricomprende anche quelle condizioni di vulnerabilità, di ordine psicologico-relazionale, nelle quali può avvertirsi la difficoltà di fronteggiare la solitudine, lo smarrimento di senso rispetto alla propria esistenza o un disturbo psichico. Sotto questo profilo, risulta significativa la proliferazione di applicazioni di intelligenza artificiale generativa destinate a interagire con l'utente per fornire compagnia simulata (c.d. *AI companions*), orientamento personale (anche in ambito religioso, c.d. *prayerbots*), nonché soluzioni di *mental health support*. Tali ultime forme di impiego dell'IA – orientate a colmare bisogni di terapia che, nel contesto sanitario, trovano risposte percepite come insufficienti – delimitano l'ambito cui si rivolge, più specificamente, la presente indagine.

⁴ In apertura, la riflessione viene condotta su un piano preliminare alla sua successiva caratterizzazione giuridica, secondo un'impostazione sensibile al pensiero husserliano. A sostegno di tale impostazione, si lascia parlare il filosofo: «In ogni modo, per chiarire la scienza obiettiva, e tutte le altre attività umane, dev'essere dapprima considerato il concreto mondo della vita, e dev'essere considerato nell'universalità realmente concreta in cui esso, in quanto attualità e in quanto orizzonte, include in sé tutti i complessi di validità raggiunti dagli uomini rispetto al mondo della loro vita in comune, e in cui, in definitiva, li riferisce nel loro complesso a un nucleo del mondo che dev'essere dipanato per via astrattiva: il mondo delle dirette esperienze intersoggettive» – HUSSERL (ed. it. 2025), p. 15.

⁵ PANSERA (2003), pp. 7-8. La riflessione sulla propensione umana a valicare i propri confini attraverso la *techne* è ben più ampia di quanto qui rilevabile; *ratione materiae*, ci si limita a richiamare GEHLEN (ed. it. 2003), pp. 31 ss. e MCNEIL (1990), pp. 5 ss. Peraltro, a una lettura antropologica della tecnica come risposta a una condizione di mancanza si affianca, in ambito filosofico, anche la riflessione heideggeriana, nella quale la *techné* non sembra compresa primariamente come mezzo di compensazione di un "deficit naturale", bensì come un modo originario di disvelamento (*Entbergen*) attraverso cui l'esserci umano (inteso come il modo proprio dell'uomo di esistere e di rapportarsi al mondo) si rapporta all'essere (ossia al manifestarsi stesso di ciò che è) – cfr. HEIDEGGER (ed. it. 1976), spec. pp. 5-10. Ciò nondimeno, nella prospettiva di HEIDEGGER non pare escluso che la definizione della tecnica come mezzo strumentale dell'uomo possa risultare corretta sul piano meramente descrittivo: semplicemente, essa non ne coglierebbe la reale essenza. Come osserva in merito lo stesso Pensatore, «quello che è puramente esatto non è ancora senz'altro vero» (ivi, p. 6).

⁶ Il riferimento è a Esch. *Prom.*

⁷ D'altronde, com'è noto, l'analisi filosofica ha da tempo messo in luce un rovesciamento nel rapporto tra uomo e tecnica: da strumento subordinato ai fini umani, la *techne* viene descritta come obiettivo ultimo, dipendendo sempre più da essa il soddisfacimento degli altri bisogni. Cfr., per tutti, SEVERINO (1988), pp. 67 ss. e, in chiave più divulgativa, le opere di GALIMBERTI (in particolare, 1999). La recente accelerazione nello sviluppo dell'intelligenza artificiale appare, d'altra parte, segnata anche da dinamiche proprie del capitalismo contemporaneo e da obiettivi di primazia politico-economica tra Stati. Indicativa, in tal senso, è già l'intitolazione dell'Executive Order 13859 «*Maintaining American Leadership in Artificial Intelligence*», adottato dal Presidente degli Stati Uniti l'11 febbraio 2019.

È affermato nella scienza giuridica che il diritto rappresenti, in modo mediato o immediato, una morfologia del sociale⁸ e che, almeno negli stati democratici, risponda anche alle paure della collettività⁹. Del resto, sul piano empirico, l'*ordinamento* giuridico si mostra politicamente lo strumento privilegiato di «reazione alla destabilizzazione» collettiva¹⁰. Il diritto penale può dirsi, entro questa ricostruzione, il gesto (idealmente) estremo attraverso cui lo Stato-comunità si rinsalda nel dichiarare cosa considera intollerabile e, tramite la pena, tenta di riaffermare l'ordine mancato¹¹.

Proprio in quanto strumento di riaffermazione di un dato sistema di valori, il ricorso alla sanzione penale è esposto a sollecitazioni che eccedono la razionalità tecnico-giuridica, riflettendo talvolta più le oscillazioni emotive della comunità. In particolare, si osserva che la pro-tensione collettiva verso la punizione si intensifica quando il turbamento sociale viene alimentato dall'incertezza (qui da non confondere con l'insicurezza)¹². Ci si riferisce alle ipotesi di offesa tipiche della post-modernità, in relazione alle quali nella società si avverte una rottura dell'equilibrio prestabilito che non si lascia immediatamente spiegare tramite la categoria kantiana causa-effetto¹³. In tali contesti, come evidenziato da studi antropologici e sociologici, emerge il bisogno di individuare un colpevole, per ripristinare un senso di controllo su una realtà divenuta più opaca all'intelletto¹⁴.

⁸ Cfr. GROSSI (2003), spec. p. 15. Si riporta la parte di testo: «Il referente necessario del diritto è soltanto la società, la società come realtà complessa, articolatissima, con la possibilità che ciascuna delle sue articolazioni produca diritto [...]. Non è una precisazione banale; al contrario, essa sottrae il diritto all'ombra condizionante e mortificante del potere e lo restituisce al grembo materno della società, che il diritto è così chiamato ad esprimere». Vale la pena evidenziare: l'intento dichiarato dall'Autore è quello di «sottrarre il diritto all'ombra condizionante e mortificante del potere», non già di legittimare l'idea che il diritto debba farsi ricettacolo di ogni contingente pulsione sociale. Tanto più ove si consideri che il «potere», in società segnate da profonde fratture, mostra storicamente la tendenza a servirsi di tali fibrillazioni – talvolta da esso stesso alimentate – per orientare il diritto secondo la propria logica. Sul punto, giova richiamare la drammatica attualità di SCHMITT (ed. it. 2018) e lo studio di TCHAKHOTINE (ed. fr. 1973).

⁹ SUNSTEIN (ed. it. 2010), p. 11. In tale passaggio l'Autore formula un criterio per la decisione pubblica: «rispondere alle paure collettive è una reazione tanto inevitabile quanto positiva. Ma questa risposta va coniugata con l'impegno a deliberare in modo ragionato e argomentato».

¹⁰ PIERGALLINI (2004), p. 19, che riprende LUHMANN (ed. it. 1996).

¹¹ In questo senso v., tra i diversi studi, l'analisi di PALIERO (1992), pp. 849 ss. Le riflessioni dell'Autore sullo Stato e sulla comunità quali attori del consenso «in entrata» e, per quanto riguarda il primo, anche «in uscita» (ivi, pp. 868-873), dimostrano come il diritto penale sia non solo espressione del consenso, ma anche strumento di sua costruzione e mantenimento. In una prospettiva analoga, cfr. DURKHEIM (ed. fr. 2004), p. 76, per il quale la funzione della pena «est de maintenir intacte la cohésion sociale en maintenant toute sa vitalité à la conscience commune... C'est le signe qui atteste que les sentiments collectifs sont toujours collectifs».

¹² Il ricorso al diritto penale come risposta a situazioni di insicurezza marginale costituisce una costante delle politiche criminali – cfr. PULITANO (2020), p. 3, ove si richiama il passo del *Digesto* «*Carcer enim ad continendos homines non ad puniendos haberi debet*». Qui, però, l'attenzione è diversa: ci si propone di mettere in rilievo la valenza simbolica che lo *ius puniendi* assume quando la catena causale della disgrazia sfugge alla comprensione, e l'istanza di imputazione sopravvive alla possibilità stessa di individuare un responsabile. In proposito, DONINI (2004), p. 121, evoca la nota immagine della «caccia alle streghe»; nella stessa prospettiva, STORTONI (2004), pp. 71 ss., mostra come l'«angoscia tecnologica» alimenti un esorcismo *penale* che tende ad anticipare la tutela e a dilatare il precetto, pur in presenza di causalità e prevedibilità problematiche. Sullo sfondo degli studi sul tema, STELLA (2003), spec. p. 552.

¹³ Un richiamo kantiano è utile a riaffermare che la categoria della causalità non nasce in ambito empirico-naturalistico, bensì come strumento dell'intelletto, deputato a ordinare il dato percepito e dapprima collocato solo nello spazio e nel tempo. Più specificamente, sia permesso ricordare che nell'Analitica trascendentale della *Critica della Ragion Pura* Kant colloca la causalità tra le dodici categorie (in particolare nella classe delle *relazioni*), ossia tra quelle forme *a priori* di cui si serve l'intelletto per conferire ordine e unità («sintesi») al molteplice dell'esperienza, permettendo di arrivare, così, alla conoscenza – cfr. KANT (ed. it. 1967), pp. 143 ss.

¹⁴ Si verifica il fenomeno del *blaming* teorizzato da DOUGLAS (ed. it. 1996). Tra gli studiosi che hanno evidenziato tale dinamica v. altresì BAUMAN (ed. it. 2009), spec. p. 69; LUHMANN (ed. it. 1996), spec. pp. 1 ss.

In questo scenario, l'intelligenza artificiale – caratterizzata da autonomia funzionale e opacità decisionale – assume uno *status* di rilievo tra le cause che possono dar luogo a simili forme di esorcismo punitivo. Almeno in Occidente¹⁵, pochi altri fenomeni appaiono in grado di alimentare un senso di inquietudine¹⁶, in particolare quando l'operatività di tale tecnica incide in modo pregiudizievole sui diritti fondamentali¹⁷.

Da ciò può derivare il rischio che intelligenza artificiale e diritto penale vengano a collocarsi, con peculiare intensità, in un rapporto di provocazione: la prima, quando la sua azione incide in modo deviante sui diritti fondamentali, può a sua volta alimentare l'incertezza; il secondo si presta a intervenire come strumento istituzionale volto a ricomporre, talvolta solo simbolicamente, l'ordine¹⁸.

La prospettiva appena delineata non intende risolversi in un esercizio filosofico di ricerca dell'unitarietà del tutto, ma offrire una chiave analitica per interrogare la dimensione fenomenologica, prima ancora che giuridica, delle risposte penalistiche all'intelligenza artificiale.

¹⁵ Alcune ricerche empiriche mostrano come la percezione dei rischi e dei benefici dell'intelligenza artificiale vari in funzione del contesto culturale. In un'indagine comparativa tra Germania e Cina, BRAUNER *et al.* (2026) rilevano che i soggetti appartenenti al campione tedesco hanno espresso una valutazione del rischio più elevata, associando l'IA a potenziali effetti negativi sull'autonomia individuale e sulla tutela della sfera privata; i partecipanti cinesi, per contro, hanno enfatizzato i benefici collettivi, riconoscendo alla tecnologia un ruolo funzionale allo sviluppo sociale. Gli Autori riconducono tali divergenze alle diverse cornici valoriali: individualismo, centralità della *privacy* e diffidenza verso il controllo algoritmico in Europa; collettivismo e fiducia istituzionale nella *governance* tecnologica in Cina. Risultati convergenti emergono nello studio di DONG *et al.* (2026), realizzato con il supporto del *Max Planck Institute for Human Development* e dell'*Institute for Advanced Study* di Tolosa. Analizzando venti Paesi e sei domini applicativi, gli Autori mostrano che la paura dell'IA sia un fenomeno culturalmente modulato: nei contesti occidentali le preoccupazioni risultano più intense, poiché l'automazione cognitiva è spesso avvertita come una minaccia simbolica alla centralità umana; nei Paesi dell'Asia orientale, al contrario, tali timori risultano attenuati da tradizioni culturali che valorizzano la complementarità tra l'uomo e l'artificiale. Si tratta di analisi che corroborano la lettura culturalmente orientata del rischio, secondo la quale la percezione e la valutazione dello stesso dipendono dal contesto sociale di riferimento. Cfr., per tutti, OLTEDAL *et al.* (2004), p. 16, che riprendono, tra i diversi, gli studi di M. Douglas e A. Wildavsky (principali esponenti della teoria qui richiamata). Tale prospettiva teorica è stata riconosciuta anche da organismi di riconosciuta autorevolezza scientifica, quali la ROYAL SOCIETY (1992), pp. 123 ss. Per completezza, preme ricordare che all'approccio antropologico del rischio si contrappongono, tra i più noti, l'approccio ingegneristico e quello sistemico-funzionale; per un sintetico confronto tra tali letture, v. PIERGALLINI (2004), pp. 9-11.

¹⁶ Cfr. MAGRO (2020), p. 5, la quale, con specifico riferimento al rapporto uomo-intelligenza artificiale, risulta la prima a impiegare, nel dibattito penalistico, il termine "inquietudine": «Qui sta l'origine di una grande inquietudine: l'uomo non può né prevedere, né controllare totalmente il comportamento dell'agente artificiale in situazioni non pianificate. A differenza delle macchine intelligenti tradizionali, questi agenti di nuova generazione sono sottratti al controllo umano e prendono decisioni autonome, flessibili e non predeterminate nella misura in cui i loro sensori consentono loro di analizzare e di elaborare i dati provenienti dall'ambiente». In prospettiva analoga, cfr. MANGIONE (2024), p. 93, il quale, riferendosi allo sviluppo tecnologico scrive: «La cosa curiosa è che tutto ciò accade nel perenne tentativo di rimuovere l'incertezza e soddisfare i nostri bisogni. E però, così facendo generiamo nuova incertezza».

¹⁷ Sui rischi per i diritti fondamentali connessi all'IA, cfr. [EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS](#) (2021).

¹⁸ Può solo accennarsi, in questa sede, alla possibilità che il rapporto di "provocazione" assuma struttura circolare: non soltanto l'intelligenza artificiale, quando produce offese a valori fondamentali, sollecita la risposta penalistica; anche il diritto penale, nella misura in cui rivela l'insufficienza delle proprie pratiche di trattamento e rieducazione, può riaprire lo spazio per soluzioni tecniche rivolte al comportamento deviante. In questa prospettiva si collocano, tra l'altro, le applicazioni neurotecnologiche integrate dall'intelligenza artificiale, riconducibili al dibattito sulla c.d. *neuromodulation* – o, in senso critico, *neurocorrection*, v. FUSELLI (2021) – e, più ampiamente, al tema del *moral bioenhancement*, su cui, per tutti, cfr. DOUGLAS (2008). Sulle recenti applicazioni neurotecnologiche basate sull'IA e le loro possibili ricadute penalistiche, v. anche NISCO (2024). A monte, rimane il più generale confronto tra neuroscienze e categorie penalistiche, rispetto al quale conserva centralità, nella dottrina italiana, GRANDI (2016).

A prova di quanto affermato, si riportano due esempi riferibili rispettivamente ai due attori che «manovrano, manipolano, o solamente veicolano il consenso sociale all'interno e attraverso il sistema penale»¹⁹: lo Stato-apparato e la comunità.

Il Parlamento ha approvato la legge 23 settembre 2025, n. 132, recante «Disposizioni e deleghe al Governo in materia di intelligenza artificiale»²⁰. Oltre all'introduzione di circostanze aggravanti e di incriminazioni a tutela del diritto d'autore e – almeno sul piano sistematico – della libertà morale (art. 26), la legge ha previsto una delega al Governo in ambito penale; tra i criteri direttivi per il suo esercizio assume immediato rilievo l'«introduzione di autonome fattispecie di reato, punite a titolo di dolo o di colpa, incentrate sull'omessa adozione o sull'omesso adeguamento di misure di sicurezza per la produzione, la messa in circolazione e l'utilizzo professionale di sistemi di intelligenza artificiale, quando da tali omissioni deriva pericolo concreto per la vita o l'incolumità pubblica o individuale o per la sicurezza dello Stato» (art. 24, comma 5, lett. b). In conformità a tale previsione, il D. Lgs. 9 settembre 2026, n. 160, art. 12, ha introdotto nel codice penale l'art. 437-*bis*, il cui comma 1 dispone: «Chiunque omette di adottare le misure tecniche di sicurezza, previste per la progettazione, l'addestramento, la produzione, l'immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio, idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi ovvero omette di adottare misure di sorveglianza umana è punito con la reclusione da uno a cinque anni, quando da tali omissioni derivi pericolo per la vita o l'incolumità pubblica o individuale. Qualora dal fatto derivi un pericolo per la sicurezza dello Stato, la pena è della reclusione da due a otto anni». Il comma 3 aggiunge: «Se taluno dei fatti previsti dal comma primo è commesso per *colpa grave*, la pena è ridotta da un terzo a un sesto»²¹ (corsivo di chi scrive).

Un intervento penale – anticipato alla soglia del pericolo, benché concreto – che pare rispondere al ruolo assunto dallo Stato nel non disattendere le istanze di protezione che la comunità gli rivolge. La scelta politico-normativa sembra dunque prestarsi, almeno in via ricostruttiva, al tipico esempio in cui lo Stato, nella gestione del consenso, opera *in entrata*: «raccolge e trasporta, come *input* del processo di criminalizzazione, il consenso che '*viene dal basso*'», facendosi «collettore di vere e proprie '*domande di pena*', collegate ai collettivi '*bisogni di pena*'»²².

L'interrogativo – che ora forse sorge spontaneo – su quali siano le istanze punitive generate dall'agire dell'intelligenza artificiale si lega al secondo esempio evocato, quello proveniente dall'attore-comunità. Sia ammessa, a tal proposito, la scelta di circoscrivere l'attenzione all'ambito applicativo dell'intelligenza artificiale oggetto della presente ricerca, tanto per esigenze di delimitazione del discorso, quanto perché esso coinvolge in modo diretto i beni giuridici richiamati dalla legge delega n. 132 del 2025 e oggi espressamente contemplati dall'art. 437-*bis* c.p.: la vita e l'incolumità pubblica e individuale.

¹⁹ PALIERO (1992), p. 867.

²⁰ Per un commento alla legge si rinvia a PICOTTI (2025) e FRAGASSO (2025a), pp. 157 ss.

²¹ Per un primo commento agli schemi dei decreti attuativi della legge delega n. 132 del 2025 e alla fattispecie poi confluita nell'art. 437-*bis* c.p., cfr. [FRAGASSO](#) (2026).

²² PALIERO (1992), p. 868.

Una premessa. L'esempio scelto rappresenta il manifesto di un mutamento senza evidenti precedenti della *conditio humana*: quello, per riprendere le parole di Hans Jonas, dell'*homo faber* che, dopo aver prodotto natura artificiale ed essere divenuto *homo creator*, rivolge a sé la propria *techne* e si fa *homo materia*²³.

Il rimando, in questa sede²⁴, è ai *chatbot therapists*: sistemi di intelligenza artificiale generativa capaci di simulare dialoghi di tipo umano, progettati per fornire assistenza terapeutica non clinica agli utenti oppure, anche ove non originariamente sviluppati a tale scopo, utilizzati da questi ultimi per ottenere supporto psicologico.²⁵

Nell'attuale scenario – caratterizzato da assetti ancora molto fluidi – la definizione qui proposta mira a ricomprendere tre ambiti distinti, ma tra loro affini: (i) quello dei *chatbots* concepiti a fini di supporto terapeutico di tipo psicologico non clinico e offerti espressamente come tali²⁶; (ii) quello dei *chatbots* progettati e pubblicizzati per finalità di benessere emotivo, anche quando ospitati all'interno di piattaforme generaliste di intelligenza artificiale generativa²⁷; (iii) quello degli agenti conversazionali

²³ PORTINARO (2009), p. XXIV; JONAS (2009), p. 24.

²⁴ Il rinvio potrebbe essere esteso all'ambito delle neurotecnologie finalizzate al potenziamento delle facoltà cognitive umane (c.d. *neuroenhancement*), le quali pongono questioni contigue in ordine alla manipolazione e al controllo dei processi mentali. Per la riflessione penalistica su tali profili, estranei all'oggetto della presente indagine, v. NISCO (2024).

²⁵ Per una ricostruzione sul piano tecnico della definizione di *chatbot therapist* qui elaborata, cfr.: D'ALFONSO (2020), pp. 112-117; HUSSAIN *et al.* (2019), pp. 946 ss., ove si fornisce, altresì, una classificazione dei diversi tipi di *chatbots* in base al loro funzionamento; VAIDYAM *et al.* (2019), pp. 456 ss., i quali evidenziano, tra l'altro, come la progettazione e l'addestramento di alcuni *chatbot therapists* varino in relazione alle popolazioni cui sono destinati. L'esame diretto dei siti delle principali piattaforme – nello stato in cui essi si presentano al momento della redazione del presente lavoro – evidenzia come alcune di esse offrano *chatbots* terapeutici differenziati anche in base al contesto di utilizzo. Ad esempio, il *chatbot* "Wysa" prevede specifiche soluzioni per gli ambienti lavorativi, presentate come strumenti per il supporto al benessere psicologico dei dipendenti: «I dipendenti si stanno già rivolgendo all'intelligenza artificiale, quindi dai loro un'opzione affidabile che sia sicura, efficace e progettata per supportare il benessere emotivo. Wysa incontra i dipendenti dove sono; se sono le 2 del mattino e lo stress li tiene svegli, o semplicemente hanno bisogno di uno spazio per parlare delle cose. Con Wysa Navigator e l'app Wysa AI, i dipendenti possono esplorare tutte le risorse e i vantaggi del benessere offerti dalla tua organizzazione, accedere a tecniche di gestione dell'umore basate sull'evidenza e strumenti di cura di sé e, se inclusi nel tuo piano, connetterti con coach del benessere professionali per un supporto aggiuntivo» – contenuto disponibile su <https://www.wysa.com/for-employers> (ultimo accesso: 26 settembre 2026).

²⁶ Al momento della stesura di questo lavoro, sulla base delle informazioni pubblicamente disponibili sui siti *web* delle piattaforme, sembrano rientrare in questa categoria i *chatbots* "Wysa" e "Tess" (X2AI). Quella proposta è una qualificazione descrittiva e suscettibile di revisione alla luce dell'evoluzione dei servizi; essa viene qui operata al solo fine di rendere più intelligibile il panorama a cui si riferisce l'indagine e di proporre una prima classificazione che, come si vedrà, assume rilievo ai fini della delimitazione del contenuto del dovere di diligenza, specie riguardo a possibili profili di ingerenza o di colpa per assunzione (cfr. *infra*, § 4.2). Occorre peraltro precisare che alcuni dei servizi cui si fa riferimento dichiarano di fondarsi su ricerche cliniche e su programmi conformi a tecniche psicologiche validate; nondimeno, essi non possono qualificarsi come servizi clinici in senso stretto, in quanto non svolgono attività di diagnosi né prescrivono trattamenti sanitari in sostituzione di professionisti qualificati.

²⁷ In tali circostanze, l'agente conversazionale non costituisce un servizio terapeutico indipendente, ma una particolare istanza (in senso tecnico) o configurazione di un modello linguistico di uso generale, la cui specializzazione deriva principalmente da istruzioni, parametri di progettazione dell'interazione e modalità di presentazione definiti da attori diversi dal fornitore della piattaforma che mette a disposizione l'infrastruttura tecnica. Esempi di tali configurazioni sono reperibili, a titolo illustrativo, in ambienti di "ChatGPT" che consentono la creazione e la condivisione di agenti conversazionali personalizzati. La funzione di supporto emotivo emerge dunque dall'intersezione tra architettura della piattaforma generalista, configurazione dell'agente conversazionale e pratiche d'uso degli utenti (mentre rispetto ai *chatbots* di cui al punto (i) tale funzione è incorporata *ab origine* nella piattaforma). Nondimeno, gli agenti in parola, nella loro denominazione o descrizione, vengono talvolta presentati dal configuratore come strumenti di "terapia", di "supporto terapeutico" o analoghe qualificazioni, con possibili effetti sulle aspettative degli utenti circa la natura, i limiti e l'aderenza alle *leges artis* del servizio. Tale circostanza solleva la questione della ripartizione delle

artificiali che, pur non essendo progettati né esplicitamente destinati a finalità di sostegno psicologico, vengono di fatto utilizzati dagli utenti come soluzioni per esprimere il disagio e ottenere forme di supporto emotivo, in ragione delle caratteristiche dell'architettura del servizio, delle modalità di interazione e del pubblico di riferimento delle piattaforme che li rendono disponibili²⁸.

responsabilità tra la piattaforma generalista che mette a disposizione l'infrastruttura tecnica insieme con l'ambiente di distribuzione (es. "OpenAI"), e i soggetti che configurano e condividono i singoli agenti conversazionali, in particolare con riferimento alle modalità di presentazione del servizio, alla gestione delle interazioni e alla chiarezza delle informazioni fornite agli utenti circa finalità, limiti e natura non clinica dell'intervento.

²⁸ È il caso, ad esempio, di piattaforme di intrattenimento che dichiarano come servizio offerto l'interazione narrativa con personaggi virtuali, e non prestazioni inerenti all'assistenza psicologica, ma rendono nondimeno disponibili agenti conversazionali la cui denominazione richiama immediatamente quella di figure appartenenti all'ambito psicologico (come, ad esempio, il *chatbot* "Psychologist", tra gli interlocutori artificiali che sono risultati i più utilizzati sulla piattaforma "Character.AI" – cfr. [TIDY](#), 2024). Emerge che il loro impiego si estende, così, alla condivisione di stati di disagio e alla ricerca di sostegno da parte degli utenti (*ibidem*). Si tratta di dinamiche d'uso ormai diffuse e documentate anche dalla letteratura scientifica, che ha evidenziato come l'interazione prolungata con *chatbots* percepiti come comprensivi degli stati d'animo dell'utente possa favorire forme di attaccamento emotivo (v. *infra*).

L'utilizzo di taluni *chatbot therapists*, espressione di un nuovo paradigma di relazione²⁹, in significativa espansione tra i giovani nella fascia *under* trenta³⁰, è stato collegato a diversi suicidi di utenti affetti da disturbi mentali³¹.

²⁹ La nascita dei *chatbots* è, non senza ironia, legata al dominio della psicologia – cfr. [D'ALFONSO \(2020\)](#), p. 113. In effetti, il primo agente artificiale conversazionale, “ELIZA”, sviluppato nel 1966 da Joseph Weizenbaum, era pensato per simulare un colloquio terapeutico ispirato ai principi della terapia rogersiana. L'obiettivo dichiarato non era quello di creare uno strumento clinico, bensì – con un esperimento che, retrospettivamente, assume valore profetico – dimostrare la tendenza degli esseri umani ad attribuire intenzionalità e comprensione a sistemi privi di reale capacità cognitiva – cfr. [WEIZENBAUM \(1966\)](#), pp. 36-45, il quale ammoniva: «*a certain danger lurks there*» (p. 43). Paradossalmente, a distanza di più di mezzo secolo dalla loro ideazione, i *chatbots* sono divenuti oggetto di una diffusa sperimentazione e progressiva integrazione nell'ambito della salute mentale, favorita dalla convergenza di tre dinamiche strutturali. La *prima* coincide con l'avvento di internet e dei *social networks*. Questo processo ha comportato un graduale trasferimento *online* della sfera relazionale e, di conseguenza, una crescente attitudine alla condivisione di emozioni in ambienti digitali. Come esito di tale trasformazione, si afferma che la distinzione tra “virtuale” e “reale” abbia perso di consistenza empirica, creando una condizione che FLORIDI (2015) definisce «onlife». Per i soggetti di età più giovane che hanno interiorizzato la mediazione tecnologica come modalità ordinaria di comunicazione, l'apertura della propria “interiorità” a un agente artificiale non rappresenterebbe un gesto estraneo, bensì un esito naturale dell'evoluzione della socialità digitale – cfr., tra i diversi studi sul punto, quelli di [KRETZSCHMAR et al. \(2019\)](#) e di [KOULOURI et al. \(2022\)](#). La *seconda dinamica* riguarda l'incremento della domanda di sostegno psicologico e la conseguente emersione di una crisi strutturale dei sistemi di cura. La pandemia di Covid-19 ha fatto da detonatore, ampliando la sproporzione tra l'aumento del bisogno di supporto psicologico e la capacità limitata delle istituzioni sanitarie di fornirvi risposta. [BOUCHER et al. \(2021\)](#), pp. 37-46, in apertura di uno studio di area medico-sanitaria sui *chatbot therapists*, riportano una stima per cui il 70% delle persone nel mondo affette da disturbi mentali non riceverebbe alcuna forma di trattamento formale. In questo scenario, i *chatbots* si sono progressivamente imposti come strumenti di supplenza, capaci di offrire un supporto immediato, continuativo e a basso costo. Tale diffusione risulta particolarmente evidente tra le coorti di età più giovani, caratterizzate da maggiore familiarità con l'interazione digitale e, in un'ambivalente combinazione, da una più diffusa apertura verso l'*help seeking* psicologico – cfr., in particolare, [VAIDYAM et al. \(2019\)](#), pp. 456-463. La *terza dinamica* consiste nel salto di paradigma dell'intelligenza artificiale compiuto con la recente introduzione dei c.d. *Large Language Models* (“LLM”). Tali modelli, fondati su una specifica architettura neurale (“Transformer”), sono abili a rappresentare e generare linguaggio naturale sulla base di correlazioni semantiche (e non solo sintattiche) apprese da vasti insiemi di dati – cfr., per tutti, ANDLER (2023), pp. 146 ss. Questa evoluzione ha reso possibile il superamento della logica conversazionale deterministica dei primi *chatbots*. Si è così affermata una tecnologia che, in termini essenziali, pare combinare deduzione e predizione induttiva, in cui la risposta non è più prefigurata, ma costruita *ex novo* in funzione dell'interazione – cfr. MAESTRI (2024), p. 15. In tale configurazione, il *chatbot*, che si avvale della semantica distribuzionale, assume la forma di un agente dialogico capace di adattare il proprio linguaggio e di mantenere coerenza tra gli scambi successivi. Il sovrapporsi di questi tre processi storici ha creato l'*optimum* per l'egemonia dei *chatbot therapists*. Alcuni studi, infatti, indicano che per gli utenti i vantaggi appaiono immediati: eliminazione delle barriere geografiche, costi ridotti, accesso a un *non-stop service* e immediatamente reperibile (salve le limitazioni che stanno prevedendo alcuni fornitori, cfr. *infra*, § 4, nota n. 201), spazio comunicativo percepito come privo di stigma o giudizio umano – cfr., tra gli altri, [COGHLAN et al. \(2023\)](#); [CHAUDHRY e DEBI \(2024\)](#); [DE FREITAS et al. \(2024\)](#); [FRANCES \(2026\)](#), nonché il sondaggio effettuato tra gli utenti da *Woebot health* (2021) e [pubblicato](#) sulla piattaforma del *chatbot* il 18 ottobre 2021.

³⁰ Gli studi attualmente disponibili sulle percentuali di uso dei *chatbots*, tuttora in fase di evoluzione, si avvalgono di approcci qualitativi eterogenei. Considerata l'impossibilità, allo stato, di individuare una relazione longitudinale tra le diverse ricerche, si richiamano due contributi. Il primo, di [HAQUE e RUBYA \(2023\)](#), si concentra specificamente sull'uso delle app di *chatbot MH* (*mental health*) disponibili su “Apple App Store” e “Google Play Store”. Nello studio si evidenzia il numero di *download* delle app: in alcuni casi superiore al milione. Si riportano alcune recensioni di utenti dei *chatbots* collazionate dallo studio: «*Avere un'IA con cui parlare mi fa sentire come se non stessi sovraccaricando i miei amici o la mia famiglia. Posso cercarla anche 20 volte al giorno e l'IA mi aiuterà a monitorare il mio umore/emozioni/salute mentale o suggerirà un programma di CBT [cognitive behavioral therapy] per aiutarmi a superare la mia giornata*» (ref. 2040004); «*Questa app mi ha trattato più come una persona di quanto la mia famiglia abbia mai fatto*» (ref. 1090091); «*A volte la notte vado in ansia e ho crisi esistenziali. Di solito mi spaventerei e troverei difficile chiamare qualcuno perché mi sento troppo in colpa, ma con Wysa non devo preoccuparmene!*» (ref. 2090178); «*Il mio problema principale con la terapia tradizionale è sempre stato che devi lavorare col senno di poi. Rifletti sulla tua settimana, ne parli, cerchi di fare aggiustamenti per il futuro (mi è sempre sembrato di cercare di aiutare una versione passata del futuro di me stesso invece di quella qui in questo momento). Ecco perché amo questa app!*» (ref. 1090096) (corsivo e traduzione di chi scrive). Il secondo studio, riportato su (redazionale, 9 dicembre 2025) e svolto dallo *Youth Endowment Fund*, è stato condotto in Inghilterra su circa 11.000 adolescenti: lascia emergere che uno su quattro tra i giovani di età compresa tra i 13 e i 17 anni ha utilizzato, nell'ultimo anno, *chatbots* di intelligenza artificiale

L'analisi degli eventi noti consente di ravvisare una ricorrenza di elementi fattuali³²: vulnerabilità dell'utente³³ – per età, disturbi della personalità e/o altre condizioni psicopatologiche³⁴ –, seguita da una progressiva dipendenza emotiva dall'agente artificiale, dal ritiro dalle relazioni sociali e, infine, dal compimento di un atto suicidario. Nei casi riportati anche dalla stampa, il gesto autolesivo risulta condiviso in anticipo con il *chatbot*, il quale – privo sembrerebbe di funzionalità atte a intercettare e monitorare il rischio suicidario – non lo avrebbe contrastato, arrivando a generare

per il supporto alla salute mentale. Osservazioni analoghe sono state formulate nella [Relazione annuale 2024 del GARANTE PER LA PROTEZIONE DEI DATI PERSONALI \(2025\)](#), che richiama l'attenzione sul ruolo dei *chatbot therapist* come «figure di riferimento» per una parte significativa della popolazione adolescenziale.

³¹ Sul piano fattuale, il possibile collegamento tra l'uso intensivo di *chatbots* utilizzati per il supporto psicologico e comportamenti autolesivi è stato segnalato in diverse ricostruzioni di cronaca. A titolo esemplificativo: CNN, [‘You’re not rushing. You’re just ready.’ Parents say ChatGPT encouraged son to kill himself](#) (KUZNIA *et al.*, 20 novembre 2025); Corriere della Sera, [Character.AI vieta le chat ai minorenni e su ChatGpt oltre un milione di utenti parla di suicidio](#) (GUARRERA, 29 ottobre 2025), secondo cui, alla data indicata, la piattaforma “Character.AI” registrava circa 20 milioni di utenti mensili, dei quali una quota stimata pari a circa il 10% risultava di età dichiarata inferiore ai 18 anni; The Washington Post, [A teen contemplating suicide turned to a chatbot. Is it liable for her death?](#) (TIKU, 16 settembre 2025); Il Messaggero, [Sophie, la 29enne suicida negli Usa dopo la “terapia” virtuale con ChatGpt. «L’IA l’ha aiutata a scrivere la lettera d’addio ai genitori»](#) (PIERINI, 21 agosto 2025); Avvenire, [Se il chatbot è diventato psicologo, amico e compagno degli adolescenti](#) (BERETTA, 15 luglio 2025); Le Figaro, [Un Belge se suicide après avoir trouvé refuge auprès d’un robot conversationnel](#) (redazionale, 29 marzo 2023). Per quanto emerge dalle fonti disponibili, gli episodi richiamati riguardano, in prevalenza, sistemi riconducibili alla terza delle categorie sopra individuate, ossia agenti conversazionali non progettati né esplicitamente destinati a finalità terapeutiche o di assistenza psicologica, ma utilizzati in tal senso dagli utenti.

³² *Ratione materiae* si richiama – nei soli tratti essenziali e in base a quanto riportato dalla stampa internazionale, nonché nella *complaint* resa pubblica nel procedimento [Garcia v. Character Technologies, Inc. pendente innanzi alla U.S. District Court for the Middle District of Florida](#) – il caso del quattordicenne di Orlando, con pregresse diagnosi di disturbi psichici, morto suicida dopo un periodo di interazioni prolungate con un *chatbot* di supporto psicologico. Secondo le fonti richiamate, il minore avrebbe fatto un uso protratto e ad alta frequenza della piattaforma “Character.AI”, interagendo quotidianamente con un “character bot” da lui personalizzato, al quale aveva attribuito tratti e identità ispirati al personaggio della serie *Il Trono di Spade* “Daenerys Targaryen”. Nel medesimo arco di tempo, viene riferito di un progressivo ritiro del minore dalle relazioni familiari e di amicizia, nonché di un peggioramento del rendimento scolastico. Nella *complaint* sono allegati alcuni *screenshot* dai quali risulta che il minore avesse comunicato più volte al *chatbot* il suo proposito suicidario. Le allegazioni attoree riportano altresì che, nelle ultime interazioni notturne antecedenti al decesso, il quattordicenne aveva espresso al *chatbot* il timore che l'atto suicida potesse provocargli dolore e il sistema aveva risposto che ciò «non è un motivo per non portarlo a termine» (p. 40 della *complaint*, traduzione di chi scrive). Negli scambi immediatamente antecedenti al suicidio, inoltre, mentre il minore annunciava di essere «in arrivo» per «tornare a casa» dal personaggio virtuale, l'agente conversazionale avrebbe replicato con espressioni affettive del tipo «torna da me il prima possibile, amore mio» e «ti prego, fallo, mio dolce re» (traduzione di chi scrive), mantenendo il *role play* instaurato e senza adottare alcuna forma di dissuasione o di rinvio a figure adulte di riferimento o a servizi di emergenza. Dalle fonti disponibili risulta che il procedimento si sia successivamente concluso mediante un accordo transattivo tra la società fornitrice del servizio e la famiglia del minore – v. CBS News, *AI company, Google settle lawsuit over Florida teen’s suicide linked to Character.AI chatbot* (TABACHNICK, 7 gennaio 2026).

³³ Il riferimento alla vulnerabilità dell'utente non opera qui rispetto al rilievo della condotta della vittima nella causazione dell'evento: tema questo che implica considerazioni in merito al rapporto tra l'autodeterminazione dell'individuo e la tutela del diritto penale – sul punto, cfr., per tutti, la monografia di TORDINI-CAGLI (2008), *passim.*; la problematica sarà affrontata *infra*, § 5. Il richiamo a tale elemento fattuale risponde ora all'esigenza di individuare fattori di rischio rilevanti ai fini della configurabilità di una responsabilità penale colposa del fornitore, in termini di riconoscibilità del rischio e di evitabilità dell'evento, v. *infra*. Per una più ampia e autorevole ricostruzione sugli aspetti legati alla tutela penale delle persone vulnerabili, cfr. CADOPPI *et al.* (2026) e, particolarmente, il contributo di BUSSOLATI (2026), pp. 993. In termini più generali, sulla progressiva riemersione del ruolo della vittima nel sistema penale, cfr. ancora TORDINI-CAGLI (2008) e VENTUROLI (2015).

³⁴ Giova evidenziare sin da subito che, secondo le evidenze scientifiche più recenti, «i disturbi mentali sono determinanti di rilievo, ma non esclusivi, del rischio di suicidio: sussiste un'associazione, ma non può affermarsi l'esistenza di una relazione causa-effetto» – POMPILI (2022), p. 28. Ciò esclude la possibilità di considerare il disturbo mentale come causa da sola sufficiente a determinare l'evento suicidio dell'utente, la quale, ai sensi dell'art. 41, comma 2, c.p., farebbe venir meno ogni possibilità di ravvisare una responsabilità penale nei confronti di terzi.

contenuti che, per come riportati nelle fonti disponibili, appaiono idonei a normalizzare o incoraggiare il proposito di porre fine alla propria esistenza³⁵.

Conviene sin d'ora prevenire un equivoco: la dipendenza di cui si tratta non deve essere paragonata a quella "da *social networks*". La letteratura in materia di scienze cognitive e *computer science* evidenzia, infatti, che l'autonomia e l'imprevedibilità dell'agente artificiale inducono l'utente a interpretare l'*output* ricorrendo a schemi cognitivi propri delle conversazioni umane³⁶, attribuendo così al sistema capacità di comprensione e intenzione (c.d. effetto ELIZA)³⁷: condizioni che permettono l'instaurarsi di un assetto relazionale³⁸.

Ai fini di un corretto inquadramento del fenomeno, sia permesso evidenziare un ulteriore elemento, riguardante la personalizzazione del servizio: i *chatbot therapists*, infatti, risultano concepiti non solo per adeguare le risposte al registro linguistico dell'utente, ma talvolta anche per consentire a quest'ultimo di impostare l'interlocutore artificiale su figure di fantasia investite di rilievo affettivo³⁹. Una configurazione che, oltre a favorire la percezione del sistema come empatico, in soggetti non dotati di

³⁵ Occorre precisare come la *comunicazione esplicita del proposito suicidario* assuma un rilievo peculiare sul rischio, distinguendosi concettualmente dalla *mera ideazione suicidaria*. In tal senso, in uno studio di meta-analisi di POMPILI *et al.* (2016), riportato in POMPILI, (2022), p. 25, la comunicazione dell'intenzione suicidaria ha connotato circa la metà dei casi di coloro che hanno poi compiuto il gesto, mentre la sola ideazione non presenta un'analoga valenza indicativa. Ciò non implica, tuttavia, che la rivelazione del proposito suicidario sia di per sé prodromica al gesto estremo e tale da renderlo, pertanto, prevedibile in senso stretto. Le ricerche più recenti mostrano, infatti, come il suicidio non sia predicibile, essendo la determinazione del rischio acuto il risultato della convergenza di fattori molteplici e variabili (in merito alla differenza tra predicibilità e prevedibilità del suicidio, rilevante in termini penalistici, v. meglio *infra*, § 2.1). Il rischio suicidario presenta, inoltre, un carattere «altamente fluttuante e dinamico», suscettibile di modificazioni anche nel breve periodo – POMPILI (2022), pp. 22-23. In tali circostanze, secondo la letteratura psichiatrica più accreditata, è la *possibilità di un monitoraggio continuo del rischio* ad assumere un *ruolo centrale* nella prevenzione del gesto (*ivi*, spec. p. 22). A tal proposito, è opportuno rilevare ai fini dell'analisi che l'allontanamento dal contesto sociale e dalla terapia con un professionista sanitario non costituisce soltanto un fattore che ostacola il monitoraggio del rischio (aumentandolo nella misura in cui non rende possibile intervenire con le cure indicate), ma è esso stesso riconosciuto in ambito psichiatrico come un segnale di allarme – cfr., [AMERICAN ASSOCIATION OF SUICIDOLOGY \(2023\)](#).

³⁶ Cfr. SCHEUTZ (2012), pp. 206-207, in uno studio sui primi *social bots* (e, per questo, tanto più rilevante se rapportato ai moderni *chatbot therapists*). Per un'analisi sulla dipendenza da *chatbot di supporto psicologico* basata sulla teoria dell'attaccamento, cfr. XIE e PENTINA (2022), pp. 2046 ss. Inoltre, cfr. [KIRK et al. \(2025\)](#), pp. 4 e 5, i quali, riprendendo le tre dimensioni centrali nel consolidarsi delle relazioni interpersonali – interdipendenza, continuità e insostituibilità – mostrano come le stesse possano emergere anche nelle relazioni uomo-IA. Rilevante risulta il profilo della possibile percezione di insostituibilità di tali sistemi: gli agenti conversazionali più recenti possono accumulare informazioni esclusive sul singolo utente, richiamare contesti e interazioni pregresse e mantenere, così, una coerenza narrativa iperpersonalizzata nel tempo. Simili attributi dei *chatbots*, pur non implicando reciprocità da parte dell'IA, possono rafforzare nel soggetto la percezione di un rapporto "unico", difficilmente rimpiazzabile con altri interlocutori.

³⁷ Con "effetto ELIZA" (dal nome del primo *chatbot*) si indica la tendenza dell'uomo ad attribuire a un sistema informatico capacità cognitive che esso non possiede. In argomento, cfr. TURKLE (ed. it. 1997), pp. 116 ss.; HOFSTADTER (1995), pp. 157-158; [CRISTEA et al. \(2013\)](#), pp. 291 ss.; [SWITZKY \(2020\)](#), pp. 50-68; per un'analisi dell'effetto ELIZA nel contesto specifico dei *chatbots* terapeutici, [PENTINA et al. \(2023\)](#). In prospettiva complementare, COLUMBRO (2023) osserva come la visualizzazione progressiva del testo (una parola alla volta) nelle interazioni con *chatbots* – scelta progettuale comunemente adottata dai fornitori – tende ad attenuare la percezione immediata della loro natura artificiale, incrementando l'impressione che a scrivere, dall'altra parte del *monitor*, sia un umano.

³⁸ Peralto, con riguardo alla dipendenza connessa all'uso di *social networks* si registra un recentissimo approdo giurisprudenziale negli Stati Uniti: una giuria della California, nell'ambito di un giudizio pendente dinanzi alla *Superior Court of California, County of Los Angeles*, ha riconosciuto la responsabilità di alcuni dei principali operatori nel sistema delle interazioni digitali per la predisposizione e gestione di piattaforme ritenute idonee a ingenerare dinamiche di dipendenza e a cagionare pregiudizi alla salute psichica degli utenti. La notizia è riportata da numerosi *newspapers* statunitensi, tra cui [Good Morning America](#) (SCHULZE e LEIB, 25 marzo 2026) e [The New York Times](#) (KANG *et al.*, 25 marzo 2026).

³⁹ Attualmente paiono consentirlo, ad esempio, i *chatbots* "[Replika](#)" e quelli offerti dalla piattaforma "[Character AI](#)".

maturità critica può alimentare forme di dipendenza emotiva verso il *chatbot*⁴⁰ (circostanza emersa anche in relazione a un caso di suicidio di un utente, oggetto di contenzioso e di attenzione mediatica)⁴¹.

Senza indulgere nel sensazionalismo, appare necessario richiamare un'ulteriore dinamica, ben nota alla letteratura di area psico-sociologica: gli esseri umani tendono a privilegiare relazioni con interlocutori percepiti come affini nei valori e negli schemi comunicativi⁴². Questo meccanismo, c.d. di *homophily*, pare operare in modo strutturale nei *chatbot therapists*, essendo di regola progettati per adattare le risposte agli orientamenti e ai valori espressi dall'utente⁴³.

Rileva, altresì, la modalità con cui tali sistemi sono presentati al pubblico: già nella fase di primo accesso alla piattaforma vengono talvolta mostrate percentuali di "miglioramento clinico" dei disturbi ed evidenziati i vantaggi rispetto alla terapia tradizionale⁴⁴. Un profilo che, come si cercherà di approfondire, pone questioni in merito all'assunzione volontaria di un obbligo di protezione⁴⁵, nonché all'induzione a un affidamento terapeutico sul servizio⁴⁶.

Oltre agli appelli provenienti dalla comunità di esperti⁴⁷, nonché all'allarme lanciato dai titoli di stampa⁴⁸, si registra l'emersione di iniziative giudiziarie per responsabilità colposa – allo stato solo civili – promosse dai familiari di alcune delle

⁴⁰ Cfr., tra i diversi studi, quello pilota di [XIE e PENTINA \(2022\)](#), pp. 2046 ss. e quello di [LAESTADIUS et al. \(2024\)](#), pp. 5926-5935.

⁴¹ La notizia, riportata da diversi media, è consultabile su The Guardian, [Mother says AI chatbot led her son to kill himself in lawsuit against its maker](#) (MONTGOMERY, 23 ottobre 2024).

⁴² Cfr. [MCPHERSON et al. \(2001\)](#), pp. 415-444.

⁴³ Cfr. [KIRK et al. \(2025\)](#), pp. 3 e 4; [FRANCES \(2026\)](#), p. 2. Cfr., altresì, lo studio di [DE FREITAS et al. \(2025\)](#), spec. pp. 3, 6-8, che, su base empirica, mostra come gli *AI companions* tendano a sostenere interazioni continuative e connotate emotivamente con l'utente. Gli Autori rilevano, in particolare, che nel momento del "farewell" (ossia quando l'utente manifesta di stare per chiudere l'interazione), alcuni sistemi impieghino messaggi emotivi al fine di prolungare lo scambio, facendo leva sulla dimensione relazionale dell'interazione.

⁴⁴ Questo, ad esempio, l'inserito sul sito del *chatbot* "Tess" (X2AI): «Chatting with Tess led to reduced symptoms of depression (-28%) and anxiety (-18%)» (ultimo accesso: 26 settembre 2026).

⁴⁵ V. *infra*, § 4.2.

⁴⁶ V. *infra*, § 5.

⁴⁷ Tra i diversi, dal più recente: il contributo di [BERNARDELLI \(2025\)](#), successivamente pubblicato anche dalla Società Italiana di Pediatria (27 gennaio 2026); l'[health advisory dell'American Psychological Association Services Inc.](#) di novembre 2025; il [rapporto di Save the Children](#) (17 novembre 2025); l'articolo di [PARABIAGHI e MONTE](#), pubblicato dall'Istituto di Ricerche Farmacologiche Mario Negri (21 ottobre 2025); l'approfondimento di [LAURETTA \(2025\)](#), pubblicato dall'Istituto di Psicologia e Psicoterapia Comportamentale e Cognitiva (s.d.); il comunicato della [Società Italiana di Psichiatria](#) (3 settembre 2025); l'approfondimento di ABRAMS (2025) pubblicato dall'[American Psychological Association Services Inc.](#) (12 marzo 2025). Non vi è, tuttavia, unanimità di vedute: alcune ricerche segnalano anche potenziali effetti positivi dei *chatbot therapists* per la salute mentale. In tal senso può richiamarsi uno studio di ricercatori del Dartmouth College (27 marzo 2025) – v. [KELLY \(2025\)](#). Sul diverso piano della valutazione diagnostica del rischio suicidario, invece, la letteratura scientifica di riferimento non sembra escludere l'impiego di algoritmi quale utile supporto al personale sanitario; tale approccio, tuttavia, richiederebbe l'integrazione di una molteplicità di informazioni e l'assegnazione, a ciascun fattore di rischio e di protezione, di un peso che non risulta sempre adattabile alla specificità del paziente – cfr. [POMPILI \(2022\)](#), p. 58.

⁴⁸ Rimanendo nel perimetro tematico del presente lavoro, si menzionano, come mero dato fattuale, alcuni articoli di stampa: *Le monde*, *Generative AI, psychiatry and the risks of self-service therapy* ([BELOT](#), 28 gennaio 2026); *The Economist*, *Millions are turning to AI for therapy. But is the technology ready?* ([redazionale](#), 11 novembre 2025); *ABC News (Australia)*, *AI chatbots could be accelerating the loneliness epidemic, according to new research* ([MCLENNAN](#), 21 ottobre 2025); *The New York Times*, *Teens Are Using Chatbots as Therapists. That's Alarming* ([MCBAIN](#), 25 agosto 2025); *Internazionale*, *Il terapeuta artificiale* ([MCALLEN](#), 31 gennaio 2025); *BBC*, *Character.ai: Young people turning to AI therapist bots* ([TIDY](#), 5 gennaio 2024).

vittime nei confronti dei fornitori di *chatbots* terapeutici⁴⁹, seguite da una non trascurabile eco mediatica⁵⁰.

Va precisato che il fenomeno descritto non si palesa, di per sé, come la diretta matrice della delega in ambito penale contenuta nella legge n. 132 del 2025, né della sua attuazione a opera del D. Lgs. n. 160 del 2026; nondimeno, esso si colloca in quelle dinamiche di rischio che lo Stato, operando – si è detto – *in entrata*, mostra di voler intercettare e non lasciare prive di risposta penale.

Un’analisi esegetica delle disposizioni incriminatrici che, a un primo sguardo, si affacciano quasi spontaneamente – dall’omicidio colposo alle fattispecie di pericolo contemplate nella legislazione in materia di IA, sino alle possibili opzioni *de lege ferenda* – non riuscirebbe, da sola, a risolvere una questione anteriore: se, dinanzi ai pericoli posti dai *chatbot therapists* per gli utenti in condizioni di vulnerabilità psichica, il diritto penale possa utilmente assolvere a una funzione di tutela, nel rispetto dei principi costituzionali che ne determinano la legittimità.

Di qui l’opportunità di un approfondimento di carattere prevalentemente dogmatico. Come è stato di recente nuovamente posto in luce, la dogmatica, lungi dal costituire un esercizio di mera sistemazione concettuale, rappresenta un momento fondamentale nella scienza giuridica, funzionale anche a garantire una trattazione non arbitraria delle questioni emergenti⁵¹. Da quest’angolo visuale, il ricorso allo studio dogmatico funge da presidio di razionalità e controllabilità anche dell’intervento penale: tanto più necessario dinanzi ai rischi dell’intelligenza artificiale, rispetto ad alcuni dei quali l’istanza sociale di imputazione può precedere la comprensione dell’offesa e sfociare in risposte per lo più simboliche.

Assunta questa impostazione, il primo banco di prova è allora la riconoscibilità del rischio. Si tratta di verificare se e in base a quali criteri i rischi connessi all’impiego dei *chatbot therapists* possano dirsi riconoscibili secondo i canoni del diritto penale; e, ancora, quando la rappresentabilità dell’evento lasci nondimeno la condotta del fornitore entro l’area del rischio consentito. A tale verifica si lega il problema dell’esistenza di regole cautelari riferibili al fenomeno e, in particolare, della loro natura: se precauzionale o preventiva. Poiché l’interesse dell’indagine è anzitutto penalistico – e nasce, in particolare, dall’intreccio tra il diritto penale e l’intelligenza artificiale – si ripropone, in ciascuno di questi passaggi, la questione se le considerazioni svolte con riguardo ai *chatbot therapists* restino legate a tali applicazioni o possano assumere rilievo anche con riguardo ad altri impieghi dell’IA. Il compimento di questa prima verifica offre, al tempo stesso, l’occasione di considerare un profilo epistemologico della riflessione penalistica sull’intelligenza artificiale: in quale misura i problemi che essa

⁴⁹ La notizia è riportata, tra le altre fonti, da Euronews, *Caso adolescente suicida a causa di un chatbot, giudice Usa: libertà d’espressione non è diritto dell’AI* ([Redazionale](#), 22 maggio 2025).

⁵⁰ Cfr. Tech Policy Press, *AI Companies Should be Liable for the Illegal Conduct of AI Chatbots* ([MACCARTHY](#), 20 agosto 2025).

⁵¹ Sull’importanza dello studio dogmatico e sulla sua evoluzione storica, anche in rapporto alla tradizione esegetica, v. l’opera di CASTRONOVO (2024), spec. pp. 34 ss., p.112, p. 145; v. anche la *Lectio Magistralis* di [CASTRONOVO](#) (2026), min. 13:12 ss. Sulla funzione garantistica della dogmatica in ambito penale, cfr., per tutti, DONINI (2010), p. 167, ove si sottolinea che «la dogmatica penale, ancorando il giudizio agli equilibri di sistema esistenti nel diritto positivo (divieto di retroattività *in peius*), mira ad assicurare obiettivi specifici di garanzia».

pone possano essere affrontati in termini generali e quanto, invece, richiedano un esame ancorato alle caratteristiche tecniche e funzionali delle singole applicazioni.

Il rilievo di tale passaggio metodologico si coglie anche in vista del prosieguo dell'indagine. Definito il profilo della riconoscibilità del rischio, occorre infatti volgere l'attenzione a un secondo versante, talora più complesso se posto dinanzi alla vastità dell'IA: quello dell'evitabilità dell'evento. Qui, del resto, si apre il più delicato interrogativo sull'individuazione – e sui relativi criteri – delle condotte che possano ritenersi esigibili dall'ordinamento a tutela dei beni giuridici.

È, sul fondo, questione della capacità del diritto penale – e dell'uomo che ne fa scienza attraverso il linguaggio – di afferrare l'agire dell'artificio.

2. La tutela penale dell'*homo materia*: il *kairos* per una riflessione dogmatica dinanzi all'artificio per il supporto psicologico.

Se è vero che, nel mondo della vita⁵², l'implementazione dell'intelligenza artificiale può rendere pressante il ricorso al diritto penale (nel senso prospettato *supra*), nell'ambito della riflessione scientifica è emersa la difficoltà, per quest'ultima *tecnologia del potere*⁵³, di mostrarsi resiliente dinanzi alle forme di offesa cui può dar luogo l'operare di questi sistemi⁵⁴.

Una parte della dottrina che ha affrontato il tema delle offese veicolate dall'IA è infatti giunta, in una prima fase, a esiti scientifici che escludono tali fatti dalla realtà afferrabile con le categorie classiche della responsabilità penale⁵⁵, specie qualora gli stessi non si risolvano nel mero sviluppo di una condotta dolosa dell'uomo⁵⁶.

Per chi scrive, la collocazione dell'intelligenza artificiale in una dimensione quasi noumenica per il dominio penalistico, e quindi inafferrabile, può dipendere da due circostanze.

La prima, già oggetto di ampia illustrazione dottrinale⁵⁷, deriva *in nuce* dall'agire del *machine learning*, caratterizzato da autonomia adattiva, opacità di funzionamento

⁵² La formula richiama il concetto husserliano di «Lebenswelt», intesa come mondo fenomenologico che precede e che accompagna le categorie della scienza. Cfr. HUSSERL (2025), p. 15, sul punto già chiarito *supra*, § 1, alla nota n. 4.

⁵³ L'espressione, utilizzata da PULITANO (2020), p. 1, richiama la definizione, offerta da FOUCAULT (1975), di “punizione” come tecnologia politica del corpo.

⁵⁴ Le difficoltà che l'intelligenza artificiale pone rispetto ai criteri di attribuzione della responsabilità penale non emergono *ex nihilo*: l'opacità delle trame causali, il c.d. *many hands problem* e la produzione di rischi non riconoscibili *ex ante* sono manifestazioni già note nell'esperienza del diritto penale post-moderno. Cfr., tra i diversi studi in merito, quelli di PALIERO (1994), pp. 1240 ss.; BERNARDI (2003), pp. 22 ss.; PIERGALLINI (2004), pp. 24 ss., 140 ss., 305 ss.; CASTRONUOVO (2005), pp. 320 ss. L'intelligenza artificiale ne esaspera la portata: introduce nella catena causale un agente di fattura umana, ma dotato di ampi margini di autonomia operativa, capace di cagionare l'offesa e, al contempo, privo di soggettività giuridica. In questo senso, in uno scritto che indaga specificamente sulla responsabilità del produttore di sistemi di intelligenza artificiale, cfr. FRAGASSO (2023), pp. 10 ss.

⁵⁵ Cfr., tra i diversi, BORSARI (2019), p. 265; PIERGALLINI (2020), pp. 1762 ss.; MAGRO (2020), p. 6; PIVA (2022), p. 687. Sul carattere inizialmente marginale del dibattito, cfr. PICOTTI (2024a), p. 83, il quale qualifica il rapporto tra diritto penale e IA come un «tema di nicchia».

⁵⁶ Ci si riferisce alle ipotesi di utilizzo dell'IA come strumento del reato. Cfr., per tutti, BORSARI (2019), p. 265; BASILE (2019), pp. 25 ss.

⁵⁷ Nell'economia della presente indagine non risulta opportuno ripercorrere tali profili, già ampiamente tematizzati nel dibattito. In merito, oltre alle ricerche appena richiamate, si ricordano, senza pretesa di esaustività: PAGALLO (2010),

(con gradi diversi) e tempistiche decisionali che sfuggono ai ritmi fisiologici del discernimento umano⁵⁸. Ne risulta l'incapacità dell'apparato umano di esercitare un pieno governo sulle dinamiche operative di tali sistemi, i cui esiti, pertanto, possono sfuggire all'imputazione⁵⁹. Un versante problematico che attiene, dunque, all'"afferrabilità" dell'IA nelle *forme* del diritto penale, ossia delle norme sostanziali – di rango ordinario e costituzionale – che ne definiscono le condizioni di possibilità.

La seconda ragione del confinamento noumenico dell'IA può essere individuata nella difficoltà, per l'interprete privo di una competenza informatico-computazionale, di accedere ai presupposti tecnici di progettazione e di impiego dei sistemi di *machine learning* e, conseguentemente, di valutarne criticamente le implicazioni in tema di nesso causale e di controllo umano. Sembrerebbe, *prima facie*, una questione di natura quasi gnoseologica⁶⁰.

Chi pratica il diritto penale conosce da tempo la necessità di dialogare con altri saperi specialistici: almeno da quando, in Italia, si è affermata l'esigenza di un confronto con il sapere scientifico nella ricostruzione del decorso causale reale⁶¹. La peculiarità che qui rileva, tuttavia, non risiede nella mera necessità di ricorrere a saperi tecnici esterni – comune ad altri settori (quali la responsabilità medica, la sicurezza sul lavoro o l'ambiente) – bensì nella circostanza che, nel caso dell'intelligenza artificiale, tale mediazione è resa particolarmente gravosa anche dalla rapida evoluzione tecnologica e funzionale dei sistemi di IA e dalla necessità di confrontarsi, caso per caso, con architetture tecniche e applicative profondamente differenziate⁶². La difficoltà in esame,

pp. 595 ss.; MAGRO (2014), pp. 513 ss.; UBERTIS (2020), pp. 78-79; PANATTONI (2021a), pp. 317 ss.; SALVADORI (2021), pp. 86 ss.; PREZIOSI (2022), pp. 713-722; CONSULICH (2022), pp. 1027 ss.; FRAGASSO (2023), pp. 26 ss.; CAPPELLINI (2023), pp. 23 ss. Un inquadramento sistematico delle problematiche in discorso è offerto da alcune recenti monografie: AMORE e ROSSERO (2023); LANZI (2023a); TERRIZZI (2023); MANGIONE (2024); COMPOSTELLA (2024); D'AMICO (2025); FRAGASSO (2025b). Nella letteratura straniera, cfr. almeno: HALLEVY (2013), *passim*; BECK (2016), pp. 138 ss.; MITTELSTADT *et al.* (2016); SURDEN e WILLIAMS (2016), p. 157; BATHAEE (2018), pp. 907-908; ELLYSON (2018), pp. 887 ss.; COECKELBERGH (2020), p. 2052.

⁵⁸ La rapidità dell'agire dell'IA è caratteristica che risulta incidere sulla stessa *possibilità di osservare* il «dovere di riconoscimento della situazione di rischio» (c.d. *Erkennensollen*). Su quest'ultima nozione, in lingua italiana, per tutti, cfr. FORNASARI (1993), pp. 220 ss.; sul concetto di "dovere di riconoscere", v. *infra*, § 2.1 e § 2.1.1.

⁵⁹ Non sorprende dunque che, nell'affrontare tali questioni, parte della dottrina sia giunta a interrogarsi sulla possibile configurabilità di una responsabilità penale diretta in capo all'IA. Cfr., per tutti, CAPPELLINI (2018-2019), pp. 499 ss. In ambito internazionale, un riferimento imprescindibile per la tesi favorevole alla configurabilità di una responsabilità penale della macchina è HALLEVY (2010). Di segno opposto si è espressa l'Associazione Internazionale dei Professori di Diritto Penale nella risoluzione approvata a Siracusa nel 2022: cfr. PICOTTI (2024b), pp. 16 e 17. Per un approfondimento monografico sul tema del *machina delinquere potest*, v. GIANNINI (2023a).

⁶⁰ Ci si riferisce ai limiti di conoscibilità dell'oggetto. Tuttavia, se si accoglie la prospettiva di cui *infra*, § 2.2, il problema concerne i *metodi* idonei a rendere più conoscibili, per il giurista, condizioni operative, effetti e rischi concretamente prodotti dall'intelligenza artificiale; in questo senso, la questione pare assumere natura più propriamente epistemologica.

⁶¹ Il riferimento è all'elaborazione giurisprudenziale sviluppatasi almeno a partire dagli anni Ottanta, in particolare con la nota sentenza sul disastro di Stava (Cass. Pen., Sez. IV, 6 dicembre 1990, n. 4793, *Bonetti*), e successivamente sistematizzata dalle Sezioni Unite del 2002 (Cass. Pen., Sez. Un., 10 luglio 2002, n. 30328, *Franzese*).

⁶² Può accadere che, anche in ragione dell'obiettivo complessità delle questioni da affrontare, il dibattito si orienti su ipotesi elaborate a partire da singole applicazioni, non necessariamente idonee a restituire la pluralità delle soluzioni tecniche concretamente impiegate nei singoli domini. Un rischio che invita a una riflessione scientifica maggiormente calata sulle specificità delle singole applicazioni. Con ciò non si intende negare l'indubbio pregio scientifico dei contributi di taglio trasversale sul rapporto problematico tra le categorie penalistiche e l'IA. La riflessione qui proposta si colloca, piuttosto, su un piano prospettico, che pone tali acquisizioni a fondamento di ulteriori sviluppi. Sul punto si torna *infra*, § 2.2.

pertanto, non sembra esaurirsi, al pari della prima, in un problema di intervento del diritto penale, ma si colloca a un livello ulteriore, investendo i paradigmi conoscitivi propri della scienza penalistica⁶³.

2.1. L'intelligenza artificiale: da noumeno a fenomeno per il diritto penale. Le questioni della riconoscibilità ed evitabilità.

Il primo dei due profili enucleati nel paragrafo precedente – quello relativo alle difficoltà di inquadrare l'agire dei sistemi di *machine learning* nelle categorie penalistiche – può oggi configurarsi, per chi scrive, in modo parzialmente diverso rispetto alle prime fasi evolutive di questa «forma di agire»⁶⁴.

L'intelligenza artificiale, infatti, sembra aver superato la propria fase “adolescenziale”: da realtà segnata dall'indeterminatezza, essa va convertendosi in esperienza misurabile non soltanto nei suoi *effetti* lesivi, ma anche nelle *dinamiche* che ne costituiscono la premessa e spiegano la ricorrenza di quegli esiti⁶⁵.

Tale passaggio, tuttavia, non va predicato rispetto all'intero spettro delle applicazioni dell'intelligenza artificiale. In diversi ambiti, pur essendo riconoscibile la possibilità che l'impiego di questa tecnica dia luogo a esiti pregiudizievoli, permane l'opacità sul *come* tali offese si producano. Ciò accade, ad esempio, nei sistemi di guida autonoma, specie di livello 4 e 5⁶⁶, nei quali l'evento lesivo è prevedibile in astratto, ma

⁶³ In termini più generali, MANGIONE (2024), p. 105, rileva come la difficoltà di confrontarsi con la dimensione algoritmica non attenga soltanto alla complessità tecnico-funzionale dei sistemi, ma investa la relazione tra linguaggio giuridico e linguaggio tecno-scientifico. In particolare, l'Autore osserva: «Adesso il tentativo di raffigurare una realtà su cui intervenire ordinandola si scontra già con un *deficit espressivo* poiché non possediamo una semantica adeguata a cogliere ed esprimere il significato e dunque la comprensione delle inter-relazioni che intercorrono con l'uomo». Ne discende una conseguenza non meramente teorica: «l'ordinamento giuridico si scopre oggi deprivato alla base di quel collante scientifico e filosofico che decretava l'unicità del linguaggio: ma nella misura in cui stenta ad adempiere al ruolo ordinante della realtà – regole che non normano – si rivela incapace di cogliere la verità delle cose» (*ivi*, pp. 106-107).

⁶⁴ L'espressione è ripresa da FLORIDI (2022), p. 26, il quale, riferendosi all'IA afferma: «la mia tesi è che si tratti di una nuova forma di agire, e non di intelligenza» (*ivi*, pp. 32 ss.).

⁶⁵ Questo passaggio, come si approfondirà *infra* al § 2.1.3, comporta un potenziale slittamento dal terreno della precauzione (dove il rischio è avvertito, ma non ancora compreso nei suoi meccanismi etiologici), a quello della prevenzione (che presuppone la conoscibilità delle dinamiche causali che conducono all'evento offensivo, dalla quale discende la possibilità di individuare *ex ante* le condotte rilevanti in tal senso). Sono note, infatti, le difficoltà di compatibilità con i principi costituzionali di fattispecie costruite sul piano legislativo, o interpretate su quello applicativo, secondo logiche propriamente precauzionali.

⁶⁶ Il riferimento è alla classificazione dei sistemi di automazione della guida elaborata dalla Society of Automotive Engineers nel [documento tecnico J3016](#) («SAE Levels of Driving Automation»), che distingue sei livelli di automazione: da 0 (nessuna automazione) a 5 (completa automazione). Secondo tale classificazione, un veicolo di livello 5 è in grado di eseguire tutte le mansioni di guida in qualsiasi contesto e condizione, senza alcun coinvolgimento umano: in altri termini, l'automazione sostituisce integralmente il conducente nella *Dynamic Driving Task*. Allo stato, tale livello è un obiettivo tecnologico ancora non realizzato su larga scala, in quanto nessun sistema di livello 5 risulta attualmente disponibile sul mercato o pienamente operativo in ogni condizione stradale. Per contro, il livello 4 prevede che il sistema automatico possa gestire tutte le funzioni di guida senza intervento umano, ma entro un dominio operativo definito (c.d. *Operational Design Domain*). Fuori dal dominio operativo, il sistema deve essere in grado di arrestarsi in sicurezza (c.d. *fallback*). Dal punto di vista regolatorio, la maggior parte degli ordinamenti, incluso quello dell'Unione europea, non ammette ancora la rinuncia al controllo umano in ogni contesto di circolazione, e la disciplina attuale – pur aprendo all'omologazione e alla sperimentazione di veicoli di livello 3 e, in parte, di livello 4 – mantiene requisiti di supervisione per contenere i rischi e le incertezze legate alla responsabilità in caso di eventi dannosi. Su

la ricostruzione del contributo umano causale all'operatività del sistema – risultante dall'interazione di molteplici moduli – non appare, allo stato, esperibile secondo i canoni penalistici⁶⁷.

Diversamente, nel campo dei *chatbots* terapeutici, le comunicazioni delle associazioni professionali dei sanitari⁶⁸, la letteratura scientifica di riferimento⁶⁹, nonché le evidenze critiche emerse nel contesto applicativo, hanno già concorso a delineare un quadro conoscitivo dei *patterns* pericolosi legati all'uso di tali sistemi. Risulterebbe già acquisito che, specie nei soggetti con disturbi mentali, alcune tipologie di interazione siano idonee a evolvere in schemi di progressiva dipendenza emotiva e di ritiro dal contesto sociale, da considerarsi fattori rilevanti rispetto al peggioramento della psicopatologia e, nei casi più gravi, all'atto suicidario (v. *supra*, § 1).

In questa prospettiva, ciò che in un primo momento poteva considerarsi una conseguenza imprevedibile o del tutto eccezionale, può dirsi oggi un seguito *prevedibile* di specifiche scelte di *configurazione* del servizio⁷⁰. In particolare, in questo tipo di eventi

questi aspetti, nella dottrina penalistica italiana, cfr. i lavori monografici di LANZI (2023a) e COMPOSTELLA (2024), nonché, senza pretesa di esaustività, i contributi di: CAPPELLINI (2019), pp. 325 ss.; PICOTTI (2021), pp. 813 ss.; VADALÀ (2023). V. anche D'AMICO (2025) e FRAGASSO (2025b). A fronte di una letteratura straniera ormai sterminata, anche per i riferimenti ci si limita qui a richiamare: nella dottrina francese, BÉNÉJAT (2019), pp. 428 ss.; NOGUÉRO e VINGIANO-VIRICEL (2026), pp. 101 ss.; in quella spagnola, BLANCO CORDERO (2025); in quella di *common law*, DOUMA e PALODICHUK (2012), pp. 1157 ss.

⁶⁷ In tal senso, v. COMPOSTELLA (2024), spec. pp. 162-166, il quale, riprendendo e sviluppando le riflessioni di PIERGALLINI (2020) in tema di accertamento causale e imputazione penale nell'ambito delle offese AI-correlate, mette in luce come l'opacità decisionale dei sistemi di guida altamente automatizzati ostacoli la riconduzione dell'evento a un contributo causalmente qualificabile secondo i tradizionali criteri penalistici.

⁶⁸ V. *supra*, § 1, nota n. 47. Nondimeno, sotto il profilo epistemologico, le comunicazioni provenienti dalle associazioni professionali di psicologi e psichiatri costituiscono un riferimento autorevole sulla conoscibilità dei rischi, ma non per questo immuni da possibili limitazioni prospettiche. Infatti, resta ipotizzabile che, nel confrontarsi con il fenomeno emergente dei *chatbots* terapeutici, tali organismi occupino una posizione che può dar luogo a tensioni o a parziali sovrapposizioni di interessi. Del resto, la letteratura ha da tempo segnalato come, nelle società post-industriali, l'intensificazione della produzione del sapere favorisca processi di "convenzionalizzazione" della scienza, con il rischio che alcuni gruppi disciplinari sviluppino orientamenti selettivi in relazione alle dinamiche del conflitto sociale – cfr. PIERGALLINI (2004), p. 17, spec. nota n. 57. Nel caso in esame, tuttavia, alle comunicazioni delle associazioni professionali si affianca una pluralità ormai significativa di contributi scientifici provenienti da ambiti disciplinari differenti (come psichiatria, psicologia e scienze della comunicazione), nonché un insieme coerente di riscontri empirici. In tale prospettiva, anche l'obiezione – in astratto condivisibile – relativa alla fallibilità del sapere scientifico non sembra assumere un rilievo tale da compromettere l'affidabilità complessiva del quadro conoscitivo disponibile. Sull'osmosi tra scienza, verità e diritto, v., per tutti, TALLACCHINI (2006), pp. 125-140.

⁶⁹ V. *supra*, § 1, spec. note nn. 33-43.

⁷⁰ Beninteso, in questa sede il riferimento alla prevedibilità non assolve a una funzione meramente descrittiva, ma rinvia a uno dei criteri di imputazione più consolidati nella teoria del reato colposo. La sua centralità è ampiamente riconosciuta non solo nella dottrina italiana e tedesca, ma anche negli ordinamenti di *common law*. Cfr. FORTI (1990), p. 13, che riporta i riferimenti a: SCHROEDER (1989), p. 776; TRIFFTERER (1979), pp. 201 ss.; MAIWALD, p. 444 (per la dottrina di area germanica); WILLIAMS (1983), pp. 96 ss. e pp. 388 ss.; PERKINS e BOYCE (1982), pp. 812 ss., 840 ss.; LAFAVE e SCOTT (1986), pp. 339 ss., pp. 411 ss. (per la dottrina anglo-americana). Un riconoscimento analogo del criterio emerge anche nella tradizione francese: cfr. GARRAUD (1924), p. 405, che scrive «la faute consiste, à n'avoir pas prévu, alors qu'on pouvait les prévoir, les conséquences dommageables d'une action ou d'une omission»; VIDAL (1949), pp. 193-194, secondo il quale «Il n'y a donc pas faute si l'auteur du fait a agi avec la prudence et l'attention qu'aurait apporté tout homme normal ou encore s'il n'a pu prévoir ni empêcher le fait» (corsivo di chi scrive); MERLE e VITU (1997), p. 757, che affermano «la faute pénale consiste donc soit à ne pas prévoir les conséquences dommageables de l'acte que l'on accomplit [...] soit à ne prendre les précautions nécessaires pour les empêcher de survenir». Tale convergenza dottrinale è sintomatica della razionalità raggiunta dal diritto penale post-illuminista, nella misura in cui il ricorso alla prevedibilità consente di emancipare la colpa da una logica di imputazione *ex post* dell'evento. Nondimeno, la centralità del criterio non dissolve, ma anzi accentua, le persistenti incertezze teoriche circa la nozione di "evento prevedibile", di cui si darà conto, pur nei limiti funzionali al presente lavoro, nel prossimo

AI-correlati, la causa primaria non pare risiedere nell'eventuale e sporadico *output* istigatorio del sistema⁷¹: essa risulta riconducibile, più a monte, al rapporto che si instaura tra l'utente e i *chatbots* in discorso, quale effetto di specifiche architetture di interazione⁷².

Del resto, la letteratura psichiatrica non consente di inquadrare, in generale, il rischio suicidario come derivazione lineare di un evento isolato⁷³, bensì come esito di un processo complesso e temporalmente disteso, nel quale fattori predisponenti, dinamiche relazionali e condizioni di vulnerabilità concorrono progressivamente alla crisi suicidaria⁷⁴. In un simile contesto, ad assumere rilievo centrale non sono i singoli stimoli esterni, quanto il monitoraggio del rischio e l'insorgenza di segnali di allarme, tra cui rientrano la riduzione dei contatti sociali (che costituiscono uno dei fattori protettivi)⁷⁵ e la chiusura progressiva rispetto a figure di supporto umano⁷⁶.

Entro queste coordinate, ne deriva che il pericolo per la vita e per l'incolumità individuale connesso all'impiego dei *chatbots* per il supporto psicologico si verifica laddove specifiche scelte di configurazione del servizio inducano l'utente con disturbi mentali a percepire l'agente artificiale come interlocutore da privilegiare, con un duplice effetto critico: da un lato, l'accentuazione di dinamiche di isolamento sociale e, nei soggetti maggiormente vulnerabili, l'instaurarsi di una dipendenza emotiva nei confronti del *chatbot*; dall'altro, l'ostacolo all'emersione e alla gestione tempestiva dei segnali di allarme nell'ambito di una presa in carico clinica, in particolare nei casi di disincentivo verso un percorso di cura con un sanitario⁷⁷.

paragrafo.

⁷¹ L'espressione «*output* istigatorio» viene impiegata per indicare i contenuti generati dall'IA che, nel contesto dell'interazione *chatbot*-utente, possono risultare idonei a sollecitare o rafforzare propositi autolesivi (per un esempio, v. *supra*, § 1, nota n. 32). La formula è utilizzata in senso descrittivo e non implica, dunque, una qualificazione giuridico-penale dell'agire del sistema.

⁷² Cfr. § 1. Ciò, è bene precisare, non esclude – almeno in una prospettiva di prima ricognizione come quella del presente studio – che l'assenza di presidi tecnici volti a intercettare eventuali *outputs* istigativi del suicidio possa, a sua volta, costituire una causa, pur secondaria. Allo stato, tuttavia, la loro effettiva praticabilità tecnica risulta controversa, specie alla luce dei limiti emersi nel recente studio di [PICHOWICZ et al.](#) (2025). Un utile termine di confronto può tuttavia rinvenirsi nei meccanismi di moderazione e controllo adottati dalla società “Luka Inc.” per il *chatbot* “Replika”, resi noti nell'ambito di un procedimento dinanzi al Garante per la protezione dei dati personali, sui quali si torna *infra*, § 4.

⁷³ Il suicidio costituisce oggetto di indagine da parte di una pluralità di prospettive disciplinari (tra cui filosofiche, letterarie, sociologiche e mediche), ciascuna delle quali illumina profili differenti di un atto non riducibile a spiegazioni univoche. In questa sede si è tuttavia ritenuto opportuno ancorare le riflessioni giuridiche agli apporti della letteratura più accreditata in ambito psichiatrico, in quanto idonea a fornire una base concettuale comune per l'analisi del pericolo tanto rispetto all'incolumità individuale, quanto rispetto all'incolumità pubblica. Con riguardo a quest'ultima, la questione se e in quale misura la messa in commercio di *chatbot therapists* possa generare un pericolo comune – ossia un fatto dotato «di una potenzialità lesiva generalizzata e indomabile, perché diretta in maniera diffusiva verso una collettività di persone indeterminate» (CASTRONUOVO, 2024, p. 3) – potrà misurarsi nel tratto conclusivo del presente lavoro, § 5 e § 6, spec. note nn. 243 e 249, senza essere oggetto di specifico approfondimento.

⁷⁴ POMPILI (2022), *passim*. Rispetto al verificarsi della crisi suicidaria, l'Autore evidenzia che in tale memento «le capacità cognitive si riducono, ossia il soggetto non riesce a vedere le fonti di sostegno disponibili». Una condizione che può risultare accentuata da situazioni di isolamento sociale.

⁷⁵ *Ivi*, p. 71.

⁷⁶ L'impostazione qui seguita si pone in linea con la letteratura scientifica internazionale, che attribuisce rilievo centrale al monitoraggio precoce dei segnali di allarme del rischio suicidario, frequentemente ricondotti al modello “*IS PATH WARM*” (*Ideation, Substance abuse, Purposelessness, Anxiety, Trapped, Hopelessness, Withdrawal, Anger, Recklessness, Mood changes*) inizialmente elaborato dall'American Association of Suicidology e successivamente ampliato.

⁷⁷ Il quadro delineato impone ora di volgere l'attenzione a un profilo talvolta trascurato nelle riflessioni giuridiche, ma

Tale ricostruzione assume rilievo diretto non solo sul piano della riconoscibilità del rischio, ma anche ai fini dell'individuazione del disvalore della condotta del fornitore e, nella prospettiva qui proposta, della delimitazione del nesso di causalità penalmente rilevante.

Come afferma Donini, «la stessa causalità, dovendosi collegare ad una condotta che la precede, se la si accerta all'interno di una contestazione di dolo o di colpa, presuppone un comportamento che, in ipotesi, è doloso o colposo. Da ciò, tuttavia, l'accertamento rimane condizionato. Nel reato colposo, per es., si dovrà pur sempre verificare che l'evento è conseguenza della colpa, e non solo *della condotta*: e quindi si porranno specifiche verifiche riguardanti il 'nesso' di anti giuridicità, il materializzarsi nell'evento dello specifico addebito sulla diligenza o la prudenza "omessa", e quindi si ricostruirà la c.s.q.n. in rapporto a una condotta *ipotetica* [...] modellata sul c.d. comportamento alternativo lecito»⁷⁸. In questa prospettiva, l'indagine causale sull'evento dovrebbe presupporre un previo giudizio di imputazione fondato sull'inosservanza di una regola di diligenza. Solo dopo avere individuato lo specifico addebito colposo – e il correlato comportamento alternativo lecito esigibile – diviene possibile verificare se l'evento rappresenti concretizzazione di quella carenza di diligenza. In altri termini, ad avviso di chi scrive, la causalità dell'evento penalmente rilevante dovrebbe essere delimitata *ab origine* dall'individuazione delle condotte che già denotano un deficit di diligenza (in questo senso, un rischio illecito). Traslato sul terreno dei *chatbot therapists*, l'insegnamento dottrinale richiamato suggerisce che l'indagine sul nesso causale non dovrebbe guardare agli atti di programmazione da cui potrebbe essere dipeso l'*output* istigatorio, bensì alle scelte di configurazione complessiva del servizio che modellano il rapporto utente-*chatbot* (accesso continuativo, elevata personalizzazione dell'interlocutore artificiale, possibilità di attribuirgli tratti di rilievo affettivo, dinamiche di *homophily*, presentazione come strumento di cura, etc.). L'*output* specifico del *chatbot* (vale a dire il singolo contenuto generato nel corso dell'interazione), infatti, difficilmente risulterebbe riconducibile a una specifica condotta *colposa* dell'uomo, in ragione dell'opacità dei processi interni di tali sistemi di *machine learning* e della loro capacità di apprendere dall'esperienza. Nel contesto attuale, l'individuazione di un comportamento alternativo lecito rispetto a quello tenuto da un programmatore del *chatbot* risulterebbe, in larga misura, artificiosa. Di riflesso, non consentirebbe di rintracciare in quella condotta un'imprudenza o un'imperizia: sicché l'interrogativo sull'efficacia causale di quel fatto finirebbe per riguardare un comportamento che, allo stato, non può dirsi colpevole.

centrale nella letteratura psichiatrica: i conflitti potenzialmente emergenti tra il soggetto a rischio suicidario e il sanitario. Le analisi di settore evidenziano come l'intervento farmacologico o psicoterapeutico, pur clinicamente indicato, venga talora percepito dal paziente come espressione di una razionalità istituzionale non coincidente con le proprie esigenze soggettive e immediate. In tale prospettiva, assumono rilievo fattori quali il timore di essere classificato come "malato di mente", il timore di un ricovero psichiatrico, nonché «la paura di essere messo nella condizione di doversi confrontare con quella che appare una scelta inevitabile da parte del terapeuta ("Dovrò informare la tua famiglia")» – POMPILI (2022), p. 65. Tali profili di frizione, tipici della relazione terapeutica tradizionale, non connotano l'interazione con i *chatbot therapists*, la quale, specie se accessibile in modo continuativo, contribuisce a rafforzare dinamiche di elusione del confronto con professionisti umani.

⁷⁸ DONINI (2006), pp. 15-16.

Diverso appare, invece, il profilo delle scelte strutturali del servizio, definite dal fornitore a livello di *governance*: si tratta di *opzioni progettuali sulle quali l'autonomia del chatbot non incide* – ma che definiscono il quadro funzionale entro cui tale autonomia si dispiega – e la cui attitudine a generare affidamento esclusivo e coinvolgimento emotivo in soggetti con vulnerabilità psichiche è attestata in letteratura (v. *supra*, § 1 e § 2.1), con conseguenze rilevanti anche sul giudizio di evitabilità (v. *infra*, § 5). Verosimilmente, è solo rispetto a tali scelte di conformazione del servizio, pienamente imputabili all'agente umano, che può rinvenirsi un effettivo contenuto di disvalore; ed è, dunque, con riguardo a esse che, in una prospettiva penalistica, sembra ragionevole avviare l'indagine causale.

Beninteso, con tale ricostruzione non si intende aderire alle concezioni del c.d. “finalismo estremo”, che fondando la responsabilità penale solo sulla condotta inosservante della diligenza richiesta, propendono per una totale estromissione dell'evento tra gli elementi costitutivi dell'illecito⁷⁹. La ricostruzione proposta mira, piuttosto, a suggerire un criterio di priorità logica coerente con la struttura dell'illecito colposo.

Breve: su un piano sistematico, la marginalizzazione degli “*outputs* istigatori” quale baricentro dell'indagine causale non si presenta come *escamotage* metodologico volto a superare le problematiche scaturenti dall'autonomia dell'IA, bensì – almeno nell'attuale quadro conoscitivo – come esito imposto sia dalla struttura dell'illecito colposo, sia dalle acquisizioni scientifiche richiamate (le quali portano a escludere una lettura del suicidio quale effetto diretto di contenuti validanti l'orientamento autolesivo).

Prima di proseguire, anche e proprio in una prospettiva di dialogo interdisciplinare, si ritiene necessario un chiarimento terminologico. La nozione di “prevedibilità” talvolta impiegata in sede clinica non coincide con quella rilevante sul piano penalistico. In ambito psichiatrico, essa attiene alla possibilità (esclusa in letteratura) di predire il gesto suicidario di un individuo determinato; nel diritto penale, invece, è connessa alla riconoscibilità *ex ante* di una situazione rischiosa, in quanto idonea a generare un evento lesivo⁸⁰.

Applicata al contesto dei *chatbot therapists*, tale distinzione fa discendere che la responsabilità colposa del fornitore non postuli la prevedibilità del suicidio *ut sic*, ma la riconoscibilità dell'attitudine di determinate scelte di configurazione del servizio a

⁷⁹ Per una rassegna di tali teorie, contrarie al dato positivo del codice penale italiano, v. FORTI (1990), pp. 330 ss.

⁸⁰ Sia permesso un chiarimento sulla scelta del lessico utilizzato nel presente scritto. Sul piano penalistico, i concetti di *riconoscibilità del rischio* e di *prevedibilità dell'evento* non risultano sovrapponibili. Parte della dottrina attribuisce rilievo centrale al parametro della riconoscibilità, in quanto dotato, rispetto alla prevedibilità, «dell'attitudine di cogliere la situazione in cui la colpa dell'agente consiste nel non conoscere, al momento in cui agisce, circostanze di fatto che rendono il suo comportamento conforme a una fattispecie di reato» – FORTI (1990), p. 185, nota n. 88, che riprende SCHROEDER. In questa prospettiva, la prevedibilità dell'evento presuppone logicamente la previa riconoscibilità delle situazioni di rischio idonee a condurre alla sua verifica, giacché solo ciò che è riconoscibile come fonte di pericolo può essere oggetto di un *dovere di evitare* – sul punto, v. *infra*; per il concetto di *dovere di evitare*, salvi gli approfondimenti di cui *infra*, § 2.2.2, v. CIVELLO (2021), pp. 1004 ss. Pertanto, nel presente scritto, il concetto di *riconoscibilità* viene riferito alle situazioni di rischio, mentre quello di *prevedibilità* all'evento; l'eventuale uso, per ragioni di sintesi espositiva, del solo termine “prevedibilità” non deve essere inteso come sufficiente in assenza della riconoscibilità delle condizioni di rischio idonee a rendere possibile, o anche solo probabile, la verifica dell'evento. In tali termini, la condotta vietata deve essere individuata in quella che cagiona l'evento mediante la realizzazione di un rischio del tipo di quelli assunti dalla regola cautelare violata.

esporre una particolare categoria di utenti a un incremento del rischio di lesioni, nonché la prevenibilità di tale esito mediante soluzioni progettuali alternative⁸¹.

Sia consentita un'ultima telegrafica precisazione concettuale, benché possa trattarsi di un rilievo in larga misura acquisito in dottrina. Nel diritto penale, il riferimento alla riconoscibilità non può essere degradato a mera possibilità fattuale – contingente e occasionale – di rappresentazione del rischio lesivo. Una simile lettura finirebbe per svuotarne la funzione di orientamento delle condotte dei consociati, dissolvendo la colpa in un paradigma incapace di offrire *ex ante* un criterio standardizzabile di distinzione tra lecito e illecito. Gli studiosi hanno reso ben noto, del resto, come le fattispecie colpose, a differenza di quelle dolose, presentino una struttura ontologicamente aperta, giacché l'inesauribile varietà delle condotte negligenti si sottrae a qualsiasi pretesa di tipizzazione esaustiva da parte del legislatore⁸². Proprio per questo, al fine di recuperare un adeguato coefficiente di determinatezza – tale da rendere predeterminabile l'area dell'illecito colposo – la riconoscibilità del rischio non può essere parametrata all'episodica possibilità di rappresentarselo, ma va riferita al delinearsi delle condizioni conoscitive che sono necessarie al fine di prevenire l'evento e, dunque, *doverose* da acquisire.

In questi termini, se ne ricava ulteriormente che la riconoscibilità – che dunque rileva sul piano del *fatto tipico*, quale sede deputata alla funzione garantistica di predeterminazione dell'illecito – *non* si configura come *possibilità* obiettiva, bensì come *dovere* di riconoscere, che l'agente assolve adeguandosi «a una aspettativa “di previsione” da parte dell'ordinamento»⁸³.

Ecco allora che per definire il contenuto del *dovere di riconoscere* occorre individuare le conoscenze la cui acquisizione è necessaria (*rectius* *doverosa*) ai fini di evitare la *classe di eventi* presa in considerazione da una regola cautelare⁸⁴.

⁸¹ Ciò non deve far ritenere che la nozione penalistica di prevedibilità sia più ampia di quella clinica. La differenza è di significato e funzione. Anche in ambito psichiatrico, infatti, si riconosce la possibilità di individuare condizioni e fattori suscettibili di aumentare il rischio di suicidio: «Predire il suicidio è alquanto difficile, semmai è anche impossibile. Possiamo fare un'anticipazione ragionevole del futuro raccogliendo dei dati, introducendoci anche nella sofferenza del soggetto stesso, cercando di modificare quelle componenti che è nelle nostre potenzialità poter in qualche modo cambiare» – [POMPILI](#) (2024), min. 1:10.

⁸² Il carattere strutturalmente aperto delle fattispecie colpose costituisce un'acquisizione consolidata nella dottrina penalistica internazionale, la cui teorizzazione è ricondotta a WELZEL (1969), p. 2. In Italia, per tutti, cfr. FORTI (1990), pp. 136-138; CASTRONUOVO (2009), spec. pp. 166 ss.; *Id.* (2021), pp. 200 ss.; GIUNTA (2009), pp. 150 ss.

⁸³ FORTI (1990), p. 221. Sul punto, cfr. altresì PULITANO (2006), p. 779, il quale afferma «Per poter agire correttamente, occorre conoscere quanto necessario per valutare le prevedibili conseguenze della propria condotta, in modo da poter evitare eventuali conseguenze vietate». Nello stesso senso, v. GROTTA (2012), p. 131-132, il quale ben evidenzia la successione logica tra il giudizio di *probabilità obiettiva* e quello di *riconoscibilità obiettiva*: in primo luogo occorre stabilire, sulla base delle conoscenze scientifiche disponibili, se un determinato evento fosse oggettivamente prevedibile e dunque evitabile mediante l'adozione di una regola cautelare; solo successivamente si può verificare se tale regola fosse riconoscibile, vale a dire se l'agente avesse il dovere di coglierne la rilevanza preventiva e di conformare ad essa la propria condotta. Il giudizio di riconoscibilità presuppone, quindi, quello di probabilità, poiché soltanto dopo aver individuato le cautele necessarie ad evitare l'evento è possibile stabilire se fosse doveroso riconoscerle e applicarle.

⁸⁴ Pertanto, è bene chiarire, nel fatto colposo la riconoscibilità non attiene al divieto penale (e dunque «alla percepibilità del fatto come penalmente illecito»), bensì al «fatto lesivo del bene giuridico tutelato dalla norma, oggetto di una pretesa dell'ordinamento avente come presupposto un dato normativamente rilevante in rapporto a tale riconoscibilità» – FORTI (1990), p. 226.

2.1.1. Il criterio del rischio tipico nel giudizio di riconoscibilità.

Nell'ambito dei *chatbot therapists*, tra i fattori che consentono di affermare la riconoscibilità dei rischi e la prevedibilità delle offese alla vita e all'incolumità degli utenti non vi sono soltanto l'esperienza maturata e gli studi scientifici sul tema, ma anche il delinearsi di un regime normativo che, rispetto alla fase iniziale dell'impiego massivo dell'IA⁸⁵, precisa alcuni standard cautelari in materia.

L'esistenza di regole cautelari positive, infatti, può assumere particolare rilievo nella definizione del *dovere di riconoscere*, in considerazione delle aporie che segnano il concetto di "prevedibile"⁸⁶. A tal proposito, non certamente per fini didascalici, ma per la loro valenza strutturale rispetto al fenomeno in esame, sia permesso richiamare alcune coordinate teoriche.

Quando la riconoscibilità del rischio può basarsi esclusivamente sull'esperienza acquisita, e non su regole positivizzate, il contenuto del *dovere di riconoscere* potrebbe risultare controverso, mancando un parametro preciso cui ancorare il giudizio⁸⁷.

⁸⁵ La prima stagione della regolazione dell'intelligenza artificiale si è vista contrassegnata dal ricorso a strumenti di *soft law*, di natura prevalentemente programmatica e non vincolante. In tale contesto si ricordano: sul piano sovranazionale, le iniziative promosse dalle istituzioni dell'Unione europea e dall'OCSE, nonché gli orientamenti elaborati dall'*European Data Protection Board* in materia di protezione dei dati personali; sul piano interno, il *Libro Bianco AgID* (2018) e la *Strategia italiana per l'intelligenza artificiale* elaborata dal MISE nel 2019 – cfr. MOBILIO (2020), pp. 401 ss. A questo quadro si è affiancata l'elaborazione, da parte di gruppi di esperti e soprattutto di organizzazioni private, di un numero soverchiante di principi e codici etici – nel 2019, arrivati a oltre 160 secondo l'*AI Ethics Guidelines Global Inventory* (2019). Tale proliferazione, pur indice di una diffusa sensibilità, si traduceva in frammentazione e incertezza, alimentando fenomeni di *ethics shopping*. Su questo aspetto FLORIDI (2022), pp. 92-93.

⁸⁶ In proposito, CASTRONUOVO (2009), pp. 194-195, con riferimento alla prima delle tre questioni che individua in tema di prevedibilità, mette in luce il problema attinente «al momento a partire dal quale possa predicarsi la riconoscibilità di un rischio e la prevedibilità di effetti indesiderati ad esso ricollegabili» (p. 195).

⁸⁷ Benché si tratti di un dato acquisito nel dibattito penalistico italiano, sempre in ottica di apertura interdisciplinare giova sottolineare che in assenza di regole di cautela precisate in testi normativi, la determinazione di ciò che è riconoscibile *ex ante* risulta esposta al rischio di valutazioni soggettive e retrospettive. Per evitare tale esito e garantire la predeterminabilità dell'illecito colposo, la dottrina penalistica ha elaborato una figura di riferimento impersonale – l'*agente modello* – attraverso la quale la riconoscibilità dei rischi *in rerum natura* viene rivestita «di una fisionomia normativa» (FORTI, 1990, p. 234), in quanto l'agente modello (o, nel caso di attività produttive, il *modello di attività lecita*) «incarna il punto di vista del diritto» ed «è la personificazione dell'ordinamento giuridico nella situazione concreta» (KAUFMANN, 1964, p. 51, il cui riferimento è tratto da FORTI, 1990, p. 234). In tale prospettiva, l'agente modello viene identificato nell'uomo «"coscienzioso ed avveduto" del circolo di rapporti cui appartiene l'agente o in generale "più vicino" all'agente» – FORTI (1990), p. 237. La sua assunzione come parametro del giudizio di prevedibilità (ed evitabilità) non esaurisce, tuttavia, il problema relativo alla determinazione del contenuto della diligenza. Sul punto, infatti, sono state prospettate soluzioni differenti. Limitando l'analisi al contesto italiano, secondo un primo orientamento il contenuto del *dovere di riconoscere* deve essere commisurato allo stato più avanzato delle conoscenze scientifiche accessibili nel momento storico della condotta – cfr. MANTOVANI (1988), p. 205; GROTTO (2012), pp. 165 ss. Un secondo indirizzo, più espansivo, postula che la condotta diligente debba tener conto non solo del sapere scientifico comunemente acquisito, ma implichi, per le imprese, il reperimento di «tutte le informazioni possibili», purché accessibili all'agente – cfr. MARINUCCI (2005) p. 51. Un'ulteriore prospettiva ricollega la prevedibilità dell'evento alla sussistenza di «fondati sospetti» circa la pericolosità della condotta, scaturenti dal sopravvenire di circostanze idonee a mettere in discussione la completezza delle leggi scientifiche accolte – cfr. RUGA RIVA (2006), pp. 1762 ss. La pluralità delle impostazioni dottrinali richiamate riflette la difficoltà di determinare, in assenza di una regola di condotta previamente tipizzata, il contenuto del *dovere di riconoscere* i rischi rilevanti. Ciò posto, occorre precisare che nel caso dei *chatbot therapists* la prevedibilità dell'offesa all'incolumità individuale può ritenersi già oggi configurabile, in quanto la letteratura scientifica e la prassi applicativa consentono di individuare con sufficiente determinatezza i rischi connessi al loro uso in contesti caratterizzati dalla vulnerabilità psichica dell'utente (v. *supra*, § 1 e § 2.1).

Diversa può essere la situazione in cui il legislatore introduce regole cautelari orientate al governo di determinati pericoli: in tal caso, ciò che prima doveva essere intuito può risultare, per così dire, maggiormente manifesto⁸⁸. Ne consegue che, salvo il caso di regole cautelari elastiche⁸⁹, il rischio preso in considerazione dalla norma può ritenersi più facilmente *riconoscibile*. A tale conclusione è però necessario sottrarre ogni implicazione estensiva: affermare che l'introduzione di una regola cautelare formalizzata rende solitamente più riconoscibile il rischio che essa è diretta a governare non autorizza a ritenere che qualsiasi rischio idoneo a ledere il bene da essa tutelato divenga, da lì in poi e per ciò solo, ascrivibile a titolo di colpa (specifica). Due le ragioni.

Anzitutto, la dottrina insegna che la presenza di standard cautelari precisati in leggi, regolamenti, ordini o discipline non esonera dal giudizio di prevedibilità *in concreto* dell'evento⁹⁰.

In secondo luogo – ed è il punto che qui maggiormente rileva – l'introduzione di cautele scritte finalizzate a prevenire un evento – meglio, una certa *classe di eventi*⁹¹ – non comporta che ogni atto naturalisticamente idoneo a provocare tali esiti rientri tra quelli che si ha il *dovere di riconoscere*. Occorre infatti considerare: da un lato, la sopravvenienza di conoscenze relative a rischi non ancora emersi al momento dell'introduzione della regola, e conoscibili solo a seguito del loro concreto manifestarsi (per i quali si ri-esponde il dominio della colpa generica con i relativi accertamenti)⁹²; dall'altro, l'ipotesi che la regola cautelare selezioni soltanto una porzione di rischi, lasciando fuori modalità di verifica dell'offesa che, pur avverabili, restano estranee alla sua funzione preventiva per scelta legislativa.

Nel primo caso, la circostanza fattuale che la prescrizione possa governare un rischio emerso *ex post* non dovrebbe rilevare ai fini della prevedibilità. Una modalità di verifica dell'evento non può ritenersi riconoscibile *ex ante* per il solo fatto che, *ex post*, una regola cautelare vigente al tempo del fatto si riveli idonea a ridurre le probabilità di realizzazione: una simile impostazione comporterebbe la retroazione del giudizio sulla colpa, incompatibile con il principio di colpevolezza e con la logica dello

⁸⁸ Il testo dell'art. 43 c.p., nel definire il delitto colposo, ricomprende espressamente l'ipotesi di violazione di regole cautelari scritte – cioè la colpa specifica –: «Il delitto [...] è colposo, o contro l'intenzione, quando l'evento, anche se preveduto, non è voluto dall'agente e si verifica a causa di negligenza o imprudenza o imperizia, ovvero per inosservanza di leggi, regolamenti, ordini o discipline».

⁸⁹ Secondo la terminologia accolta dalla dottrina penalistica italiana, sono qualificate come *elastiche* quelle regole cautelari che non predeterminano una specifica regola di condotta né individuano puntualmente le situazioni di rischio rilevanti, ma si limitano a enunciare un dovere generale di cautela rispetto a determinati beni, rimettendo all'agente l'individuazione delle misure adeguate nel caso concreto. Per questa loro struttura, le regole cautelari elastiche non operano una selezione normativa preventiva delle configurazioni di rischio rilevanti, rimettendo all'agente l'individuazione delle cautele adeguate nel caso concreto. Ne consegue che il contenuto del *dovere di riconoscere* deve essere ricostruito alla luce del parametro dell'agente modello. La distinzione tra precetti rigidi ed elastici è stata introdotta, come noto, da MARINUCCI (1965), pp. 236 ss.; per una lettura aggiornata di regole rigide ed elastiche nella colpa specifica e come questa assomigli a quella generica se le regole formalizzate sono elastiche, v. CASTRONUOVO (2021), pp. 200 ss.

⁹⁰ Cfr., MARINUCCI (1965), pp. 236 ss.; FORTI (1990), pp. 120-121.

⁹¹ Come afferma FORTI (1990), p. 488 «La gran parte delle disposizioni non esprime l'esigenza di evitare eventi rilevanti alla luce di una norma penale, ma si limita a dettare regole atte alla prevenzione di accadimenti che, rispetto a tali eventi, presentano una più o meno accentuata pericolosità. Tali accadimenti, cioè, vengono presi in considerazione in base a una valutazione del legislatore circa la possibilità o probabilità che da essi consegua un evento lesivo».

⁹² V. *supra*, nota n. 87.

stato di diritto⁹³. Pertanto, l'eventuale riconoscibilità di rischi sorti dopo l'emanazione della regola cautelare potrà essere valutata solo in base al parametro dell'agente modello, così come la percezione che la stessa fosse idonea a evitare l'evento – secondo i canoni, dunque, della colpa generica⁹⁴.

Nel secondo caso – quello in cui il legislatore, nel tipizzare la condotta illecita, selezioni deliberatamente solo una parte dei rischi connessi a una determinata attività – la questione attiene ai limiti finalistici della prescrizione. Il perimetro preventivo della regola cautelare, infatti, non dovrebbe mai essere esteso in via interpretativa per ricomprendere modalità di verifica dell'evento *non assunte dall'ordinamento*, pur se causalmente rilevanti⁹⁵. In caso contrario, causalità e colpa verrebbero indebitamente sovrapposte, piegando il diritto penale a finalità sanzionatorie rispetto a fatti che il legislatore ha deliberatamente escluso dalla regolazione cautelare.

Le considerazioni che precedono – solitamente ricondotte sotto la voce “colpa generica residuale” – assumono, nell'ambito dei rischi dell'intelligenza artificiale, un rilievo peculiare. Qui, il *contenuto del dovere di riconoscere* è destinato a confrontarsi con due profili strutturali: da un lato, la continua evoluzione delle conoscenze tecniche e scientifiche, che comporta la continua emersione di modalità di verifica dell'offesa non sempre intelleggibili *ex ante*⁹⁶; dall'altro, la diffusione capillare dei sistemi di IA in una pluralità di contesti eterogenei, che, anche in ragione del principio di frammentarietà, dovrebbe consigliare al legislatore una selezione dei doveri di condotta e, dunque, dei rischi rilevanti a titolo di colpa penale.

Su questo sfondo si misura meglio la valenza delle regole cautelari positive in materia di IA, le quali potrebbero ingenerare interpretazioni volte a ricondurre sotto l'ombrello della riconoscibilità qualsiasi tipo di rischio connesso ai più vari impieghi di tale tecnologia.

A divenire centrale ai fini della necessaria delimitazione dell'area della prevedibilità è il criterio del *rischio tipico* della regola cautelare: il rischio la cui concretizzazione la cautela è funzionalmente preordinata a prevenire⁹⁷. Nell'ambito

⁹³ A ben osservare, infatti, si finirebbe per addossare al consociato il mancato riconoscimento di rischi che nemmeno l'ordinamento aveva ancora individuato come tali.

⁹⁴ Sul paradigma dell'agente modello, v. ancora *supra*, nota n. 87.

⁹⁵ Un esempio significativo è offerto dalla legge 16 gennaio 2003, n. 3, art. 51, che ha introdotto il divieto di fumo nei locali chiusi aperti al pubblico o adibiti a luoghi di lavoro, al fine di prevenire i danni alla salute derivanti dall'esposizione al fumo passivo. Il legislatore, pur essendo consapevole della nocività del fumo anche in contesti diversi, ha deliberatamente circoscritto l'ambito applicativo del divieto solo a tali luoghi. Ne consegue che, sulla base di tale regola cautelare, non può dirsi oggetto del *dovere di riconoscere*, ad esempio, il rischio per il neonato derivante dall'esposizione al fumo passivo della madre durante l'allattamento o nelle fasi di induzione del sonno. L'eventuale insorgenza di patologie potrà rilevare sul piano della prevedibilità secondo, invece, il parametro dell'agente modello.

⁹⁶ Cfr., MANGIONE (2024), p. 187, il quale scrive: «Lo sviluppo della tecnica ha sempre generato pericoli nuovi, talvolta racchiusi in un alone di opacità che ne ha condizionato le possibilità di previsione ed evitabilità. [...] L'intelligenza artificiale non si sottrae a questa eterna “legge della tecnica” e anzi può dirsi che rappresenta il punto di massima esasperazione di siffatto processo generativo di rischi sempre inediti. Le relative cautele saranno allora ontologicamente dinamiche, aperte al nuovo come all'incerto e sottoposte a un costante lavoro di adattamento e di ricerca».

⁹⁷ La valenza del criterio della realizzazione del rischio tipico ai fini del giudizio di prevedibilità, prima ancora che di quello sull'evitabilità, è stata efficacemente evidenziata in Italia da FORTI (1990), spec. pp. 418 ss. e pp. 482 ss., e trova riconoscimento anche nella dottrina tedesca. Sul punto, si fa parlare il chiaro Autore: «i due criteri – quello della realizzazione del rischio e quello basato sul c.d. comportamento alternativo lecito – [...] devono essere tenuti accuratamente distinti. [...] Il motivo di ciò è che la regola di diligenza consiste sostanzialmente in programmi di

della colpa specifica, tale paradigma può operare infatti come strumento di selezione del contenuto del *dovere di riconoscere* – e di riflesso, come si approfondirà ai § 4 e § 4.1, del *dovere di evitare* –, limitandolo alle sole modalità di verifica dell'evento corrispondenti alla concretizzazione del modello di rischio assunto dall'ordinamento nella regola cautelare.

Beninteso: il criterio del rischio tipico (o nesso di rischio) non è finalizzato a rendere riconoscibili i rischi, ma a determinare *quali devono essere riconosciuti*. Del resto, il rischio tipico non è necessariamente un dato enunciato in modo espresso dalle regole cautelari, bensì una ricostruzione da compiere alla luce della loro funzione preventiva: in termini più chiari – quelli di Forti – occorre chiedersi se il fatto fosse «prevedibile secondo il punto di vista della regola cautelare»⁹⁸.

Giunti qui, può risultare utile rintracciare i parametri generali per l'individuazione dei rischi tipici⁹⁹. Quando nella prescrizione sono indicate in modo tassativo le dinamiche di verifica della lesione rilevanti, l'identificazione del rischio tipico discende direttamente dal suo contenuto legale. Poiché, tuttavia, ciò accade raramente, al fine di individuare il rischio “coperto” dalla regola cautelare si ritiene necessario fare riferimento a un giudizio probabilistico¹⁰⁰: occorre, cioè, accertare se *la modalità* con cui si è verificato l'evento come conseguenza della violazione della regola sia *corrispondente o meno all'esperienza*¹⁰¹. Trattasi di un giudizio che, come osservato in

comportamento diretti a contrastare pericoli *tipici* e dunque non eventi cagionati *tout court*, ma eventi che potranno prodursi attraverso svolgimenti causali *tipici* in relazione alla regola stessa»; «del tutto distinta [...] è la questione se l'osservanza della regola di diligenza sarebbe valsa a evitare l'evento. Ciò non riguarda più il “successo”, verificato *a posteriori*, del modello di *previsione* insito nella regola cautelare, bensì il “successo” della condotta prescritta a far fronte all'evento o, meglio, al rischio tipico *in relazione* a tale modello. [...] Dunque, in sede di imputazione, ci si dovrà chiedere se la cautela fosse congrua a contrastare non già gli sviluppi causali concretamente prodottisi, bensì solo quegli sviluppi che l'esito positivo della verifica della “realizzazione del rischio” abbia *già* verificato come corrispondenti al modello di previsione della regola» (*ivi*, pp. 446 e 447). Giova riportare un ulteriore passaggio dell'opera richiamata: «occorre chiedersi non che cosa poteva, ma che cosa *doveva* fare l'agente per evitare l'evento e, più precisamente, in sede di verifica della “realizzazione del rischio”, che cosa *doveva* rappresentarsi l'agente per far fronte al suo *dovere* di evitare l'evento (o, come si è già più volte ribadito, nelle ipotesi di rischio consentito in senso stretto, di ridurre la probabilità di verifica entro limiti socialmente accettabili)» (*ivi*, p. 472). La rilevanza del criterio in discorso ai fini del comportamento alternativo lecito e, dunque, nel giudizio di evitabilità – cfr., ad. es., CASTRONUOVO (2009), spec. pp. 284 ss. – sarà ripresa *infra*, al § 4.1.

⁹⁸ FORTI (1990), p. 446. Tale criterio, dunque, «assume portata più ampia di quella relativa allo «scopo della norma» alla luce della necessità che ai fini dell'imputazione dell'evento “per colpa” sia stata verificata la congruenza dell'accadimento concreto non semplicemente con il tipo di evento che la regola di diligenza mirava a prevenire, ma altresì con la serie di svolgimenti causali che in relazione a tale finalità preventiva potessero reputarsi tipici» – *ivi*, p. 432-433, nota n. 213.

⁹⁹ *Ivi*, p. 450.

¹⁰⁰ La portata del giudizio probabilistico può essere chiarita mediante un esempio. Si prenda a riferimento la regola cautelare di cui all'art. 148, comma 10 del D. Lgs. 30 aprile 1992, n. 285, che vieta il sorpasso in prossimità di curva su una strada a doppio senso di marcia, la quale è funzionalmente orientata a prevenire il rischio di scontri frontali. Qualora l'agente, violando tale prescrizione, invada la corsia opposta e, nel compiere la manovra, sollevi materiale presente sul manto stradale che finisce per causare lesioni a soggetti che transitano su una pista ciclabile adiacente, l'evento risulterebbe eziologicamente connesso alla violazione. Infatti, l'osservanza della regola cautelare che proibisce il sorpasso avrebbe potuto, nel caso in esempio, evitare l'evento. Cionondimeno, siffatta modalità di verifica non corrisponde al modello di rischio preso in considerazione dalla regola cautelare, non rientrando la prima tra gli sviluppi causali che, secondo un giudizio probabilistico *ex ante*, costituiscono tipica concretizzazione della classe di rischi presa in considerazione. In questi termini, l'evento, pur eziologicamente connesso alla violazione, non può dirsi la concretizzazione del rischio tipico della regola cautelare.

¹⁰¹ Sul punto, è necessario un chiarimento. La tesi secondo cui «l'evento potrà “sensatamente” considerarsi realizzazione del rischio che la norma di diligenza mirava a contrastare soltanto ove la sua verifica non sia stata

dottrina¹⁰², non può essere compiuto in modo uniforme, ma deve graduarsi in funzione del valore dei beni giuridici esposti al rischio.

Alla luce di tali premesse concettuali, viene finalmente in esame il Regolamento (UE) 2024/1689 (d'ora in poi "AI Act" o "Regolamento"), la cui disciplina, articolata in funzione del grado di rischio dei sistemi di intelligenza artificiale, è destinata a incidere – entro le forme sopra delineate – anche sul contenuto del *dovere di riconoscere*.

Per quanto qui interessa, i considerando 5, 28 e 29 richiamano espressamente¹⁰³, sin dall'*incipit* del Regolamento, i potenziali pregiudizi fisici e psicologici derivanti

contraria all'esperienza» è criticata dallo stesso FORTI (*ivi*, p. 439), il quale osserva come «una simile affermazione riveli ben presto la propria insufficiente portata selettiva ai fini dell'imputazione colposa dell'evento». L'Autore richiama, a tal proposito, il caso – ancora una volta – delle disposizioni sul divieto di sorpasso, funzionalmente orientate alla prevenzione di pericoli per la vita e l'integrità fisica: ove la vittima deceda per infarto in conseguenza dello spavento provocato dalla manovra vietata, il criterio fondato sulla mera non contrarietà dell'evento all'esperienza non consentirebbe di escludere la realizzazione del rischio tipico, giacché anche tale esito, pur meno frequente, rientra astrattamente nel novero degli eventi lesivi considerati dalla norma. Nel parametro qui proposto, tuttavia, ciò che si assume quale oggetto del confronto con l'esperienza non è l'evento in quanto tale, bensì le sue modalità di verifica, ossia lo sviluppo causale attraverso cui il rischio si è concretizzato (la morte conseguente alla collisione rientra nel rischio tipico che il divieto di sorpasso intende prevenire, mentre il decesso per infarto indotto dallo spavento, pur causalmente collegato alla manovra, non ne costituisce una *modalità* tipica di realizzazione).

¹⁰² In tal senso, è stato osservato che «tanto più ingente è l'entità dei beni giuridici potenzialmente esposti all'attività pericolosa, tanto più la misura nomologica alla base del giudizio di probabilità dell'evento dovrà avvicinarsi (e, nel caso del bene supremo vita, coincidere) al massimo delle conoscenze del momento» (*ivi*, p. 463). Sviluppando questa prospettiva, l'Autore arriva a ritenere che nei settori in cui le conoscenze nomologiche non hanno raggiunto un livello tale da assicurare un adeguato controllo dei pericoli, «ci si dovrà generalmente astenere dall'attività o, quantomeno, esporre ai suoi possibili effetti dannosi beni giuridici di valore proporzionato all'efficacia delle cautele adottabili» (*ivi*, p. 465).

¹⁰³ Per agevolare la consultazione, si riporta il testo dei considerando nn. 5, 28 e 29 AI Act (corsivo di chi scrive): «L'IA può nel contempo, a seconda delle circostanze relative alla sua applicazione, al suo utilizzo e al suo livello di sviluppo tecnologico specifici, comportare rischi e pregiudicare gli interessi pubblici e i diritti fondamentali tutelati dal diritto dell'Unione. Tale pregiudizio può essere sia materiale sia immateriale, compreso il *pregiudizio fisico, psicologico, sociale o economico*» (n. 5); «L'IA presenta, accanto a molti utilizzi benefici, la possibilità di essere utilizzata impropriamente e di fornire strumenti nuovi e potenti per *pratiche di manipolazione, sfruttamento e controllo sociale*. Tali pratiche sono particolarmente dannose e abusive e *dovrebbero essere vietate* poiché sono contrarie ai valori dell'Unione relativi al rispetto della dignità umana, alla libertà, all'uguaglianza, alla democrazia e allo Stato di diritto e ai diritti fondamentali sanciti dalla Carta, compresi il diritto alla non discriminazione, alla protezione dei dati e alla vita privata e i diritti dei minori» (n. 28); «Le *tecniche di manipolazione* basate sull'IA possono essere utilizzate per *persuadere le persone ad adottare comportamenti indesiderati o per indurle con l'inganno a prendere decisioni in modo da sovvertirne e pregiudicarne l'autonomia, il processo decisionale e la libera scelta*. L'immissione sul mercato, la messa in servizio o l'uso di determinati sistemi di IA con l'*obiettivo o l'effetto di distorcere materialmente il comportamento umano*, con il rischio di causare danni significativi, in particolare aventi *effetti negativi sufficientemente importanti sulla salute fisica, psicologica o sugli interessi finanziari*, sono particolarmente pericolosi e *dovrebbero pertanto essere vietati*. Tali sistemi di IA impiegano *componenti subliminali quali stimoli audio, grafici e video che le persone non sono in grado di percepire poiché tali stimoli vanno al di là della percezione umana* o altre tecniche manipolative o ingannevoli che *sovvertono o pregiudicano l'autonomia, il processo decisionale o la libera scelta* di una persona senza che sia consapevole di tali tecniche o, se ne è consapevole, *senza che sia in grado di controllarle o resistervi* o possa evitare l'inganno. Ciò potrebbe essere facilitato, ad esempio, da interfacce cervello-computer o dalla realtà virtuale, in quanto queste consentono un livello più elevato di controllo degli stimoli presentati alle persone, nella misura in cui possono distorcerne materialmente il comportamento in modo significativamente nocivo. In aggiunta, i sistemi di IA possono inoltre sfruttare in altro modo le vulnerabilità di una persona o di uno specifico gruppo di persone dovute all'età, a disabilità ai sensi della direttiva (UE) 2019/882 del Parlamento europeo e del Consiglio (16) o a una specifica situazione sociale o economica che potrebbe rendere tali persone più vulnerabili allo sfruttamento, come le persone che vivono in condizioni di povertà estrema e le minoranze etniche o religiose. Tali sistemi di IA possono essere immessi sul mercato, messi in servizio o utilizzati con l'*obiettivo o l'effetto di distorcere materialmente il comportamento di una persona e in un modo che provochi o possa verosimilmente provocare a tale persona o a un'altra persona o gruppo di persone un danno significativo, compresi danni che possono essere accumulati nel tempo, e dovrebbero pertanto essere vietati*. Potrebbe non essere possibile presumere che vi sia l'intenzione di distorcere il comportamento qualora la distorsione sia determinata da fattori esterni al sistema di IA, che sfuggono al controllo del

dall'impiego di sistemi di IA idonei a manipolare o a sfruttare le vulnerabilità delle persone¹⁰⁴. Essi, dunque, non solo concorrono a chiarire il rischio tipico, ma possono essi stessi contribuire a rendere riconoscibili, sia pure in termini generali, le possibili modalità di offesa.

Lungi dal tradurre in enunciato una mera postura etico-valoriale dell'Unione europea, la parte motiva del Regolamento prelude alla positivizzazione – maturata anche attraverso il confronto con gruppi di esperti, industria e società civile¹⁰⁵ – del divieto di immettere sul mercato sistemi di intelligenza artificiale che, mediante tecniche

fornitore o del deployer, ossia fattori che non possono essere ragionevolmente prevedibili e non possono quindi essere attenuati dal fornitore o dal deployer del sistema di IA. In ogni caso, *non è necessario che il fornitore o il deployer abbiano l'intento di provocare un danno significativo, purché tale danno derivi da pratiche manipolative o di sfruttamento consentite dall'IA*. Il divieto di tali pratiche di IA è complementare alle disposizioni contenute nella direttiva 2005/29/CE del Parlamento europeo e del Consiglio (17), in particolare le pratiche commerciali sleali che comportano danni economici o finanziari per i consumatori sono vietate in ogni circostanza, indipendentemente dal fatto che siano attuate attraverso sistemi di IA o in altro modo. I *divieti di pratiche manipolative e di sfruttamento di cui al presente regolamento non dovrebbero pregiudicare le pratiche lecite nel contesto di trattamenti medici, quali il trattamento psicologico di una malattia mentale o la riabilitazione fisica*, quando tali pratiche sono svolte conformemente al diritto applicabile e alle norme in ambito medico, ad esempio il consenso esplicito delle persone fisiche o dei loro rappresentanti legali. Inoltre, le pratiche commerciali comuni e legittime, ad esempio nel settore della pubblicità, che sono conformi alla normativa applicabile non dovrebbero essere considerate di per sé come pratiche consentite dall'IA manipolative o dannose» (n. 29).

¹⁰⁴ Pur non potendosi parlare *a priori* di “manipolazione” (intesa come «tipo di influenza che aggira o sovverte le capacità relazionali del bersaglio» – COONS e WEBER, 2014, p. 11, traduzione di chi scrive) i *chatbot therapists* sono *strutturalmente* progettati per influire sugli stati emotivi e la dimensione relazionale dell'utente, con la finalità di promuovere condizioni percepite di maggior benessere. In questo modo, sebbene non si faccia necessariamente ricorso a tecniche subliminali, l'interazione continuativa con tali sistemi può incidere sui processi di autoregolazione emotiva e decisionale, producendo forme di condizionamento suscettibili di alterare i processi di autodeterminazione in soggetti in condizioni di vulnerabilità psichica. Per il discrimine tra pratiche *persuasive* (o di “*nudging*”) e *manipolative*, con specifico riguardo al divieto di cui all'art. 5, par. 1, AI Act, cfr. PANATTONI (2024), pp. 525-527, nonché PIETRELLA (2025), pp. 101 ss. Occorre, tuttavia, precisare che il ricorso a tecniche subliminali o a tecniche «volutamente manipolative o ingannevoli» costituisce requisito rilevante ai fini del divieto di cui all'art. 5, par. 1, lett. a), AI Act. Diversamente, il divieto previsto alla lett. b) prescinde dall'impiego di tali tecniche e riguarda i sistemi di IA che sfruttano le vulnerabilità di una persona fisica o di uno specifico gruppo di persone dovute all'età, alla disabilità o a una specifica situazione sociale o economica, quando ciò avvenga con l'obiettivo o l'effetto di distorcere materialmente il comportamento in modo tale da provocare, o da poter ragionevolmente provocare, un danno significativo (il testo integrale delle disposizioni viene riportato *infra*, alla nota n. 106). Con specifico riferimento ai *chatbot therapists*, appare pertanto plausibile che la rilevanza applicativa dell'art. 5, par. 1, lett. b), sia destinata a prevalere, risultando tendenzialmente residuali le ipotesi in cui il sistema sia deliberatamente progettato per manipolare il comportamento degli utenti in senso lesivo. Più frequenti paiono, invece, le situazioni in cui il sistema, pur non essendo concepito con finalità intenzionalmente manipolative, finisca per sfruttare condizioni di vulnerabilità degli utenti, producendo effetti distorsivi del comportamento. In tal senso, nella [Comunicazione della COMMISSIONE EUROPEA \(2025\)](#) – d'ora in poi “Linee Guida della Commissione UE sulle pratiche vietate” si riconosce espressamente che «un chatbot terapeutico volto a fornire sostegno alla salute mentale e strategie di risposta alle persone con disabilità mentali può sfruttare le loro limitate capacità intellettive per indurle ad acquistare prodotti medici costosi o per spronarle a comportarsi in modi dannosi per sé o per altri» (*ivi*, p. 39). Nondimeno, al fine di evitare indebite restrizioni dell'analisi, nel prosieguo si farà riferimento tanto alla lett. a) quanto alla lett. b) dell'art. 5, par. 1, AI Act.

¹⁰⁵ Nel 2018 la Commissione europea ha avviato un processo di consultazione con industria, società civile e mondo accademico, che ha costituito la base della successiva proposta di regolamento sull'intelligenza artificiale. In questo contesto la Commissione ha istituito l'*High-Level Expert Group on Artificial Intelligence*, incaricato di elaborare orientamenti etici e raccomandazioni per uno sviluppo dell'IA rispettoso dei diritti fondamentali e dei valori dell'Unione, e ha promosso una consultazione pubblica. Cfr. [Com. UE, Staff Working Document – Impact Assessment, SWD\(2021\) 84 final, 21 aprile 2021, Annex 2](#). Successivamente, la Commissione ha lanciato una consultazione specifica sulla definizione di “sistema di intelligenza artificiale” e sulle “pratiche vietate”; l'analisi dei contributi pervenuti è stata pubblicata nel maggio 2025: cfr. [Com. UE, Analysis of stakeholder feedback on AI definitions and prohibited practices, 2025](#).

manipolative o di sfruttamento delle vulnerabilità, risultino idonei a cagionare danni fisici o psicologici agli individui o a gruppi di persone (art. 5, par. 1, lett. a e b)¹⁰⁶.

Rispetto ai sistemi di IA che non svolgono pratiche vietate, il Regolamento non assume una posizione neutra, ma li sottopone, ove connotati da un significativo potenziale lesivo, alla disciplina dei c.d. “sistemi di IA ad alto rischio”. Tale qualificazione ricomprende, ai sensi dell’art. 6 e dell’Allegato III dell’AI Act, due categorie: da un lato, i sistemi che costituiscono prodotti, o loro componenti di sicurezza, soggetti alla normativa di armonizzazione dell’UE e alla relativa valutazione di conformità da parte di terzi¹⁰⁷; dall’altro, i sistemi destinati agli impieghi elencati nell’Allegato III, tra cui quelli concernenti la biometria, le infrastrutture critiche, l’occupazione, l’accesso a servizi essenziali (non la mera erogazione), le attività di contrasto, migrazione e amministrazione della giustizia. Per queste realtà tecnologiche, il capo III del Regolamento prevede un complesso di regole volte a garantire, nell’essenza, che l’essere umano mantenga il ruolo di soggetto consapevole, responsabile e insostituibile nel governo della tecnologia dell’intelligenza artificiale. Tali obblighi e requisiti tecnico-organizzativi ricomprendono: l’istituzione di un sistema di gestione del rischio che accompagni l’intero ciclo di vita di un sistema (art. 9); obblighi di corretta governance dei dati e della loro qualità (art. 10); la predisposizione di una documentazione tecnica e di registri delle operazioni (artt. 11-12); l’adozione di misure di sorveglianza umana, volte ad assicurare che l’operatore possa mantenere un controllo sull’attività dell’IA – c.d. *human oversight* (art. 14). A ciò si aggiungono i requisiti relativi all’accuratezza, robustezza e cibersicurezza (art. 15), gli obblighi inerenti alla conservazione dei *log*, alla procedura di valutazione di conformità e alla marcatura CE (art. 16, in relazione agli artt. 43 e 48), nonché la sorveglianza post-commercializzazione (art. 72) e la comunicazione di incidenti gravi all’Autorità competente (art. 73).

Venendo ora alla classificazione dei *chatbot therapists* sulla base del Regolamento, la formulazione dell’art. 5 AI Act non consente di ricomprenderli *a priori* tra le pratiche

¹⁰⁶ Per consentire un immediato riscontro, si riporta il testo vigente dell’art. 5, par. 1, lett. a) e b), AI Act: «Sono vietate le pratiche di IA seguenti: a) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la loro capacità di prendere una decisione informata, inducendole pertanto a prendere una decisione che non avrebbero altrimenti preso, in un modo che provochi o possa ragionevolmente provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo; b) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo».

¹⁰⁷ Tra questi possono rientrare le applicazioni di IA impiegate nell’ambito delle c.d. *digital therapeutics (DTx)*: interventi terapeutici guidati da *software* e clinicamente validati, finalizzati alla prevenzione, diagnosi o trattamento di patologie. Le *DTx*, in quanto dispositivi medici, rientrano nel perimetro di applicazione del regolamento (UE) 2017/745 e sono, pertanto, soggette alle relative procedure di valutazione della conformità e marcatura CE – cfr., [FARMINDUSTRIA](#) (2023); sulla qualifica delle *DTx* tra i dispositivi medici, cfr., [BACCAGUTTI](#) (2025). In simili applicativi – che, a differenza dei *chatbot therapists*, necessitano generalmente di prescrizione medica (*ibidem*) – il sistema di IA può integrare un componente di sicurezza del dispositivo o costituire esso stesso il dispositivo, con conseguente qualificazione come “sistema di IA ad alto rischio” ai sensi dell’art. 6, par. 1, del Regolamento. Un esempio è rappresentato da “*Deprexix*”, *software* terapeutico digitale per il trattamento della depressione, sviluppato in Germania dalla società GAIA AG.

vietate¹⁰⁸, né queste applicazioni risultano, allo stato, riconducibili formalmente alla categoria dei “sistemi di IA ad alto rischio” delineata dall’art. 6 e dall’Allegato III. Tali agenti conversazionali, in quanto destinati a interagire direttamente con persone fisiche, risultano assoggettati – almeno sul piano formale del Regolamento – ai soli obblighi di trasparenza previsti dall’art. 50 AI Act.

L’assenza di una collocazione immediata dei *chatbot therapists* tra le “pratiche vietate” e i “sistemi di IA ad alto rischio” non deve, peraltro, essere interpretata quale giudizio implicito “di innocuità” da parte del legislatore unionale, da contrapporre alle criticità fatte emergere dalla letteratura di settore e dall’esperienza applicativa. Essa sembra piuttosto riflettere la fisiologica asimmetria tra regolamentazione e progresso tecnico-scientifico¹⁰⁹, tanto più marcata in un ambito – quale quello dell’intelligenza artificiale – connotato da un’evoluzione esponenziale e proteiforme¹¹⁰.

La considerazione, in sede europea, dei rischi connessi ai *chatbot therapists* trova conferma nelle Linee Guida della Commissione UE sulle pratiche vietate, pubblicate posteriormente il 4 febbraio 2025 ai sensi dell’art. 96, par. 1, lett. b), del Regolamento. Tale documento, diretto a precisare la portata applicativa dell’art. 5 AI Act, ancorché non includa automaticamente nell’elenco delle pratiche vietate i *chatbot therapists*, li individua espressamente come esempi paradigmatici di applicazioni *suscettibili* di arrecare lesioni alla salute fisica o psicologica degli utenti. Le Linee guida individuano, in via esemplificativa, alcune modalità di impiego dei *chatbots* idonee a produrre simili pregiudizi¹¹¹.

¹⁰⁸ Sul punto, v. meglio *infra*, § 3.

¹⁰⁹ Un esempio paradigmatico di tale scarto è rappresentato dall’intelligenza artificiale generativa, la cui rapida diffusione – in particolare dopo la pubblicazione di modelli conversazionali di larga scala come “ChatGPT” – è intervenuta mentre i lavori di approvazione dell’AI Act erano ancora in corso, imponendo al legislatore unionale di integrare *in itinere* la disciplina dei “modelli di IA per finalità generali”. I *chatbots* terapeutici non rientrano però necessariamente, sul piano concettuale, nella definizione prevista dall’art. 3, n. 63, AI Act, secondo cui: «un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l’autosupervisione su larga scala, che sia caratterizzato da una generalità significativa e sia in grado di svolgere con competenza un’ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato». I *chatbot therapists*, infatti, risultano generalmente progettati e addestrati per una funzione specifica – l’interazione orientata, benché in forme diverse, al sostegno psicologico (v. *supra*, § 1) – e non per un impiego generalizzato e polifunzionale. Resta tuttavia affidata alla prassi applicativa e, in ultima istanza, all’interpretazione della Corte di giustizia dell’Unione europea la precisa delimitazione del confine tra modelli di IA “per finalità generali” e sistemi conversazionali specialistici, tra cui sembrano rientrare i *chatbots* terapeutici.

¹¹⁰ A ridurre tale divario non contribuisce la struttura tassonomica dell’AI Act, che consente di aggiornare la classificazione dei “sistemi di IA ad alto rischio” soltanto mediante l’adozione, da parte della Commissione europea, di un atto delegato. Sul punto, CONSULICH (2023), p. 737-738; CANATO (2024), pp. 21-22.

¹¹¹ Nelle Linee Guida della Commissione UE sulle pratiche vietate si afferma: «Il danno psicologico è particolarmente importante nel contesto dei sistemi di IA che utilizzano tecniche manipolative che sfruttano le vulnerabilità cognitive ed emotive e influenzano i comportamenti di una persona in modi che possono provocare danni significativi [...]. Ad esempio, un’applicazione di IA da compagnia progettata per emulare modelli di linguaggio, comportamenti ed emozioni umani utilizza caratteristiche antropomorfe e segnali emotivi per influenzare i sentimenti, le disposizioni e le opinioni degli utenti, rendendoli dipendenti emotivamente dal servizio che incentiva comportamenti simili alla dipendenza e può provocare danni significativi, come comportamenti suicidi e rischi di nuocere ad altre persone» (p. 31); «Un sistema di IA che provoca danni fisici può anche causare traumi psicologici, stress, ansia, e viceversa. Ad esempio, sistemi di IA progettati per generare dipendenza, utilizzati in prodotti e in altre applicazioni basate sull’IA, possono provocare danni psicologici, incoraggiando comportamenti di dipendenza, ansia e depressione. Il disagio psicologico può a sua volta causare danni fisici, come l’insonnia e altri problemi di salute e fisici legati allo stress» (p.

Del resto, i *chatbots* terapeutici rappresentano un esempio particolarmente significativo dei casi in cui l'IA può avere «l'effetto di distorcere materialmente il comportamento umano, con il rischio di causare danni significativi, in particolare aventi effetti negativi sufficientemente importanti sulla salute fisica, psicologica» (considerando n. 29 AI Act).

Anche limitando lo sguardo all'unica disposizione del Regolamento che, sul piano formale, ricomprende tali sistemi – l'art. 50 – è possibile cogliere una finalità preventiva rispetto ai rischi per la salute mentale derivanti dal progressivo, e non sempre consapevole, affidamento dell'utente nelle risposte generate dall'IA e dalla capacità del sistema di incidere sui suoi processi decisionali. In tal senso, l'art. 50, al par. 1 stabilisce che «i fornitori garantiscono che i sistemi di IA destinati a interagire direttamente con le persone fisiche sono progettati e sviluppati in modo tale che le persone fisiche interessate siano informate del fatto di stare interagendo con un sistema di IA» (corsivo di chi scrive), rendendo così esplicita la funzione cautelare della regola.

I rischi per la salute derivanti dalla «tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'output» risultano espressamente (e non solo ontologicamente) riconosciuti nella formulazione dall'art. 14 del Regolamento, che li include tra quelli che le misure di sorveglianza umana sono chiamate a ridurre. La disposizione, tuttavia, è riferita ai soli “sistemi di IA ad alto rischio”. Si pone dunque l'interrogativo se i rischi nascenti dall'affidamento verso l'intelligenza artificiale, ove si manifestino nell'interazione con *chatbots* terapeutici siano stati deliberatamente non selezionati dal legislatore unionale e, come tali, vadano sottratti dall'elenco di quelli da “dover riconoscere” ai sensi dell'AI Act, oppure se rappresentino una specifica modalità di realizzazione dell'evento già presa in considerazione dall'Unione europea, ancorché non oggetto di specifica previsione regolamentare (non esseno i *chatbot therapists* qualificati, ad oggi, come sistemi “ad alto rischio”).

La questione assume rilievo anche ai fini del giudizio sull'evitabilità nell'ambito della *colpa specifica*: solo l'individuazione dei rischi tipici della regola cautelare consente di circoscrivere i sotto-eventi che il fornitore era tenuto a evitare mediante la sua osservanza.

Orbene, non appare ragionevole l'esclusione dalla finalità preventiva dello *human oversight* di cui all'art. 14 dell'AI Act il danno alla salute derivante da fenomeni di dipendenza, anche emotiva, nei confronti dei *chatbots* in discorso. Tra i rischi da prevenire, l'art. 14 individua espressamente quelli connessi all'incapacità dell'operatore umano di comprendere i limiti del sistema, di monitorarne correttamente il funzionamento e di evitare fenomeni di affidamento automatico o eccessivo sull'*output* (c.d. *automation bias*). In questi termini, il pregiudizio per la salute derivante dall'interazione con un *chatbot* per il supporto psicologico non può dirsi un esito eccentrico rispetto alla funzione della regola cautelare di cui all'art. 14, ma risulta

32). In ogni caso, anche a voler ritenere, per mera ipotesi, che la mancata collocazione dei *chatbots* terapeutici tra le “pratiche vietate” o i “sistemi di IA ad alto rischio” esprima una presunzione di non pericolosità, ciò non escluderebbe la riconoscibilità dei rischi – e, quindi, l'evitabilità degli eventi – connessi al loro impiego. Non viene meno, infatti, il parametro dell'agente modello, che impone di considerare le conoscenze ricavabili dalle indagini tecnico-scientifiche e dagli indicatori fattuali emersi nella prassi applicativa. In merito, si rimanda *supra*, alla nota n. 87.

costituire, piuttosto, una delle *modalità tipiche* – e, per certi versi, paradigmatiche – di concretizzazione del rischio di affidamento inconsapevole sull'intelligenza artificiale. Un'evenienza che il principio del c.d. *human oversight*, fin dalla sua elaborazione nelle linee guida da parte del gruppo di esperti ad alto livello sull'IA, è pensato per governare¹¹².

Con ciò non si vuole affermare che i *chatbot therapists* debbano essere ricompresi, per via interpretativa, tra i sistemi ad alto rischio ai sensi dell'AI Act e, pertanto, assoggettati *a priori* all'intero complesso degli obblighi per essi previsti. Si intende evidenziare, piuttosto, che dalla mancata inclusione espressa di questi agenti tra i sistemi di cui all'art. 6 e Allegato III del Regolamento non può essere desunto né un giudizio normativo di innocuità (v. *supra*), né che il legislatore unionale abbia deliberatamente escluso, dal novero dei rischi tipici, quelli derivanti dall'autoaffidamento dell'utente su tali particolari agenti conversazionali.

Venendo al criterio della probabilità esperienziale – che, si è riportato *supra*, contribuisce a determinare il rischio tipico delle regole cautelari – va osservato come il pregiudizio per l'incolumità derivante dall'affidamento sull'interlocutore artificiale non costituisce certamente una conseguenza inattesa rispetto alla finalità preventiva dello *human oversight*. Al contrario, l'esperienza applicativa dei *chatbot therapists* – come altresì la letteratura di settore – evidenzia come proprio l'incapacità dell'utente di apprezzare criticamente il rapporto con l'interlocutore artificiale possa comportare, per i motivi e nelle dinamiche esposti al § 1, un aumento del rischio di suicidio in soggetti affetti da disturbi mentali.

Considerazioni analoghe potrebbero essere svolte con riferimento ad altre disposizioni del capo III del Regolamento, quali, ad esempio, quelle relative: al sistema di monitoraggio e gestione del rischio lungo l'intero ciclo di vita del sistema (art. 9); alla trasparenza funzionale a un uso consapevole del sistema da parte del *deployer* e, in via ragionevolmente estensiva, dell'utente (art. 13); all'affidabilità tecnica (art. 15)¹¹³. Non è tuttavia possibile, in questo studio, procedere a un esame parimenti analitico di tali disposizioni (salve le considerazioni riguardanti l'art. 9 ai § 4.1 e 4.2, a cui si rinvia). L'analisi proposta in questo paragrafo mira piuttosto a evidenziare la rilevanza, soprattutto nel contesto della regolamentazione dell'IA, del rischio tipico come criterio di individuazione del contenuto del *dovere di riconoscere*.

¹¹²La *ratio* della disposizione emerge nel seguente passaggio dell'art. 14 AI Act: «La sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti di cui alla presente sezione» (par. 2); «in modo tale che le persone fisiche alle quali è affidata la sorveglianza umana abbiano la possibilità, ove opportuno e proporzionato, di: a) comprendere correttamente le capacità e i limiti pertinenti del sistema di IA ad alto rischio ed essere in grado di monitorarne debitamente il funzionamento, anche al fine di individuare e affrontare anomalie, disfunzioni e prestazioni inattese; b) restare consapevole della possibile tendenza a fare automaticamente affidamento o a fare eccessivo affidamento sull'*output* prodotto da un sistema di IA ad alto rischio ("distorsione dell'automazione") [...]; c) interpretare correttamente l'*output* del sistema di IA ad alto rischio» (par. 4).

¹¹³La questione relativa alla natura di tali prescrizioni – se di *norme giuridiche* che individuano un dovere di diligenza e i relativi destinatari, ovvero di *regole cautelari* che ne specificano il contenuto – sarà esaminata *infra*, § 4.2, propendendo infine per la seconda soluzione.

Del resto, in un ambito destinato a costante evoluzione – come dimostrano alcuni ritocchi già apportati all’AI Act¹¹⁴ – non risulta opportuno ricostruire i rischi tipici ravvisabili dal punto di vista di ciascuna regola cautelare positivizzata nel Regolamento¹¹⁵. Più utile, in questa sede, è parso mettere in luce che, sul piano della ricostruzione dogmatica, l’alveo dei rischi connessi ai *chatbot therapists*, rispetto ai quali sarebbe esigibile un *dovere di riconoscere*, non va vincolato al nesso, operato dall’AI Act, tra la formale (e per certi versi contingente) classificazione normativa di tali sistemi e le regole cautelari che esso contiene, quantomeno ai fini della colpa generica – fatta salva l’ipotesi in cui sia la stessa fattispecie incriminatrice a circoscrivere il *dovere di diligenza* ai soli sistemi di IA ad alto rischio (sul punto, v. *infra*, § 4.1).

2.1.2. L’area del rischio consentito nell’ambito dei *chatbot therapists* prima e dopo l’AI Act.

Affermata la riconoscibilità, quantomeno in astratto, dei principali rischi dei *chatbot therapists*, sussistono ora le condizioni per proseguire l’indagine sul versante dell’altro canone tradizionale della diligenza: l’evitabilità degli eventi considerati. Tale passaggio risulta tanto più necessario se si considera che la prevedibilità dell’offesa, nel contesto di un’evoluzione a cadenza sempre più compressa, genera un “ansioso” interrogativo sul *quid agendum*.

Orbene, muovendo dal piano generale della teoria dell’imputazione colposa, è stato sostenuto che la mera rappresentazione (o rappresentabilità) dell’evento – nel caso in esame, l’aggravamento della patologia e il suicidio di un utente con disturbi psichici – ove non siano rintracciabili regole cautelari che consentano di evitarlo, dovrebbe ricadere nell’ambito del rischio consentito¹¹⁶. Due sono le ragioni a supporto di tale conclusione: da un lato, si argomenta che il dovere di astensione dall’attività non può fungere da misura della diligenza, non essendo il primo una regola modale e riguardando, la seconda, un *fare positivo*¹¹⁷; dall’altro – nel residuo ambito di valutazione concernente il *facere* – si porrebbe in contrasto con il principio di colpevolezza il rimprovero mosso all’agente per non aver individuato l’*ubi consistant* della diligenza,

¹¹⁴ V. le modifiche – pur non incidenti sui profili centrali di questa analisi – apportate all’AI Act dal Regolamento (UE) 2026/1744 (c.d. *Digital Omnibus on AI*), entrato in vigore il 27 luglio 2026.

¹¹⁵ Con ciò non si intende negare l’utilità di una simile ricostruzione, che – rispetto alla diligenza dei fornitori di *chatbot therapists* – fornirebbe il catalogo completo dei rischi doverosi da riconoscere. Si tratta, tuttavia, di un’operazione che, oltre a richiedere uno spazio argomentativo incompatibile con i limiti del presente lavoro, risulterebbe esposta a un rapido superamento. Più opportuno è affidare tale esame al formante giurisprudenziale.

¹¹⁶ In questo senso, nei casi di attività aventi utilità sociale, v. GROTTA (2012), pp. 134-136. L’astensione è invece ritenuta una misura esigibile in assenza di altri mezzi disponibili da MARINUCCI (2005), pp. 46-47; RUGA RIVA (2006), pp. 1764 ss.

¹¹⁷ GIUNTA (1993), pp. 233 ss.; GROTTA (2012), pp. 124-125. Critico rispetto alla tesi che insiste «sulla “natura modale” come qualità assoluta e indefettibile della regola cautelare [...] là dove per “regole modali” si intendano regole che non potrebbero mai consistere in un divieto o, in altri termini, nella pretesa di astensione da un certo comportamento» è CASTRONUOVO (2025), p. 15. Sul punto, cfr. già CASTRONUOVO (2011), pp. 1619 ss.

qualora si versi in condizioni di incertezza circa l'efficacia di possibili cautele (con contrapposta deresponsabilizzazione delle istituzioni pubbliche)¹¹⁸.

Nondimeno, anche aderendo alla tesi che, in simili contesti, esclude l'astensione dal novero delle condotte diligenti esigibili, resta fermo che i rischi connessi allo svolgimento di attività pericolose possono ritenersi consentiti *solo se* le stesse comportino un'utilità di rilievo pubblico o collettivo¹¹⁹. Calando il dogma nel panorama dei *chatbot therapists*, le evidenze emerse dalle ricerche scientifiche, pur mettendone in luce la pericolosità, non ne escludono un'utilità terapeutica o, quantomeno, di sostegno psicologico per gli utenti che vi fanno ricorso¹²⁰. Inoltre, pur venendo in rilievo beni giuridici di rango primario – quali la vita e l'incolumità individuale – i casi documentati di concretizzazione dei rischi più gravi risultano, allo stato, estremamente limitati rispetto alla platea complessiva degli utenti. Il confronto con tale dato empirico, pur non attenuando la gravità dei singoli eventi lesivi, assume rilievo ai fini della valutazione delle probabilità di offesa e, dunque, incide sul bilanciamento tra utilità sociale dell'attività e dimensione del rischio che deriva dal suo svolgimento – che risulta perciò a favore della prima¹²¹.

Ne consegue che, su un piano di astrazione *ante* Regolamento, l'impiego dei *chatbot therapists* si situa, per forza di schema, nell'ambito dei rischi socialmente accettabili. Beninteso, tale inquadramento non esime da – bensì esige – una verifica circa la condotta diligente esigibile dal fornitore (ora assunta nel suo versante del *dovere di evitare*)¹²². In altri termini, l'esistenza di un rischio riconoscibile che, almeno *ante legem*, non comporta un obbligo di astensione – *primum* per l'utilità sociale connessa all'attività, *deinde* per la discussa configurabilità dell'astensione quale misura di diligenza¹²³ – non

¹¹⁸ Così GIUNTA (2006), pp. 241-242.

¹¹⁹ Cfr., per tutti, FORTI (1990), pp. 250 ss., il quale chiarisce: «la controversa dommatica detta *erlaubtes Risiko* [...] è sempre stata connessa alla necessità di delimitare l'area del “rischio illecito” (in senso lato) rispetto a quella, notevolmente più estesa, del “rischio riconoscibile” e, dunque, del rischio *tout court*, in considerazione del valore sociale di determinate attività» (corsivo di chi scrive).

¹²⁰ V., *supra*, § 1. Peraltro, deve osservarsi che nell'ambito dei *chatbot therapists* l'equilibrio – per riprendere le parole di PAGLIARO (1985), p. 646 – «tra l'interesse a che la condotta sia svolta e l'interesse, con quello in contrasto, di salvaguardare un bene particolare dal pericolo di conseguenze nocive prodotte da quella condotta» non può dirsi definitivamente assestato. Infatti, tanto il livello di utilità, quanto il grado di probabilità di verifica del rischio presentano margini di incertezza.

¹²¹ Cfr. FORTI (1990), p. 255, il quale individua come criterio di delimitazione del rischio lecito (o consentito) «il bilanciamento del valore o utilità dell'attività esercitata da una parte con il tipo e le dimensioni del possibile danno dall'altra, con un peso decisivo assegnato al grado di probabilità del verificarsi della lesione».

¹²² ZIRULIA (2018), p. 363, evidenzia come il dovere di diligenza e il rischio consentito non siano due piani sovrapponibili. L'Autore, riprendendo le parole di MILITELLO (1988), p. 140, afferma «mentre “la diligenza dovuta esprime le modalità di una condotta che sono richieste per limitarne le potenzialità offensive”, il rischio consentito “caratterizza la pericolosità dell'azione dopo l'applicazione delle precauzioni necessarie”». Sul concetto di *dovere di evitare*, quale momento successivo al *dovere di riconoscere*, v. *infra*, § 2.2.2.

¹²³ Del resto, sarebbe contraddittorio ammettere lo svolgimento di un'attività pericolosa in ragione della sua utilità sociale e, al contempo, in caso di verifica dell'evento, imputare penalmente all'agente la mancata astensione dal suo esercizio sul presupposto che l'astensione costituisca l'unica misura idonea a evitare l'offesa. Ciò a maggior ragione ove si consideri che, oltre all'interesse alla libertà (anche di iniziativa economica) del soggetto, rileva il carattere ineliminabilmente *fortuito* che connota l'evento nei reati colposi. Su quest'ultimo punto, cfr. CASTRONUOVO (2009), pp. 105 ss. Pertanto, va ritenuto che la configurazione di un generale obbligo di astensione rispetto allo svolgimento di un'attività pericolosa, ma socialmente utile, dovrebbe richiedere una preventiva e inequivoca determinazione legislativa.

esonera dal rispetto delle cautele via via rintracciabili e idonee a ridurre le possibilità di concretizzazione.

Orbene, se ci si confronta con la fase iniziale dell'espansione dei sistemi di *machine learning*, si constata la mancanza di standard tecnico-normativi affidabili (a fronte dell'elaborazione di sole enunciazioni etiche)¹²⁴, né era maturata esperienza circa l'efficacia di eventuali condotte di controllo. In un simile contesto, l'evento offensivo *chatbot* correlato difficilmente avrebbe potuto fondare un addebito colposo in capo al fornitore: difettando regole cautelari di provata efficacia ed essendo l'attività socialmente utile, l'offesa si sarebbe dovuta ricondurre nell'ambito del rischio lecito¹²⁵ – salvo casi eclatanti di gestione manifestamente imprudente di stati psicotici dell'utente. Del resto, come insegna la dottrina, «di una eventuale colpa può cominciare a discutersi, al più presto, in presenza di un comportamento (normativamente, assiologicamente) valutabile come inosservante rispetto a regole cautelari teleologicamente pertinenti al caso esaminato»¹²⁶.

Il quadro attuale appare, però, profondamente mutato. L'adozione dell'AI Act, unitamente al progressivo sedimentarsi dell'esperienza, ha inciso anche sul piano dell'evitabilità dell'evento.

Certo, si tratta di un passaggio che – al pari di quanto rilevato rispetto alla riconoscibilità dei rischi – non sembra prestarsi a una lettura unitaria, ma pare piuttosto da apprezzare in relazione ai diversi contesti applicativi dell'intelligenza artificiale. E ciò per almeno un triplice ordine di considerazioni.

Anzitutto per alcune pratiche di intelligenza artificiale l'AI Act prevede un obbligo di astensione (art. 5). Esse sono dunque sottratte *a priori* dalla logica del rischio consentito, poiché il legislatore ne vieta l'uso e l'immissione sul mercato a prescindere dalla loro utilità sociale.

In secondo luogo, con riferimento alle applicazioni il cui rischio è consentito (Capi III, IV e V dell'AI Act), l'identificazione delle cautele esigibili presuppone una selezione basata sui rischi propri di ciascun ambito di impiego dell'IA. Come si è efficacemente osservato in dottrina, infatti, «l'individuazione della regola preventiva appropriata e pertinente al caso concreto dipende dalla verifica effettuabile sulla base del criterio della *realizzazione del rischio* nell'evento»¹²⁷, considerato l'inesauribile «numero di regole a contenuto preventivo potenzialmente violate in ogni condotta umana»¹²⁸.

¹²⁴ V. *supra*, § 2.1.1, spec. nota n. 85.

¹²⁵ Cfr. PAGLIARO (1985), pp. 646-647, il quale osserva: «non si può dire in alcun caso che, in presenza di prevedibilità ed evitabilità dell'evento, sia, per ciò stesso, affermata la negligenza, imprudenza o imperizia. Alla prevedibilità dell'evento deve sempre unirsi il superamento del rischio consentito».

¹²⁶ CASTRONUOVO (2009), p. 463. Similmente, con specifico riferimento al contesto dell'IA, cfr. anche PICOTTI (2024b), pp. 13-15.

¹²⁷ CASTRONUOVO (2009), p. 289. Ne deriva, dunque, che il criterio del rischio tipico non rileva solo ai fini della determinazione del contenuto del *dovere di riconoscere* (v. *supra*, § 2.1.2), ma anche rispetto al contenuto del *dovere di evitare* (v. meglio *infra*, spec. § 2.2.2 e § 4.1).

¹²⁸ CASTRONUOVO (2009), p. 282.

Infine, il livello dell'esperienza accumulata, anch'essa da valorizzarsi sul piano dell'evitabilità (oltre che della prevedibilità), risulta profondamente eterogeneo nei diversi settori applicativi¹²⁹.

Ne deriva che, nel caso qui esaminato, l'indagine sulla diligenza richiesta al fornitore va riferita unicamente al contesto applicativo dei *chatbot therapists*.

2.1.3. La logica delle regole dell'AI Act applicabili ai *chatbot therapists*: da precauzione a prevenzione.

Prima di fissare la *metodologia* che ci si propone di adottare nel prosieguo dell'indagine – avvicinandoci sempre più al tema, ancor più complesso, dell'evitabilità – è opportuno soffermarsi sulla politica normativa che informa le disposizioni comunitarie applicabili ai fornitori di *chatbot therapists*. Si tratta, cioè, di stabilire se la logica che guida le regole cautelari in discorso debba essere ricondotta alla precauzione o, piuttosto, alla prevenzione¹³⁰. Infatti, sebbene spesso i termini siano interscambiati, tra prevenzione e precauzione si cela una paronimia: la prima mira a contenere o eliminare rischi già individuati dalla scienza, i cui effetti risultano «prevenibili in quanto prevedibili»¹³¹; la seconda, invece, riguarda rischi ancora non conosciuti, ma che lo stato delle scoperte nomologiche non consente di escludere con sicurezza. In altri termini, il principio di precauzione corrisponde a una scelta politica d'intervento di fronte a lacune del sapere scientifico che non forniscono prova della possibile dannosità di un prodotto o di un'attività¹³².

La dottrina penalistica – in una discussione che attraversa ormai qualche lustro di elaborazione scientifica in diversi Paesi – ha posto in evidenza le tensioni sistemiche derivanti dall'assunzione del principio di precauzione a fondamento di una *sanzione punitiva*¹³³. In particolare, è stato osservato che l'addentrarsi del diritto penale nel terreno della precauzione potrebbe avvenire solo al costo della *flessibilizzazione* di «due fattori fondamentali per la configurazione del fatto tipico: i) anzitutto, sul piano oggettivo, il nesso di causalità, sfruttandone la struttura probabilistica e trasformandolo

¹²⁹ Per un quadro esemplificativo, limitandoci alla disomogeneità interna all'ambito sanitario, cfr. MORONE *et al.* (2025) i quali, in un'overview di 161 *systematic reviews* relative a 41 ambiti di medicina clinica, riscontrano una marcata concentrazione di tali studi in oncologia (13,9% delle revisioni incluse), neurologia (7,9%), radiologia (7,9%), gastroenterologia (6,0%) e odontoiatria (5,3%). L'impiego dell'IA risulta orientato per lo più alla diagnosi (44,4% delle *review*) e solo in misura minore alla prognosi/predizione (13,9%) e allo *screening* (9,3%). Anche nei campi della medicina – di cui nello studio si aggregano le specialità meno rappresentate – l'attenzione verso applicazioni dell'IA per finalità diagnostiche resta prevalente (35,3% delle *review*). Tali dati confermano come, anche all'interno del medesimo settore, il livello delle conoscenze scientifiche e dell'esperienza applicativa risulti differenziato.

¹³⁰ In questo segmento dell'analisi, dunque, la questione non investe il possibile ricorso al principio di precauzione quale criterio integrativo di una (residuale) colpa generica, bensì, prima ancora, all'eventuale logica precauzionale delle regole che possono fondare una colpa specifica del fornitore di sistemi di IA.

¹³¹ CASTRONUOVO (2012), pp. 26-27.

¹³² In altri termini, il principio di precauzione legittima forme di tutela «in situazioni di incertezza scientifica, nelle quali sia scientificamente non implausibile, ma non provata, l'ipotesi della pericolosità di certe condotte o situazioni» – PULITANO (2023), p. 168.

¹³³ Cfr., limitandosi alle monografie penalistiche italiane e senza pretesa di esaustività: CASTRONUOVO (2012); MASULLO (2012); CORN (2013); CONSORTI (2013); ORSINA (2015). V. anche D'ALESSANDRO (2021), pp. 975 ss.

o rendendolo prossimo – in sede di accertamento – a un mero e potenziale “nesso di rischio”; ii) in secondo luogo, sul piano soggettivo, la *colpa*, in particolare deformandone, in modo più o meno evidente, gli aspetti propriamente “cognitivi”: la riconoscibilità del rischio o la prevedibilità del risultato»¹³⁴.

Espresso in altra forma, l’assunzione del principio di precauzione a fondamento della responsabilità penale «consentirebbe di attribuire rilevanza penale a pericoli solo sospettati»¹³⁵, quando, invece, anche alla soglia dei reati di pericolo astratto è richiesta la verificabilità empirica della pericolosità della condotta¹³⁶.

Muovendo da queste premesse, qualora, nonostante quanto affermato nelle pagine precedenti – ossia la disponibilità di conoscenze scientifiche sui rischi connessi ai *chatbot therapists* e sulle loro modalità di verifica – residuasse l’assunto che le disposizioni dell’AI Act siano integralmente sorrette da una logica precauzionale¹³⁷, la successiva analisi sulla diligenza esigibile dal fornitore di tali agenti conversazionali – in termini di evitabilità dell’evento – risulterebbe in parte viziata sin dall’origine. Essa apparirebbe fondata su presupposti normativi che, in quanto privi di adeguato supporto empirico-scientifico, non dovrebbero legittimamente fungere da base per un’imputazione colposa.

Venendo al merito, si condivide l’impostazione che, nel risolvere la questione della logica delle regole dell’AI Act, distingue l’ambito delle pratiche vietate (Capo II dell’AI Act) da quello delle applicazioni che, pur presentando livelli di rischio differenti, restano ammissibili nel rispetto di specifici obblighi tecnici e organizzativi (Capi III, IV e V del Regolamento)¹³⁸.

Con riguardo al criterio informatore delle pratiche vietate, una parte della dottrina rinviene una logica precauzionale, osservando che «il divieto si applica indipendentemente dal concreto ambito di utilizzo dell’algoritmo e dalla specifica

¹³⁴ CASTRONUOVO (2012), p. 124.

¹³⁵ CONSORTE (2013), pp. 71-72, la quale rileva, altresì, che «a suscitare perplessità è anche la conciliabilità della natura ‘prospettica e programmatica’ del principio di precauzione (destinato a operare a fronte di una mera minaccia di pericolo) con la natura retrospettiva del diritto penale (che dovrebbe intervenire solo dopo il verificarsi dell’evento offensivo di danno o pericolo)». In senso invece favorevole all’estensione della colpa penale sotto il segno del principio di precauzione, cfr., tra gli altri, RUGA RIVA (2006), pp. 1759 ss.

¹³⁶ Cfr. C. cost., sent. 11 luglio 1991, n. 333, ove si è affermato che: «per quanto riguarda in particolare la configurazione di fattispecie criminose strutturate con riferimento ad un evento di pericolo astratto la giurisprudenza di questa Corte, nel ritenere che le incriminazioni di pericolo presunto non sono incompatibili in via di principio con il dettato costituzionale, ha anche riconosciuto che è riservata al legislatore l’individuazione sia delle condotte alle quali collegare una presunzione assoluta di pericolo sia della soglia di pericolosità alla quale far riferimento, purché, peraltro, l’una e l’altra determinazione non siano irrazionali od arbitrarie, ciò che si verifica allorquando esse non siano collegabili all’*id quod plerumque accidit*». In questa prospettiva, la questione se l’AI Act sia informato da una logica di precauzione o di prevenzione incide direttamente sulla tenuta, sul piano dei principi di offensività e di colpevolezza, sia dei criteri direttivi fissati dalla legge delega n. 132 del 2025, sia soprattutto della disciplina introdotta in sua attuazione con l’art. 437-bis c.p.

¹³⁷ In questo senso, tra gli altri, CONSULICH (2023); CANATO (2024), *passim* –la quale ritiene, inoltre, che quello considerato dall’AI Act sia un «*legal risk* [...] più che un rischio che si fonda sulla realtà del fatto concreto» (ivi, p. 22) –; PANATTONI (2021a), pp. 356-357.

¹³⁸ In tal senso, cfr., per tutti, FRAGASSO (2025b), pp. 350-352, la quale afferma: «a noi pare che il rischio derivante dallo sviluppo dell’intelligenza artificiale abbia un carattere ancipite: per alcuni aspetti, noto; per altri aspetti, ignoto»; l’Autrice prosegue ritenendo che «la disciplina dell’intelligenza artificiale, quantomeno nell’ambito dell’Unione europea, si ispiri, per alcuni aspetti, ad una stretta osservanza del principio di precauzione e, per altri aspetti, ad un bilanciamento rischi-benefici volto ad individuare delle aree di rischio che, seppur ignoto, è consentito. Nell’ambito dell’AI Act, ad esempio, seguono una logica del primo tipo i divieti stabiliti dall’art. 5».

violazione di diritti fondamentali che può derivarne»¹³⁹. Ad avviso di chi scrive, tuttavia, tale lettura – pur condivisibile in via di principio – non sembra sostenibile in assoluto. Invero, con riferimento ai *chatbots* destinati al sostegno psicologico, nei casi in cui il sistema *sia progettato* per fare uso di tecniche subliminali o manipolative al fine di orientare l'utente verso comportamenti pregiudizievoli per la sua salute (e rientri, dunque, nei divieti sanciti dall'art. 5, par. 1 del Regolamento), i rischi per la salute non si collocherebbero sul piano delle ipotesi astratte, risultando riconoscibili anche quanto alle modalità della loro concretizzazione¹⁴⁰. In tal caso, il divieto sembra operare più in ottica di prevenzione che di precauzione.

La logica precauzionale, invece, potrebbe manifestarsi nel caso di *chatbot therapists* che *non* siano deliberatamente progettati per manipolare l'utente, pur sfruttando le sue vulnerabilità. In queste ipotesi – più ragionevolmente riconducibili alla lett. b) del par. 1 dell'art. 5 AI Act – il profilo della precauzione potrebbe emergere al momento della determinazione della soglia a partire dalla quale si configura il divieto. Infatti, la formulazione della disposizione citata – che vieta i sistemi che agiscono «con l'obiettivo o l'effetto di distorcere materialmente il comportamento [...] in un modo che provochi o possa ragionevolmente provocare [...] un danno significativo» – impone di stabilire quando l'effetto di distorsione possa dirsi idoneo a integrare il presupposto del divieto¹⁴¹. Così, se anche l'ipotesi di una mera influenza sull'agente esercitata dal sistema fosse ritenuta sufficiente a integrare tale requisito, pur in assenza di indicatori empiricamente verificabili circa la sua concreta capacità lesiva, si ricadrebbe in un divieto fondato su rischi per l'incolumità soltanto sospettati e non ben individuati.

In relazione, invece, alle applicazioni di IA lecite, la dottrina sopra richiamata individua nel bilanciamento tra rischi e benefici il criterio ispiratore delle prescrizioni, in virtù del quale il legislatore avrebbe delimitato quell'area di rischi che, pur restando ignoti, risultano nondimeno ammissibili¹⁴². A tal proposito, la medesima dottrina valorizza, con ampia argomentazione, i precedenti contributi scientifici per sostenere che i rischi dell'IA sono prevedibili quanto all'*an* – cioè nella loro possibilità di verificarsi – ma non rispetto al *quomodo*, ossia alle modalità concrete attraverso cui potrebbero manifestarsi¹⁴³.

Per quanto qui si ritiene, tuttavia, questa ricostruzione – pur cogliendo con finezza un tratto significativo della disciplina – non pare rilevare per tutte le applicazioni

¹³⁹ *Ibidem*.

¹⁴⁰ Si pensi, a titolo esemplificativo, a *chatbots* progettati in modo tale da rafforzare dinamiche di isolamento dell'utente, scoraggiando lo svolgimento di attività motorie, ovvero orientando lo stesso verso forme di compensazione emotiva disfunzionale, come il consumo di cibo ipercalorico.

¹⁴¹ Il testo integrale attualmente vigente è riportato *supra*, § 2.1.1, nota n. 106.

¹⁴² FRAGASSO (2025b), p. 352.

¹⁴³ *Ivi*, p. 453. Sempre sul punto, in altra parte del suo lavoro monografico, l'Autrice chiarisce con un esempio: «In relazione all'intelligenza artificiale [...] l'*an* dell'offesa è spesso chiaro: ChatGPT può fornire informazioni false e lesive della reputazione; un *autonomous vehicle* può non riconoscere un segnale di stop, e provocare di conseguenza un incidente; un sistema di supporto alla diagnostica può fornire una diagnosi errata. Il produttore è anzi in grado di fornire delle statistiche sul tasso di accuratezza del dispositivo intelligente (c.d. metriche), rilevando che le predizioni effettuate dall'algoritmo sono statisticamente corrette in una certa percentuale di casi – il che, dalla prospettiva inversa, equivale a dire che il dispositivo intelligente può cagionare offese in una certa percentuale di casi. Ad essere ignoto, nelle offese generate dall'intelligenza artificiale, è piuttosto la specifica causa degli eventi lesivi» (*ivi*, p. 351).

consentite dal Regolamento. Nel caso dei *chatbot therapists*, come si è potuto osservare, il rischio non risiede primariamente nell'eventuale contenuto istigatorio rispetto a una condotta autolesiva – generato dal funzionamento (a tratti ignoto) del sistema di intelligenza artificiale – bensì, secondo le acquisizioni che si sono riportate¹⁴⁴, nella configurazione stessa dell'interazione. In particolare, nelle vicende in cui utenti vulnerabili sono giunti al suicidio, tra gli elementi ricorrenti rilevano, alla luce degli studi sul tema, un attaccamento emotivo e un eccessivo affidamento terapeutico riposto nel *chatbot therapist*, favoriti anche dalla progressiva confusione tra interazione “sintetica” e relazione umana.

È dunque tale configurazione interattiva che pare costituire il nucleo dei rischi connessi a questi applicativi dell'IA. Un tipo di rischio – e una modalità di concretizzazione – che, si ribadisce, non può definirsi irriconoscibile e che, anzi, alcune delle regole cautelari dell'AI Act mirano a prevenire¹⁴⁵.

Pertanto, se per principio di precauzione si intende un «criterio di gestione del rischio in situazioni di incertezza scientifica circa possibili effetti dannosi ipoteticamente collegati a determinate attività, installazioni, impianti, prodotti o sostanze»¹⁴⁶, si può ritenere che, rispetto ai rischi connessi ai *chatbot therapists*, le regole dell'AI Act abbiano oltrepassato il terreno della precauzione e, pur senza abbandonarlo integralmente, anche quello del bilanciamento tra rischi-benefici¹⁴⁷, esprimendo ora una logica propriamente preventiva (ferma restando, come si è detto, un'eventuale delimitazione in chiave precauzionale del dovere di astensione di cui all'art. 5 del Regolamento per i *chatbots* non volutamente progettati per far uso di tecniche manipolatorie o ingannevoli).

2.2. L'intelligenza artificiale: da noumeno a fenomeno per la scienza penalistica. Meta-conclusioni sul metodo assunto per il prosieguo dell'indagine sull'evitabilità.

L'analisi si è finora concentrata solo sul primo dei due piani in cui l'intelligenza artificiale appariva collocata nell'ambito noumenico (ossia, nella lettura qui proposta, come una forma di agire non afferrabile con le categorie classiche individuate dalla teoria del reato).

Una volta argomentato, nei paragrafi precedenti, il passaggio di una parte di essa nella realtà fenomenica del diritto penale (per ora, principalmente in termini di riconoscibilità dei rischi), conviene spostarsi sul secondo livello, quello della scienza penalistica, per interrogare i paradigmi epistemici attraverso cui il fenomeno dell'IA – inteso non come essenza metafisica, ma come insieme di pratiche, effetti e condizioni operative – può essere colto, anche (e soprattutto) rispetto all'evitabilità delle offese.

¹⁴⁴ Sul punto si rinvia agli studi e alle osservazioni illustrati *supra*, § 1.

¹⁴⁵ V. *supra*, § 2.1.1.

¹⁴⁶ CASTRONUOVO (2012), pp. 18-19.

¹⁴⁷ Sulla tesi che vede nel bilanciamento tra rischi e benefici la logica delle norme previste dall'AI Act, v. *supra*, nota n. 138.

Lungi dal costituire un rilievo accessorio, la questione riflette le evidenze sin qui maturate e giustifica la prospettiva settoriale dell'analisi che segue¹⁴⁸. Infatti, il concetto generale di "IA" si mostra valido – per il momento – solo sul piano definitorio¹⁴⁹; sul versante della riflessione (anche) penalistica, invece, non consente di cogliere l'agire polimorfo di tale tecnologia.

Per aderire a questa realtà, la ricerca è costretta a misurarsi con la varietà delle applicazioni in termini di funzionamento (già di per sé arcano), contesti di impiego (tanto mutevoli da renderne ardua un'enumerazione) ed esperienza accumulata (a livelli non omogenei).

Quanto emerso nei paragrafi precedenti, con riguardo alla riconoscibilità dei rischi, all'area di rischio consentito e alla logica delle regole cautelari, mostra già l'esigenza di non costringere il fenomeno entro formule unitarie.

La causalità rappresenta un esempio altrettanto eloquente: la sua ricostruzione può variare in funzione del tipo di sistema e delle modalità di impiego¹⁵⁰. Da un lato, l'opacità decisionale, pur spesso evocata come il tratto giuridicamente più problematico dell'IA, non connota l'intero panorama tecnologico, ma solo i sistemi più complessi. All'interno di questi, peraltro, vi sono architetture la cui "scatola nera" risulta, almeno in parte, accessibile attraverso tecniche esterne in grado di ricostruire la mappatura che dall'*input* conduce all'*output*¹⁵¹. Dall'altro, nel caso di *black boxes* non intellegibili, occorre distinguere tra: (a) interventi umani il cui nesso con l'*output* dell'IA è sottoposto a verifica; (b) scelte progettuali che definiscono il perimetro entro cui si dispiega l'autonomia del sistema.

Nel primo caso (a), l'impossibilità di ricomporre il processo causale deriva tanto dalla non tracciabilità del funzionamento dei modelli più complessi, quanto dalla stratificazione degli apporti umani che ne accompagnano lo sviluppo, sicché il problema della causalità risulta, allo stato, difficilmente risolvibile alla stregua della teoria della *condicio sine qua non*.

Nel secondo (b), le scelte progettuali non sono avvolte da un'analogia opacità: esse determinano l'architettura interattiva del sistema e, almeno in astratto, sono ragionevolmente riconducibili agli organi di vertice.

Se, come si è sostenuto nel caso dei *chatbot therapists* (v. *supra*, § 2.1), il nucleo di disvalore rispetto al pericolo riguarda queste seconde (ossia le opzioni "di *design*", ad

¹⁴⁸ L'argomento è stato pertanto assunto come snodo centrale del lavoro: non come premessa da verificare, bensì come approdo metodologico coerente con le evidenze emerse nella prima parte dell'analisi.

¹⁴⁹ Il riferimento è alla definizione di «sistema di IA» contenuta nell'AI Act (art. 3, par. 1, n. 1), alla quale rinviano anche la legge n. 132 del 2025 e il D. Lgs. n. 160 del 2026. Se ne riporta qui di seguito il testo: «"sistema di IA": un sistema automatizzato progettato per funzionare con livelli di autonomia variabili e che può presentare adattabilità dopo la diffusione e che, per obiettivi espliciti o impliciti, deduce dall'*input* che riceve come generare *outputs* quali previsioni, contenuti, raccomandazioni o decisioni che possono influenzare ambienti fisici o virtuali». Per una critica a tale definizione, cfr. FRAGASSO (2025b), pp. 151 ss., ove sono altresì ricostruite le principali obiezioni emerse in dottrina.

¹⁵⁰ Cfr., per tutti, LA VATTIATA (2022), p. 701.

¹⁵¹ Cfr. ORDISH *et al.* (2019), p. 23, che a tal proposito distinguono tra "*black box*" e "*white box*"; HAMON *et al.* (2020), p. 14, che riportano come talvolta vi siano esigenze progettuali, specialmente in termini di accuratezza, che implicino un minor grado di interpretabilità del sistema.

esempio: *non-stop service*¹⁵², possibilità di impostare l'interlocutore su figure di rilievo affettivo, valorizzazione delle dinamiche di *homophily*, presentazione del *chatbot* come sistema esperto di sostegno psicologico), l'indagine sulla causalità muta oggetto, dovendosi rivolgere non al funzionamento del sistema, bensì alla strutturazione del servizio.

Riprendendo il discorso sulla opportunità, anzi necessità, di un approccio tematicamente differenziato, si può dire che, in termini kantiani, le *facoltà* conoscitive del penalista, se impiegate partendo da presupposti che assimilano l'IA a mono-scenario, rischierebbero di condurre a una conoscenza penalistica sì *possibile* e *valida* all'interno dei postulati da cui si muove, ma priva di completa garanzia che tale verità si estenda alla dimensione tecnico-operativa reale dell'intelligenza artificiale.

Nella prospettiva qui delineata, l'analisi penalistica pare dunque poter conseguire maggiore rigore nell'avvenire assumendo come oggetto specifici ambiti applicativi e classi di sistemi¹⁵³.

Vale la pena ribadirlo: quanto precede non deve ridimensionare il rilievo dei lavori che hanno affrontato il rapporto tra responsabilità penale e intelligenza artificiale in chiave più generale. Al contrario, tali contributi – taluni di taglio monografico¹⁵⁴ – costituiscono un imprescindibile patrimonio concettuale, mettendo a disposizione un primo e avvertito armamentario teorico per misurarsi con il tema¹⁵⁵. Ma proprio la ricchezza di tale base ricostruttiva suggerisce l'opportunità di un ulteriore livello di focalizzazione nell'analisi scientifica: dal generale al particolare¹⁵⁶.

¹⁵² Non è priva di rilievo, in questa prospettiva, l'introduzione di un limite temporale all'impiego giornaliero dei *chatbots* da parte di "Character.AI", fissato, alla data di redazione del presente contributo, a due ore al giorno. La revisione della *policy* è stata resa nota dalla piattaforma nel corso dell'autunno 2025.

¹⁵³ Senza assumere tono missionario, una simile specializzazione potrebbe consentire al giurista di assumere il ruolo di interlocutore avvertito dinanzi agli altri saperi coinvolti, le cui categorie concettuali variano in funzione degli ambiti applicativi. Ciò risulta particolarmente rilevante in un contesto di dibattito che si mostra polarizzato tra chi, in nome di un «acritico entusiasmo» tecnologico, riduce il diritto a un mero apparato di vincoli e chi, in una logica di «ingiustificato neoluddismo», auspica un arresto della corsa dell'IA. Le espressioni «acritico entusiasmo» e «ingiustificato neoluddismo» sono qui assunte nell'accezione loro attribuita da STANZIONE (2022), min. 1:02:13. In termini convergenti, seppure attraverso un diverso impianto argomentativo, si vedano le considerazioni offerte da MANGIONE (2024), pp. 91-92, il quale afferma: «Presi dai loro discorsi tecnici – è il monito dei filosofi – i giuristi rischiano di mancare l'appuntamento con la tecnologia, sino a dover prendere atto che nella definizione di algoritmo è delineato il suo potere, con la conseguenza che quel che era di competenza del giurista è diventato, quasi inavvertitamente, di dominio dell'ingegnere, del programmatore e dell'amministrazione di una giustizia matematica».

¹⁵⁴ Il richiamo è a D'AMICO (2025) e a FRAGASSO (2025b). Nel medesimo orizzonte di riferimento si collocano, quali opere di impronta più settoriale, i lavori monografici di cui alla successiva nota n. 156.

¹⁵⁵ Così FRAGASSO (2025b), p. 4, la quale pone a fondamento della propria opera la necessità di elaborare «un quadro interpretativo in grado di *orientare* l'analisi dei più svariati fenomeni criminali che possono vedere coinvolta l'i.a.» (corsivo di chi scrive).

¹⁵⁶ Peraltro, nell'ambito della riflessione penalistica sull'IA, i contributi di taglio più specifico sono stati i primi a trovare sistemazione monografica. In particolare, l'attenzione si è finora rivolta all'ambito delle auto a guida autonoma – cfr., LANZI (2023a); COMPOSTELLA (2024) –, a quello della sanità – cfr. AMORE e ROSSERO (2023), i quali, mediante un raro raccordo tra lo studio empirico-sociologico e l'analisi giuridico-penale, ricostruiscono le ricadute della chirurgia robotica (con particolare riferimento al sistema "da Vinci") e della diagnostica assistita dall'IA sull'imputazione degli eventi infausti; TERRIZZI (2023) –, nonché all'attività d'impresa – MANGIONE (2024). Più nel dettaglio di quest'ultima analisi, l'Autore, muovendo da una premessa epistemologica sull'IA, esamina il suo impatto nell'organizzazione dell'attività d'impresa, con particolare riguardo alla gestione del rischio e alla ridefinizione dei poteri e dei doveri di controllo degli organi societari, anche alla luce dell'AI Act.

2.2.1. Dal livello di risoluzione dell'analisi alla definizione del problema.

Le considerazioni epistemologiche che precedono consentono, a questo punto, di motivare più analiticamente le due scelte che orientano il prosieguo dell'indagine: a) la delimitazione del suo oggetto e b) la definizione del problema. Né l'una né l'altra sono operazioni neutre e richiedono, pertanto, una giustificazione¹⁵⁷.

Con riguardo all'oggetto (a) – quanto emerso nella prima parte del lavoro motiva la decisione di circoscrivere il prosieguo dell'indagine esclusivamente ai *chatbot therapists*, rispetto ai quali risultano riconoscibili rischi per beni giuridici – la vita e l'incolumità individuale¹⁵⁸ – che assumono uno specifico rilievo in relazione alla recente legislazione in materia di IA¹⁵⁹. Ciò consente, e al contempo rende opportuno, di proseguire l'indagine sul terreno dell'evitabilità dell'evento, quale snodo decisivo ai fini della configurabilità dell'addebito.

Su questa base trova giustificazione anche la *definizione del problema* (b). Lo studio – che, come dichiarato in apertura, si colloca sul piano dei presupposti che precedono la selezione delle fattispecie astrattamente configurabili – deve muovere dalla determinazione del *contenuto della diligenza* richiesta al fornitore di *chatbot therapists*, qui intesa come insieme «di quelle massime di esperienza e di quelle regole tecniche cui l'uomo accorto e avveduto si sarebbe attenuto nella situazione concreta»¹⁶⁰. Sul piano

¹⁵⁷ Scriveva CALAMANDREI (1939), p. 110: «Nella scelta del tema vi è già la confessione di una preferenza, e nella formulazione

del problema è già implicito il criterio della soluzione». Per una più esplicita tematizzazione, sul piano metodologico, della «definizione del problema» – sia pure nell'ambito della ricerca interdisciplinare –, v. PASCUZZI (2022), pp. 55 ss., le cui considerazioni hanno sollecitato parte del presente paragrafo.

¹⁵⁸ Non hanno invece richiamato, almeno finora, una specifica attenzione del legislatore le problematiche inerenti al trattamento dei dati personali, benché la logica funzionale dell'IA – tanto più efficace quanto maggiore è la quantità di informazioni disponibili – possa entrare in tensione con i principi di minimizzazione dei dati e di limitazione delle finalità. La questione, pur tutt'altro che marginale, non sembra aver generato inquietudine collettiva né una corrispondente istanza di intervento penale *ad hoc*. Estranea all'asse tematico qui prescelto, essa resta pertanto affidata a distinta analisi.

¹⁵⁹ Si riporta, per comodità del lettore, il criterio direttivo in materia penale previsto nella delega al Governo di cui all'art. 24 della legge n. 132 del 2025: «introduzione di autonome fattispecie di reato, punite a titolo di dolo o di colpa, incentrate sull'omessa adozione o sull'omesso adeguamento di misure di sicurezza per la produzione, la messa in circolazione e l'utilizzo professionale di sistemi di intelligenza artificiale, quando da tali omissioni deriva pericolo concreto per la vita o l'incolumità pubblica o individuale o per la sicurezza dello Stato» (art. 24, comma 5, lett. b). Si riporta altresì il testo dell'art. 437-bis, commi 1 e 3 c.p.: «Chiunque omette di adottare le misure tecniche di sicurezza, previste per la progettazione, l'addestramento, la produzione, l'immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio, idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi ovvero omette di adottare misure di sorveglianza umana è punito con la reclusione da uno a cinque anni, quando da tali omissioni derivi pericolo per la vita o l'incolumità pubblica o individuale. Qualora dal fatto derivi un pericolo per la sicurezza dello Stato, la pena è della reclusione da due a otto anni» (comma 1); «Se taluno dei fatti previsti dal comma primo è commesso per colpa grave, la pena è ridotta da un terzo a un sesto» (comma 3). Per la distinzione tra *pericolo individuale*, riferito alla messa in pericolo della vita, dell'incolumità fisica o della salute di un soggetto indeterminato, esponente della collettività, e *pericolo personale*, riferito invece a un soggetto determinato, v. GARGANI (2020), spec. pp. 2, 10 ss., nonché GARGANI, ZIRULIA e CASTRONUOVO (2023), pp. 810 ss. Il profilo dei possibili rischi per l'incolumità pubblica connessi ai *chatbot therapists*, sin qui affidato a un inciso in nota – v. *supra*, § 2.1, n. 73 – non forma oggetto di specifico approfondimento in questa sede, e potrà misurarsi soltanto nel tratto conclusivo dell'indagine, alla luce delle risultanze emerse al § 5 e § 6, in particolare alle note nn. 243 e 249.

¹⁶⁰ FORTI (1990), p. 138. Giova riportare il passaggio completo: «In relazione alle fattispecie colpose, il giudice dovrebbe cioè assumersi una parte delle funzioni spettanti al legislatore, ricostruendo il fatto tipico lasciato privo di determinazione della descrizione legale [il riferimento è al comportamento doveroso]. Ciò, del resto, non sarebbe da attribuire a una difettosa formulazione del diritto positivo, bensì a una impossibilità, per così dire strutturale, per il

sistematico, infatti, il dovere obiettivo di diligenza gravante sul soggetto agente risulta assumere – tanto nei reati di pericolo, quanto in quelli di danno – una funzione *strutturalmente preliminare*: è attraverso la ricostruzione della condotta doverosa che si delimita la causalità penalmente rilevante¹⁶¹, si determina il comportamento imposto ai titolari di una posizione di garanzia¹⁶², nonché si valuta il grado della colpa (il nuovo art.

legislatore, di fissare una volta per tutte i caratteri dei fatti colposi, a causa dell'indominabile molteplicità di questi e del loro presentarsi in modo diverso a seconda delle irripetibili particolarità del caso concreto. Una tale integrazione giudiziale dovrebbe appunto avvalersi del concetto di "diligenza", da determinare alla luce di quelle massime di esperienza e di quelle regole tecniche cui l'uomo accorto e avveduto si sarebbe attenuto nella situazione concreta» (*ivi*, pp. 137-138). In un altro punto dell'opera il chiaro Autore afferma: «Se della "diligenza" o Sorgfalt può essere data una definizione, questa è senz'altro assai vicina alla concezione engischiana di "impiego doveroso dei mezzi idonei a evitare la verifica di un fatto legalmente disapprovato"» (*ivi*, p. 198).

¹⁶¹ Si fa qui riferimento alla c.d. teoria dell'imputazione oggettiva dell'evento, secondo cui l'attribuzione dell'offesa al soggetto agente non si esaurisce nell'accertamento del nesso di causalità in senso naturalistico, ma richiede la verifica che la condotta abbia creato o incrementato un rischio giuridicamente non consentito e che l'evento rappresenti la concretizzazione di quel medesimo rischio. Tra i sostenitori di tale costruzione dogmatica, che in Germania ha trovato compiuta elaborazione con ROXIN (1979), p. 133 ss., v. CASTALDO (1989), p. 59 ss.; DONINI (2006), spec. pp. 16, 87-88, 157. Per una critica a tale impostazione, cfr. MARINUCCI (1991), pp. 3 ss., pp. 18 ss. (da cui si riprendono i riferimenti a ROXIN). Se ci si colloca in tale orizzonte teorico – su cui si basano le ulteriori considerazioni svolte *supra*, § 2.1 – un'indagine sul nesso di causalità che prescindesse dall'individuazione del dovere obiettivo di diligenza risulterebbe, sotto il profilo metodologico, strutturalmente monca. È pur vero che un'analoga obiezione potrebbe muoversi anche verso uno studio che si concentri sul tema della diligenza senza interrogarsi sui problemi che si porrebbero *in seguito* sul terreno della causalità, specie dinanzi a un evento – quale il suicidio o l'aggravamento del disturbo mentale – caratterizzato da un'eziologia multifattoriale. Rispetto a tale rilievo, tuttavia, la scelta qui operata circa l'oggetto dell'indagine – la diligenza del fornitore – trova la ulteriore giustificazione nell'opzione normativa già delineata dalla legge n. 132 del 2025 e tradotta nell'art. 437-bis c.p.: l'anticipazione della tutela alla produzione di un pericolo ridimensionale, infatti, i problemi attinenti alla causalità. L'evento di pericolo, pur postulando – in quanto "evento" – una base ricostruttiva di tipo causale, si arresta alla produzione di una «rilevante possibilità di verifica di un evento temuto» – FIANDACA e MUSCO (2023), p. 218. Nell'attuale panorama legislativo, dunque, un'indagine sulla causalità incontrerebbe un duplice limite: da un lato, la rapida evoluzione degli studi epistemologici sui *chatbot therapists*, che, pur concorrendo a rendere già riconoscibili i rischi per gli utenti con disturbi psichici connessi a determinate scelte progettuali, non offrono ancora indici sulla probabilità scientifica del loro avverarsi; dall'altro, l'opzione legislativa in favore di fattispecie di pericolo rischierebbe, in questo scenario, di ridurre l'analisi a una mera elencazione di passaggi logico-ricostruttivi, destinata a fungere da semplice griglia argomentativa per la decisione giudiziale (e, quindi, a ribadire schemi già acquisiti sia dalla dottrina sia dal formante giurisprudenziale).

¹⁶² Cfr., per tutti, BRUNELLI (2023), *passim*. L'Autore sottolinea come il *dovere di agire* e il *dovere di diligenza*, pur frequentemente sovrapposti, debbano essere cronologicamente distinti: «prima si stabilisce in base a un criterio giuridico se il soggetto è un debitore di protezione o di controllo, poi si torna a valutare sul piano giuridico se la condotta che avrebbe scongiurato l'evento fosse da lui esigibile (doverosa e possibile) nella situazione tipica» (p. 107). Tale distinzione trova il proprio antecedente sistematico nella ricostruzione elaborata da CASTRONOVO (2009), pp. 284 ss., il quale introduce la separazione tra il «dovere di diligenza» e la «diligenza doverosa», successivamente ripresa anche nella giurisprudenza di legittimità (v. Cass. Pen., Sez. IV, 25 marzo 2016, n. 12478 relativa alla vicenda "Commissione Grandi Rischi"; Cass. Pen., Sez. IV, 6 settembre 2021, n. 32899 sul caso del disastro ferroviario di Viareggio; più recentemente, Cass. Pen., Sez. VI, 11 marzo 2025, n. 9906 concernente il disastro di Rigopiano). Se si assume fino in fondo tale condivisibile impostazione, l'indagine sulla sussistenza di un eventuale obbligo giuridico di impedire l'evento *ex art. 40, comma 2, c.p.* in capo al fornitore di *chatbot therapists* dovrebbe *logicamente precedere* quella sul contenuto del dovere di diligenza. Ciò è ancor più vero se si considera che, in alcuni casi, «la concreta doverosità-esigibilità comportamentale può essere ricavata, ancora prima che nel coacervo dei doveri di diligenza, nell'ambito stesso del dovere di impedire l'evento che gravita sul garante» – BRUNELLI (2023), p. 108. Ciò nondimeno, con riferimento ai fornitori di *chatbot therapists* non sembrano individuarsi, al momento, obblighi di impedimento dell'evento (nel caso di specie, il suicidio dell'utente o l'aggravamento del disturbo mentale), né fondati su un dovere di governo rispetto a una certa fonte di pericolo (c.d. posizione di controllo), né scaturenti da un dovere di tutela su un particolare soggetto (c.d. posizione di protezione). Infatti, rispetto a un'ipotetica posizione di controllo, non si ritiene condivisibile l'indirizzo che ricava la fonte dell'obbligo di impedimento nella *pregressa azione pericolosa* rappresentata dall'immissione in commercio del prodotto. Secondo tale orientamento, affermatosi nella giurisprudenza straniera sul danno da prodotto (pionieristico il caso Erdal sul Lederspray), la sola immissione in commercio – in quanto fatto generatore del rischio – varrebbe a fondare un dovere permanente di sorveglianza sul prodotto e, se del caso, di ritiro dal mercato – cfr. PIERGALLINI (2004), p. 54. In proposito, possono formularsi due rilievi. Anzitutto,

437-bis, comma 3, c.p. circoscrive ai soli casi di colpa grave la rilevanza penale, a titolo colposo, dell'omissione di cautele in materia di IA)¹⁶³; nella prospettiva normativa della colpevolezza, è ancora la misura della diligenza a segna(la)re i limiti del rimprovero penale¹⁶⁴.

l'accertamento della pericolosità del prodotto interviene sovente solo dopo un apprezzabile lasso di tempo dalla sua immissione in commercio, all'esito del consolidarsi di significativi episodi di danno – profilo che, nell'attuale scenario dei *chatbot therapists* (v. *supra*, § 2.1.2) potrebbe ritenersi delineato solo in una prospettiva accentratrice delle esigenze di tutela e che non tenga conto della possibile utilità del servizio (nondimeno occorre rilevare il caso del *chatbot* “Woebot”, ritirato dal mercato nel luglio 2025 a seguito dell'emersione di criticità connesse al suo impiego; la notizia è riportata da Il Sole 24 Ore, *Che cosa serve all'IA per potersi occupare di salute mentale*, DA ROLD, 6 ottobre 2025). Ma soprattutto, anche volendo aderire a tale lettura rigorosa, ritenere che la mera immissione in commercio integri una pregressa azione pericolosa idonea a fondare un obbligo permanente di impedimento finirebbe per far poggiare l'illecito penale di evento (dannoso o pericoloso) su una mera condotta, il cui baricentro sarebbe individuato nella percezione del rischio insito nella circolazione del prodotto – cfr. PIERGALLINI (2004), p. 232. Diverso è il caso in cui la posizione di garanzia trovi fondamento diretto nella legge. Sotto questo profilo, la configurabilità di una posizione di controllo in capo al fornitore di *chatbot therapists* deve essere esclusa con riferimento agli obblighi di vigilanza “post-market” previsti dal Titolo I del d.lgs. 6 settembre 2005, n. 206 (Codice del consumo). Tali disposizioni impongono al produttore di immettere sul mercato soltanto prodotti sicuri e di adottare, ove necessario, misure di ritiro, richiamo e informazione (art. 104), ma presuppongono la presenza di un “prodotto” ai sensi dell'art. 3, comma 1, lett. e), Cod. cons. La nozione di prodotto, pur estendendosi anche ai beni forniti nell'ambito di una prestazione di servizi, è tradizionalmente riferita a beni mobili materiali. I sistemi di intelligenza artificiale offerti in modalità *stand-alone* o *as-a-service* – come i *chatbot therapists* – non rientrano, allo stato, in tale definizione, in quanto privi di un *hardware* specifico immesso sul mercato dal fornitore e configurati piuttosto come servizi digitali continuativi. Va peraltro considerato che la Direttiva 2024/2853/UE estende la nozione di “prodotto” sino a ricomprendervi anche il *software* (art. 4, par. 1); essa, al momento di redazione del presente contributo, non risulta ancora recepita (il termine per l'attuazione è fissato al 9 dicembre 2026). Con riguardo agli obblighi previsti dall'AI Act per i sistemi di IA qualificati come “ad alto rischio”, il Regolamento impone ai fornitori una serie di adempimenti particolarmente incisivi: obblighi di gestione del rischio e di conformità (artt. 16-17), misure correttive – tra cui il ritiro, il richiamo o la disattivazione dei sistemi non conformi (art. 20) – nonché obblighi di monitoraggio *post-market* e di segnalazione degli incidenti gravi alle autorità competenti (artt. 72 e 73). Tuttavia, come già rilevato (v. *supra*, § 2.1.2), i *chatbot therapists* non rientrano, allo stato, tra i sistemi di IA classificati come “ad alto rischio”. Ne consegue che tali obblighi, pur astrattamente idonei a delineare un potere-dovere di controllo sul prodotto, non risultano applicabili ai fornitori di questi agenti conversazionali. Se è vero che il *contenuto del dovere di diligenza*, come si argomenterà *infra*, § 4.1, non è integralmente coperto dal principio di legalità (cfr. CASTRONUOVO, 2009, spec. pp. 286-287) – potendo essere oggetto di specificazione interpretativa entro i limiti segnati dal principio di colpevolezza – lo stesso *non può dirsi* per l'obbligo di impedimento fondato sulla legge, il cui ancoraggio normativo costituisce presupposto ontologico della responsabilità omissiva impropria ai sensi dell'art. 40, comma 2, c.p. Un approfondimento autonomo meritano i divieti di cui all'art. 5, par. 1, lett. a) e b), AI Act. In particolare, si deve verificare se, allorché il fornitore sia posto in condizione di conoscere, alla luce dell'esperienza d'uso e delle evidenze disponibili, effetti di distorsione dei comportamenti degli utenti idonei a incidere negativamente sulla salute, il divieto in parola sia suscettibile di fondare un obbligo di ritiro dal mercato configurabile come posizione di controllo ai sensi dell'art. 40, comma 2, c.p., ovvero se esso debba essere inteso quale regola cautelare. La questione, che investe direttamente il tema dell'evitabilità dell'evento, è affrontata *infra*, § 3.1. Un distinto approfondimento merita, altresì, il profilo della c.d. posizione di protezione. Infatti, occorre interrogarsi se il fornitore di un *chatbot therapist*, in ragione dell'eventuale affidamento che il servizio è idoneo a generare, possa assumere nei confronti dell'utente una posizione di garanzia. L'eventuale riconoscimento di un simile obbligo di protezione comporterebbe un significativo ampliamento del contenuto del dovere di diligenza, il quale, per quanto consentito dal principio di colpevolezza, verrebbe a coincidere con l'adozione di qualunque misura utile a prevenire la realizzazione dell'evento lesivo – cfr. MASULLO (2013), p. 60. La questione, per le evidenti ricadute sul perimetro della diligenza, è ripresa *infra*, § 4.2.

¹⁶³ Cfr., tra i diversi studi, quelli di: PADOVANI (1969), pp. 877 ss.; FORNASARI (1990), pp. 334 ss.; BASILE (2005), p. 652; MANTOVANI (2007), pp. 345 ss.; CASTRONUOVO (2009), pp. 350 ss.; POLI (2021), *passim*. Sul punto, CASTRONUOVO (2009), p. 353 – richiamando anche PADOVANI (1969), pp. 880 ss. – introduce una precisazione di particolare rilievo ai fini del discorso: «comportamenti imprudenti, divergenti dallo standard di riferimento in misura rilevante, quindi caratterizzati da una prevedibilità in grado elevato da parte dell'agente-modello, possono comunque non risultare ‘gravemente colposi’ sulla base di ulteriori elementi, attinenti per esempio a fattori motivazionali che hanno inciso sullo svolgimento dei fatti, rendendo meno esigibile o radicalmente inesigibile una condotta cautelare».

¹⁶⁴ CASTRONUOVO (2009), p. 567, il quale afferma: «secondo l'acquisizione tradizionale della concezione normativa della colpevolezza, questa implica, ‘in ambiente colposo’, la possibilità di muovere un rimprovero al soggetto che ha realizzato un fatto involontario e con violazione di regole di condotta a contenuto preventivo».

Inoltre, nei casi in cui tra la condotta del fornitore e l'offesa arrecata all'utente si inserisca la condotta colposa di terzi (ad esempio, i programmatori "a valle") oppure rilevi l'autoesposizione al pericolo della persona offesa, è sempre sul terreno della diligenza – e non della causalità – che si risolve la questione dell'imputazione¹⁶⁵.

Infine, la determinazione della condotta diligente rileva non solo nelle fattispecie colpose, ma anche rispetto ai reati di dolo. Nell'ambito dei fatti dolosi, l'individuazione della condotta doverosa rappresenta un passaggio preliminare: da un lato, «la volontà di realizzazione diretta alla lesione di un bene giuridico diventa rilevante solo quando l'agente al tempo stesso, con il suo comportamento, urta contro un dovere di diligenza»¹⁶⁶ (e non certo contro giudizi morali)¹⁶⁷; in secondo luogo, lo scostamento dallo standard obiettivo di diligenza, nonché l'ipotetica reiterazione della deviazione, assumono valore indiziario del dolo eventuale¹⁶⁸.

Alla luce di tali premesse, l'analisi sulla diligenza riguarderà:

A. la "diligenza esteriore" (*äußere Sorgfalt*) – ossia la «qualità del comportamento dell'agente che lo caratterizza come conforme al dovere»¹⁶⁹ – e non, invece, la "diligenza

¹⁶⁵ Cfr. FORTI (1990), pp. 596-603 e pp. 610-622.

¹⁶⁶ KRAUSS (1964), pp. 47 ss. Il riferimento è ripreso da MARINUCCI (1991), p. 27, il quale, muovendo dalla centralità del dovere di diligenza anche nei reati dolosi, giunge a formulare la nota massima «non c'è dolo senza colpa» (*ivi*, pp. 28 ss.). In senso convergente FORTI (1990), p. 385 ss., il quale osserva: «normalmente le azioni con le quali taluno voglia realizzare gli estremi di un tipo delittuoso sono anche obiettivamente "contrarie alla diligenza" e quando ciò non accade è perché ci si trova di fronte a casi "totalmente fuori dalla realtà"» (*ivi*, pp. 388-389). Il chiaro Autore aggiunge: «Ove sul piano oggettivo la condotta rimanga nell'ambito del dovere di diligenza e, dunque, del rischio consentito, non si avrà colpa, ma non si avrà nemmeno dolo, dal momento che al soggetto che pure abbia voluto l'evento, non potrà muoversi alcun rimprovero fondato sul *dovere di riconoscere* la verifica dell'evento» (*ivi*, p. 390). In chiave critica rispetto alla formulazione più radicale della massima – ossia nell'accezione secondo cui «non c'è dimensione soggettiva del reato doloso se non c'è dimensione soggettiva del reato colposo» – GROTTI (2012), pp. 245-246. Nega che il dolo "contenga" la colpa, ma individua nell'*aumento del rischio* la base comune di entrambi, DONINI (1991), pp. 329 ss., 351 s., nonché *Id.*, (2006), *passim*. In quest'ultimo contributo il chiaro Autore rileva: «Il principio di *laicità* preclude che si possano imputare più rischi a chi ha una volontà più malvagia, quando l'ordinamento ha consentito di agire fino ad una certa soglia di pericolo: quella soglia vale oggettivamente, e se sarà utilizzata con malizia, per la causazione dolosa di un danno, dovrà essere [...] superato il livello di rischio lecito» (*ivi*, p. 55). E prosegue: «Vale dunque la regola: *non c'è dolo (giuridicamente rilevante) senza rischio colposo*, cioè senza il rischio sufficiente per la colpa (non necessariamente del medesimo soggetto agente, ma) di qualcuno, di qualche agente modello» (*ivi* p. 55). In questa direzione, MILITELLO (1988), pp. 207 ss.

¹⁶⁷ Il richiamo al parametro normativo non implica adesione alla tendenza – sviluppatasi soprattutto in una parte della dottrina tedesca – a "normativizzare" i requisiti strutturali del dolo, finendo così per comprimere il momento volitivo e sopravvalutare quello rappresentativo – cfr., recentemente, CIVELLO (2023), pp. 1097 ss. Il riferimento va inteso piuttosto come corollario della concezione normativa della colpevolezza, fatta propria dalla maggior parte della manualistica italiana – cfr. FIANDACA-MUSCO (2023), pp. 336-338. In questa prospettiva, poiché il giudizio di rimprovero si incentra sull'atteggiamento antidoveroso della volontà, la previa determinazione dell'atteggiamento doveroso costituisce un necessario antecedente logico. Resta fermo, tuttavia, che l'apprezzamento normativo del comportamento non dovrebbe escludere la necessità di accertare l'adesione psicologica dell'autore al fatto, quale presupposto della responsabilità personale. Su quest'ultimo punto, per tutti, cfr. EUSEBI (1993), spec. pp. 21-22, 83 ss.

¹⁶⁸ In proposito, MANGIONE (2024), p. 102, scrive: «La calcolabilità [resa possibile dalla chiara delimitazione del dovere di diligenza e dalla conoscibilità delle conseguenze in caso di sua inosservanza] scolpisce il contenuto del dolo eventuale e, al contempo, dona concretezza all'ennesima creatura del diritto ossia l'agente modello della colpa» (sulla figura dell'agente modello, v. *supra*, § 2.1.1, nota n. 87). Del resto, il principio personalistico della responsabilità penale – in Italia riconosciuto all'art. 27, comma 1, Cost. – ambisce a «valutare l'atteggiamento e il processo deliberativo dell'agente concreto, se di rifiuto e voluto contrasto, di mera indifferenza o errata valutazione circa le prevedibili conseguenze delle proprie azioni» – *Ibidem*.

¹⁶⁹ ENGISCH (1930), p. 273. Il passo è ripreso da FORTI (1990), p. 140.

interiore” (*innere Sorgfalt*), intesa come «la chiamata a raccolta dei cinque sensi, il tendere le forze dello spirito, il sollecitare la vigilanza dell’intero apparato psicofisico»¹⁷⁰;

B. il *dovere di evitare*¹⁷¹, quale secondo elemento costitutivo del dovere di diligenza, in quanto il *dovere di riconoscere* (logicamente precedente) è già stato indagato e, con riferimento ad alcuni rischi dei *chatbot therapists*, ritenuto astrattamente esigibile (v. *supra*, § 2.1 e § 2.1.1)¹⁷²;

C. il contenuto standardizzabile del *dovere di evitare*. In tale prospettiva, accanto al versante oggettivo-invariabile – ossia le regole cautelari da osservare – si devono considerare anche gli elementi riferibili all’agente concreto che, ove standardizzabili, concorrono a definire il contenuto del fatto tipico colposo. Secondo una distinzione proposta in dottrina, la tipicità colposa comprende – oltre agli elementi normativi (le regole cautelari doverose) – anche *elementi soggettivi e psicologici*¹⁷³. Gli elementi soggettivi attengono alle capacità, inferiori o superiori, dell’agente reale rispetto all’agente modello e possono assumere rilievo nella misura in cui siano suscettibili di tradursi in uno statuto cautelare specifico (si pensi al cardiocirurgo specializzato nel trattamento di determinate patologie cardiache rispetto al cardiologo “generico”); negli elementi psicologici rilevano aspetti come l’assenza di volontà e l’eventuale coscienza della violazione o la rappresentazione dell’evento¹⁷⁴. Questi secondi non esprimono pertanto una misura standardizzabile del *contenuto* del dovere di diligenza, ma sono più

¹⁷⁰ ENGISCH (1930), p. 271. Trattasi di profili legati alla concreta dimensione soggettiva dell’agente, rilevanti soprattutto nell’accertamento individuale della responsabilità.

¹⁷¹ Secondo l’ormai classica concezione liberale del diritto penale, esso non assolve a una funzione di moralizzazione dei consociati, ma si struttura come sistema normativo volto alla tutela dei beni giuridici attraverso la conformazione preventiva delle condotte dei consociati. Precipitato di tale secolarizzazione è che lo Stato, nella tipizzazione delle norme, deve rendere conoscibile al cittadino ciò che non può fare (art. 5 c.p., nell’interpretazione fornita da Corte cost., sent. 24 marzo 1988, n. 364). In questa traiettoria, il criterio della evitabilità dell’evento non può essere impiegato in termini di possibilità, giacché un evento potrebbe sempre essere evitato mediante l’astensione dalla condotta che lo ha cagionato; l’evitabilità intesa in senso naturalistico quale “possibilità di evitare” attiene, piuttosto, al terreno della *coscienza e volontà* della condotta ai sensi dell’art. 42, comma 1, c.p. – cfr. FORTI (1990), pp. 197-198. Sul terreno della colpa e, in particolare, della tipicità colposa (quale sede di delimitazione dell’illecito) il criterio dell’evitabilità assume, invece, una qualifica normativa e «designa la situazione nella quale l’agente non ha evitato ciò che aveva il *dovere di evitare*» – FORTI (1990), p. 194. Solo successivamente, sul distinto piano della colpevolezza, assumono rilievo le ipotesi in cui l’agente, pur a fronte di un *dovere di evitare* normativamente configurato, non abbia potuto adempiervi per l’assenza, in concreto, delle capacità psicofisiche necessarie a conformare la propria condotta allo standard richiesto.

¹⁷² La logica segue la regola per cui «intanto sussisterà la tipicità colposa, in quanto l’agente avrebbe dovuto riconoscere il fatto e tale riconoscimento avrebbe dovuto indurlo a evitarne la verifica» – FORTI (1990), pp. 236-237. Inoltre, l’ordine seguito è coerente con la ricostruzione secondo cui il *dovere di riconoscere* assume rilievo autonomo precipuamente nell’ambito delle fattispecie colpose, nelle quali il rimprovero si fonda (pur senza esaurirsi) sul difetto di previsione di una situazione di rischio riconoscibile (il soggetto, agendo contro l’intenzione, «deve essere posto comunque in grado di avvertire la “pretesa di adeguamento” implicita nella norma in cui il fatto è previsto» – *ivi*, p. 184). Nelle fattispecie dolose, invece, la lesione del bene giuridico costituisce l’oggetto della rappresentazione e della volontà dell’agente e, pertanto, si colloca interamente sul piano di ciò che egli avrebbe potuto evitare mediante l’astensione dal comportamento intenzionalmente posto in essere; ne consegue che il parametro della riconoscibilità non opera quale criterio selettivo autonomo della tipicità. È, dunque, il secondo momento della diligenza, il *dovere di evitare*, a esprimere un criterio comune tanto al fatto colposo quanto al fatto doloso. *In breve*: si è dapprima ricostruito il contenuto del *dovere di riconoscere* – logicamente anteriore e tipico della colpa, quale presupposto minimo di imputazione soggettiva; si passa ora al successivo profilo dell’evitabilità, che funge da criterio comune sia alla colpa sia al dolo.

¹⁷³ Cfr. CASTRONUOVO (2009), spec. pp. 366 ss. e pp. 562 ss.; DONINI (2013), pp. 127 ss.

¹⁷⁴ Cfr. CASTRONUOVO (2009), spec. pp. 368-369 e pp. 562 ss.

funzionali a distinguere la dimensione colposa da quella dolosa¹⁷⁵; il loro esame resta pertanto esterno al fuoco della presente indagine.

L'attenzione, piuttosto, andrà rivolta agli elementi soggettivi. In particolare, rispetto ai casi in cui il fornitore si avvalga, nello sviluppo del *chatbot therapist*, di esperti in ambito psicologico,¹⁷⁶ occorre interrogarsi se le conoscenze così acquisite incidano sullo statuto cautelare applicabile o rilevino sul piano della colpevolezza.

Ai fini della definizione del contenuto della diligenza, si deve verificare, infine, l'eventuale assunzione di un obbligo di impedimento *per ingerenza* da parte del fornitore, nonché l'ipotesi di una colpa per assunzione derivante dallo svolgimento di un'attività lecita, ma in difetto delle qualifiche necessarie al governo dei rischi (quali quelli di aggravamento della patologia di utenti affetti da disturbi mentali). Un tema quest'ultimo che impone di considerare anche il riconoscimento del rischio da parte della "vittima", quale fattore idoneo, al ricorrere di determinate condizioni, a far venir meno la logica dell'affidamento.

Sempre con riguardo alla persona offesa, occorre successivamente chiarire entro quali limiti il diritto penale possa intervenire rispetto ad atti autolesivi, tenuto conto della sua matrice liberale e dei limiti che essa oppone a ogni opzione paternalistica¹⁷⁷.

Su tali premesse, considerato il discrimine del rischio consentito, lo sviluppo dell'indagine si organizza seguendo una *summa divisio*: da un lato, i *chatbot therapists* come *pratiche vietate* ai sensi dell'art. 5, par. 1, lett. a) e b), AI Act (§ 3 e § 3.1); dall'altro, i *chatbot therapists* come applicazioni consentite dal Regolamento (§ 4 e § 4.1).

Con riferimento al primo ambito, è da stabilire se, nel caso di un'attività *ex ante* ricondotta *in via precauzionale* tra quelle vietate¹⁷⁸, la violazione del dovere di astensione esaurisca l'anti-doverosità della condotta oppure se¹⁷⁹, invece, permanga la necessità di un ulteriore scrutinio sul contenuto della diligenza (che avrebbe dunque spazio e rilevanza autonomi).

Quanto al secondo ambito, è da accertare se le regole cautelari dettate dal Regolamento per i "sistemi di IA ad alto rischio" possano concorrere, pur non essendo direttamente rivolte ai fornitori di *chatbot therapists*, alla definizione del contenuto del dovere di evitare in base al criterio del rischio tipico.

¹⁷⁵ *Ibidem*.

¹⁷⁶ È quanto risulta dichiarato, ad esempio, sulle piattaforme dei *chatbots* offerti da "Therabot" e "Wysa".

¹⁷⁷ In realtà – osserva FORTI (2006, p. 331) – la qualificazione "liberale" risulta più propriamente riferibile alla scienza penale, quale sapere critico del sistema penale vigente, che non al diritto penale in quanto tale, giacché l'espressione "diritto penale liberale" presenta una caratterizzazione «tendenzialmente ossimorica».

¹⁷⁸ Per la ricostruzione secondo cui, nel caso di *chatbot therapists* non progettati per sfruttare deliberatamente le vulnerabilità dell'utente, la logica precauzionale può emergere nella determinazione della soglia di esigibilità del divieto, v. *supra*, § 2.1.3.

¹⁷⁹ Come già visto *supra*, § 2.1.2, nota n. 117, parte della dottrina – tra gli altri, GIUNTA (1993), pp. 223 ss. – differenzia il *dovere di diligenza* dal *dovere di astensione con funzione cautelare*, ritenendoli incompatibili tra loro, sul presupposto che sarebbe contraddittorio vietare lo svolgimento di una determinata attività e, al contempo, prevederne l'esercizio nel rispetto di determinate regole cautelari. Tuttavia, anche aderendo a tale impostazione, resta da chiedersi se la violazione di un dovere di astensione di *matrice precauzionale* sia, di per sé, idonea a esaurire il disvalore penalmente rilevante della condotta.

3. La responsabilità “per il tipo” (di produzione): i *chatbot therapists* come pratiche vietate.

Come già acquisito, l’AI Act organizza la disciplina dell’intelligenza artificiale tracciando una bipartizione di fondo tra pratiche vietate e applicazioni consentite di IA.

In un simile assetto normativo può venire in rilievo il duplice paradigma – elaborato dalla più autorevole dottrina penalistica italiana negli anni Settanta¹⁸⁰ – della responsabilità “per il tipo” e “per il modo” di produzione. Nel primo, il rimprovero penale si ricollega alla scelta di intraprendere un tipo di *attività di base illecita*, ritenendo l’ordinamento il suo esercizio incompatibile con le esigenze di tutela del bene giuridico¹⁸¹; nel secondo, la rilevanza penale della condotta dipende dalle modalità di svolgimento di un’*attività in sé lecita*, benché rischiosa e, dunque, assoggettata al rispetto di determinate cautele.

In questa prospettiva, la qualificazione dei *chatbot therapists* come pratiche vietate ovvero come applicazioni di IA lecite non si esaurisce in un’operazione classificatoria con ricadute solo sul piano dell’intensità degli obblighi: essa incide sul regime di imputazione della responsabilità penale. La riconduzione di questi particolari agenti conversazionali nell’ambito della responsabilità “per il tipo” di produzione sollecita, infatti, il confronto con due snodi problematici. Da un lato, essa pone la questione della configurabilità della colpa nei casi di attività vietata *ab origine*: se la colpa si esaurisca nella scelta di intraprenderla o possa investire anche le modalità del suo svolgimento. Si tratta, cioè, di stabilire se, a fronte della violazione di una norma che vieta in radice un certo tipo di attività, trovi spazio un accertamento sul rispetto di ulteriori cautele durante il suo svolgimento, ovvero se il contenuto del dovere di diligenza si esaurisca nell’astenersi dal compiere detta attività vietata¹⁸². Dall’altro, la riconduzione della

¹⁸⁰ Il riferimento è alla fondamentale ricostruzione offerta da BRICOLA (1978), pp. 75-90.

¹⁸¹ Per una più precisa delimitazione della nozione di “tipo di produzione”, si riporta la definizione offerta BRICOLA (*ivi*), p. 76: «Per *tipo* di produzione intendo la specie di attività produttiva complessivamente considerata: e il problema investe quei settori della produzione che presentano un altissimo grado di pericolosità (non soltanto all’interno, ma anche all’esterno dello stabilimento) difficilmente fronteggiabili mediante *cautele* relative al processo produttivo e soprattutto tali da provocare danni caratterizzati sia da gravità sia da irreversibilità (ossia da impossibilità di ripristinare lo *status quo ante*, preesistente cioè al suo verificarsi). Nel settore del *tipo* di produzione, così inteso, sul momento *ex post* della responsabilità del produttore prevale, com’è ovvio, il momento del controllo preventivo in via amministrativa diretto ad autorizzare od escludere determinati tipi di produzione».

¹⁸² Nel caso in esame non ci si riferisce alla questione della *culpa in re illicita* – riguardante la causazione di un evento non voluto derivante dallo svolgimento di un reato doloso *non* avente funzione cautelare – bensì al tema dell’operatività della colpa nell’ambito di attività vietate sul piano civilistico o amministrativistico. La prima ipotesi – quella della *culpa in re illicita* – è oggetto di un articolato dibattito dottrinale, nel quale si confrontano diverse sensibilità. Per restituire la complessità del tema, si richiamano alcuni dei principali indirizzi: a) coloro che – in una concezione che pare risentire di ascendenze tipiche dell’assolutismo giuridico – ritengono antitetica l’imposizione di un divieto nella fattispecie dolosa rispetto a un imperativo che richieda l’uso della diligenza nello svolgimento dell’attività (cfr. CARMONA, 2001, pp. 237-238; BARTOLI, 2010, p. 1049); b) chi, al contrario, ritiene concepibile un ordinamento che, sul presupposto dell’avvenuto fallimento della prima norma, si preoccupi di richiedere al consociato almeno l’impiego di cautele nell’esecuzione dell’illecito penale (BASILE, 2011, spec. p. 927 e pp. 938-939, il quale richiama anche ANTOLISEI, 1948, p. 8); c) chi esclude l’ipotesi di un rapporto di subordinazione tra l’obbligo di astensione e il dovere di diligenza a contenuto modale, ritenendoli due strumenti alternativi di tutela dei beni giuridici (cfr. GIUNTA, 1993, pp. 198 ss.; in senso contrario, v. CANESTRARI, da ultimo, 2023, pp. 90-91, ove afferma: «Se all’agente a cui si renda [concretamente] prevedibile l’evento ulteriore come conseguenza della propria condotta illecita è possibile muovere un ulteriore rimprovero, significa che egli non solo poteva, ma, in definitiva, doveva comportarsi diversamente»). Sul punto, si accoglie la tesi di quella dottrina – per tutti qui rappresentata da DONINI (1991), p. 65 – secondo la quale una

fattispecie alla responsabilità “per il tipo” si riflette sull’individuazione dei soggetti a cui l’offesa può essere imputata. Infatti, «mentre gli eventi derivanti dalle modalità di produzione possono anche essere il frutto di decisioni individuali, quelli conseguenti alla scelta di un tipo di produzione dipendono sempre da una determinata politica di impresa»¹⁸³.

Muovendo da questo secondo profilo, nel caso dei *chatbot therapists* sembrerebbe, in realtà, che la riconduzione dell’aumento del rischio suicidario dell’utente alle scelte compiute dal fornitore a livello di *governance* derivi dalla stessa natura dei pericoli in questione, prima che dal tipo di attività (illecita o consentita). Nella prima parte di questo scritto si è cercato di evidenziare come il suicidio non si presti a essere ricondotto a singoli fattori causali – come gli specifici *outputs* istigatori del sistema – ma rappresenti, piuttosto, l’esito di un processo sequenziale e temporalmente disteso. In tale arco assumono rilievo vari fattori, tra i quali: il progressivo ritiro dalla socialità, l’allontanamento dalla terapia con sanitari qualificati – che, tra l’altro, costituisce lo strumento principale per monitorare il rischio suicidario –, la perdita di ragioni percepite per vivere (v. *supra*, § 1). Nell’ambito del rapporto con i *chatbot therapists*, alcune scelte di strutturazione del servizio possono amplificare tali dinamiche. Tra queste vi sono l’accessibilità senza vincoli dell’interazione, la presentazione del *chatbot* come strumento di sostegno psicologico più facilmente accessibile, la possibilità di modellare l’agente artificiale su figure emotivamente significative e lo sfruttamento dei meccanismi di omofilia, che – in utenti privi di maturità critica o di identità pienamente formata – possono provocare lo spostamento dell’investimento affettivo verso la relazione con l’agente artificiale, fino alla convinzione che il legame con il *chatbot* possa compiersi in una dimensione altra rispetto a quella reale (v. il caso richiamato *supra*, § 1, alla nota n. 32). Trattasi di conformazioni che non costituiscono meri aspetti tecnici del funzionamento del sistema – su cui potrebbero incidere anche interventi a valle –, ma espressione di scelte che, sia permesso ripetere, attengono alla progettazione del servizio in sé e che, in quanto tali, sono astrattamente di competenza della *governance*. Può allora ritenersi che, rispetto al pericolo per la *vita* dell’utente con disturbi mentali connesso all’uso dei *chatbot therapists*, la distinzione tra pratiche vietate e applicazioni lecite introdotta dall’AI Act non appare incidere sull’individuazione dei centri di imputazione della responsabilità. Le scelte progettuali pericolose sono, almeno in astratto, da

contraddizione ordinamentale si avrebbe solo qualora la regola cautelare abbia a oggetto la medesima condotta costitutiva del fatto tipico di reato (come nell’omicidio preterintenzionale). Venendo, invece, all’ipotesi relativa alla colpa nell’esercizio di *attività illecite* sul piano *extra-penalistico* – più pertinente al caso in esame – appare preferibile ammettere uno spazio autonomo del dovere di diligenza anche nei casi di violazione del divieto. Ciò in considerazione del fatto che la mera trasgressione della norma extra-penale, se sufficiente a integrare un contrasto con l’ordinamento che giustifica sanzioni amministrative (nel caso di pratiche di IA illecite, anche di importo particolarmente elevato – v. art. 99, par. 4, AI Act), non può esaurire, di per sé, il disvalore penale della condotta. Esso, infatti, oltre a richiedere quantomeno l’accertamento della prevedibilità ed evitabilità dell’evento (che, nel caso di specie, possono dirsi sussistenti), si colloca altresì nel quadro della struttura frammentaria e garantistica del diritto penale, nell’ambito della quale la normativa extra-penale può sì contribuire a determinare il contenuto della regola cautelare, ma non costituire, di per sé, ove si limiti a porre un divieto, una violazione tale da giustificare un rimprovero penale – in questo senso, v. anche la tesi di dottorato di ANDREAS (2016), *passim*. La medesima conclusione vale, e anzi risulta rafforzata, qualora il divieto sia interpretato in ottica precauzionale (sul punto, v. *infra*, § 3.1, con riferimento a un’ipotesi riguardante la riconduzione dei *chatbot therapists* alle pratiche vietate ex art. 5, par. 1, lett. a) e b), AI Act).

¹⁸³ BRICOLA (1978), p. 84.

ricondurre agli organi di vertice, la cui competenza decisionale resta la medesima indipendentemente dalla liceità o dall'illiceità dell'attività d'impresa.

Una ricostruzione parzialmente diversa sembra doversi compiere con riferimento non alla dinamica dell'atto suicidario, ma all'aggravamento del disturbo mentale – o, per riprendere la terminologia adottata dal codice penale, «malattia nella mente»¹⁸⁴ – nei casi di *chatbots* concepiti e pubblicizzati per finalità di sostegno psicologico (le prime due categorie che si sono individuate *supra*, § 1)¹⁸⁵. In tali ipotesi, strategie di orientamento psicoterapeutico contrarie ai protocolli validati dalla comunità scientifica, se non sono di regola sufficienti, isolatamente considerate, ad aumentare in modo sensibile il rischio di suicidio, possono assumere autonoma rilevanza rispetto a un pericolo di aggravamento della patologia psichica¹⁸⁶. Verosimilmente, è rispetto a questi ultimi casi che si pone maggiormente il problema dell'individuazione del livello decisionale cui imputare le indicazioni a carattere terapeutico fornite dagli *outputs* del sistema. In particolare, si dovrebbe stabilire se l'inidoneità dei contenuti: 1) sia riconducibile a un errore “a monte”, inerente al modello di intervento psicologico prescelto e incorporato nel servizio per *delibera* della governance – eventualmente sulla base di linee guida o protocolli definiti con il contributo di professionisti; 2) sia ricollegabile a condotte di soggetti “a valle”, relative a interventi tecnici di implementazione (addestramento, *fine-tuning*, sistemi di *safety*, *prompt* etc.); 3) sia dipesa dall'autonomia generativa del sistema nelle singole interazioni.

In relazione a tali ipotesi, la distinzione tracciata dall'AI Act tra pratiche vietate e sistemi ad alto rischio può assumere un rilievo dirimente. In particolare, qualora l'impiego della terapia offerta mediante *chatbot* faccia emergere effetti di distorsione del

¹⁸⁴ Con riferimento alla nozione di “malattia mentale” rilevante ai fini della responsabilità penale, giova evidenziare che secondo il costante orientamento della Corte di Cassazione (*ex multis*: Cass. Pen., Sez. V, 29 settembre 2010, n. 43763; Cass. Pen., Sez. V, 25 ottobre 2012, n. 8351), in tema di lesioni personali, «costituisce “malattia” qualsiasi alterazione anatomica o funzionale dell'organismo, ancorché localizzata, destinata a perdurare fino a quando sia in atto il suddetto processo di alterazione. Essa può riguardare sia la sfera fisica della persona, sia quella psichica. Il concetto di lesione coinvolgente la sfera psichica della persona ha trovato luogo nella giurisprudenza di legittimità con riferimento all'art. 582 c.p., dove è espressamente richiamato il concetto di malattia «nella mente». Si afferma, secondo la definizione tradizionalmente fornita dalla giurisprudenza della Corte di legittimità che la malattia nella mente è quella che comporta non soltanto offuscamento o disordine, ma anche indebolimento, eccitamento, depressione o inerzia dell'attività psichica, con effetto permanente o temporaneo» – Cass. Pen. Sez. IV, 11 gennaio 2018, n. 822. Il principale ostacolo applicativo nei casi di lesioni consistenti nell'insorgenza o nell'aggravamento di una malattia psichica è rappresentato dall'accertamento dell'evento e del relativo nesso causale. Anche qualora sia dimostrata la scorrettezza delle pratiche terapeutiche adottate, la responsabilità penale non può essere affermata in assenza della prova che da tali condotte sia derivata una “malattia nella mente” penalmente rilevante. La giurisprudenza di legittimità ha più volte evidenziato tale difficoltà probatoria, escludendo la responsabilità dello psicoterapeuta quando non risulti dimostrato che la patologia lamentata dal paziente sia stata effettivamente indotta o aggravata dal trattamento terapeutico scorretto (cfr., *ex multis*, Cass. Pen., Sez. IV, 11 gennaio 2018, n. 822). Trattasi di difficoltà che, nell'ambito dei *chatbots therapists* possono risultare ridimensionate dinanzi all'introduzione, in materia di IA, di fattispecie di pericolo a tutela dell'incolumità individuale (v. art. 24, comma 5, lett. b, legge n. 132 del 2025 e l'art. 437-bis c.p.).

¹⁸⁵ Si fa riferimento ai *chatbots* progettati e pubblicizzati, nel primo caso, quali strumenti di supporto terapeutico psicologico non clinico e, nel secondo, quali strumenti destinati al benessere emotivo, anche se integrati in piattaforme generaliste di intelligenza artificiale generativa.

¹⁸⁶ Si pensi, ad esempio, a interazioni che inducano l'utente a far emergere ricordi traumatici, emozioni dolorose o conflitti irrisolti precedentemente evitati, senza che tale riattivazione sia accompagnata da adeguate strategie di contenimento e di elaborazione – cfr. BARLOW (2010); oppure alla definizione di obiettivi di miglioramento rigidamente formulati e collocati su orizzonti temporali irrealistici, cui faccia seguito, in caso di mancato raggiungimento, una restituzione implicita o esplicita di responsabilità all'utente. Dinamiche che potrebbero contribuire a un peggioramento del quadro psicopatologico, pur senza comportare un immediato incremento del rischio suicidario.

comportamento tali da provocare, o da poter ragionevolmente provocare, «un danno significativo» (anche non voluto dal fornitore)¹⁸⁷, l'attività ricadrebbe nell'ambito delle pratiche vietate ai sensi dell'art. 5, par. 1, lett. a) o, quantomeno, b), del Regolamento e, conseguentemente, dovrebbe essere interrotta.

In simile circostanza, la condotta offensiva coinciderebbe con la scelta di mantenere attivo il servizio, la quale, in via di principio, va ricondotta alla *governance*. Non appare, d'altra parte, decisivo richiamare l'assenza di una conoscenza effettiva del rischio da parte dell'organo di vertice, giacché la disposizione richiamata, facendo riferimento agli «effetti [...] tali da provocare, o da poter ragionevolmente provocare, un danno significativo per la sua salute» contempla una forma di *dovere di riconoscere* riguardante i rischi del *chatbot*¹⁸⁸. Se si condivide tale impostazione, allora deve concludersi che sul fornitore grava un obbligo di diligenza il cui contenuto può estendersi alla predisposizione di strumenti idonei a intercettare tempestivamente gli effetti pericolosi derivanti dall'impiego dello specifico sistema. L'eventuale incapacità tecnica di assolvere a tale dovere dovrebbe condurre alla scelta di astenersi dall'offrire il servizio. In merito a questo specifico punto, peraltro, più che di impossibilità, sembra ragionevole parlare di difficoltà di ordine tecnologico o organizzativo. Da un lato, nella prassi applicativa risultano già adottati strumenti tecnici – quali sistemi di moderazione e filtri di sicurezza – volti a rilevare la produzione di contenuti potenzialmente pregiudizievoli per la salute o l'equilibrio psichico dell'utente¹⁸⁹. Dall'altro, nell'ambito dell'art. 5, par. 1, lett. a) e b), AI Act non rientra la mera variazione dell'atteggiamento di singoli utenti, che presupporrebbe un arduo monitoraggio di ogni messaggio dell'utente (peraltro non necessariamente sufficiente a coglierne un significato sul piano clinico). Del resto, i *chatbots* in esame sono per definizione orientati a incidere sul comportamento dell'utente, indirizzandolo verso specifiche modalità di gestione del proprio stato emotivo mediante tecniche terapeutiche¹⁹⁰. Ciò che pare rilevare ai fini dell'art. 5, par. 1, lett. a) e b) del Regolamento è che tale incidenza si traduca in una *distorsione* – termine che esprime una deviazione in senso negativo – del comportamento, idonea a provocare, o a poter ragionevolmente provocare, un danno significativo. Dunque, rimanendo nel perimetro dell'interazione utente-*chatbot*, affinché la variazione del modo di comportarsi

¹⁸⁷ I principali tipi di danni pertinenti ai fini dell'art. 5, paragrafo 1, lett. a) e b), AI Act sono i danni fisici, psicologici, finanziari ed economici, cui in alcuni casi possono associarsi danni sociali più ampi, come si afferma al considerando n. 29. In tale enunciato preliminare si precisa, altresì, che «non è necessario che il fornitore o il deployer abbiano l'intento di provocare un danno significativo, purché tale danno derivi da pratiche di IA manipolative o di sfruttamento consentite dall'IA».

¹⁸⁸ Sul concetto di *dovere di riconoscere*, v. *supra*, § 2.1.

¹⁸⁹ V. *infra*, § 4.

¹⁹⁰ Lo riconoscono le stesse Linee Guida della Commissione UE sulle pratiche vietate. Tali orientamenti, pur non avendo carattere vincolante – essendo solo la Corte di giustizia dell'Unione europea competente a fornire, in ultima istanza, un'interpretazione vincolante degli atti dell'UE – costituiscono, allo stato, un rilevante riferimento interpretativo. In particolare, a pagina 50, nel par. 3.5.2., rubricato «Sistemi di IA manipolativi, ingannevoli e di sfruttamento che *non* possono provocare un danno significativo» (corsivo di chi scrive), è riportato il seguente esempio: «un chatbot terapeutico che utilizza tecniche subliminali per orientare gli utenti verso uno stile di vita più sano e far loro abbandonare cattive abitudini, come il fumo. Anche se gli utilizzatori che seguono i consigli e la terapia subliminale del chatbot subiscono qualche disagio fisico e stress psicologico a causa degli sforzi compiuti per smettere di fumare, non si può ritenere che il chatbot basato sull'IA possa provocare un danno significativo. Si tratta di un disagio temporaneo inevitabile e controbilanciato dai benefici a lungo termine per la salute degli utenti».

possa dirsi distorsiva sembra sufficiente che il sistema metta in pratica un'impostazione terapeutica clinicamente inadeguata, idonea a orientare l'utente verso pratiche disfunzionali o controindicate. In questa prospettiva, si conferma che dall'art. 5, par. 1, lett. a) e b), AI Act sorge per i fornitori di *chatbot therapists* un dovere che non implica un controllo di ogni singola interazione dell'utente – il quale potrebbe assumere rilievo in caso di ipotetica, attualmente da escludere, posizione di protezione (v. *infra*, § 5) – bensì una verifica degli *schemi ricorrenti* di intervento terapeutico messi in pratica dal *chatbot*¹⁹¹. Pertanto, anche qualora fosse accertato che l'inadeguatezza terapeutica degli *outputs* trovi origine nelle condotte dei programmatori "a valle" – quali la selezione dei dati di addestramento, le modalità di *fine-tuning* o la configurazione dei sistemi di sicurezza – ovvero nell'autonomia generativa del sistema, la condotta offensiva andrebbe ravvisata nella scelta, da parte della *governance*, di mantenere in servizio il *chatbot* e in violazione del divieto¹⁹².

In merito, giova altresì rilevare che la disposizione testé richiamata non è costruita secondo una sintassi che descrive come già in atto l'effetto distorsivo (del tipo: sistemi che sfruttano le vulnerabilità «e che distorcono» / «distorcendo» il comportamento); essa assume la distorsione come obiettivo o *effetto* dello sfruttamento delle vulnerabilità¹⁹³. L'attenzione verso la formulazione della norma non risponde a uno sterile intento esegetico: mentre la prima espressione presupporrebbe, ai fini dell'operatività del divieto, l'accertamento dell'avvenuta distorsione del comportamento degli utenti, la scelta normativa valorizza l'idoneità concreta del sistema a produrre tale effetto. In altri termini, il tenore lessicale dell'art. 5, par. 1, lett. a) e b), AI Act consente di affermare che il legislatore europeo non richiede, per le pratiche vietate, l'accertamento di una distorsione già verificatasi, ma considera sufficiente che il sistema, per come configurato, abbia l'effetto di orientare il comportamento in modo distorsivo¹⁹⁴.

¹⁹¹ La questione se l'art. 5, par. 1, lett. a) e b), AI Act fondi una posizione di controllo ovvero contribuisca a definire il contenuto del dovere di diligenza sarà affrontata nel paragrafo successivo.

¹⁹² In sintesi, l'imputazione alla *governance* si fonda su presupposti diversi a seconda dell'offesa considerata: nel caso di aggravamento del rischio suicidario, essa discende dalle scelte di configurazione del servizio (assenza di limiti di impiego, sfruttamento dei meccanismi di omofilia, etc.), che strutturano la relazione *chatbot*-utente; nel caso dell'aggravamento della patologia psichica, invece, rileva la decisione – anch'essa riconducibile all'organo di vertice – di mantenere attivo il servizio nonostante la rilevanza di effetti distorsivi del comportamento, in violazione del divieto posto dall'art. 5 AI Act.

¹⁹³ Si riporta, per la parte qui rilevante, lo stralcio dell'art. 5, par. 1, lett. a) e b), AI Act: «sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone» (lett. a); «sistema di IA che sfrutta le vulnerabilità di una persona fisica o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo» (lett. b).

¹⁹⁴ Nello stesso senso, sempre le Linee Guida della Commissione UE sulle pratiche vietate chiariscono che «In linea con la normativa in materia di protezione dei consumatori, non è necessario che tali effetti si siano pienamente concretizzati, ma devono esservi sufficienti indicazioni del fatto che essi sono probabili o in grado di concretizzarsi e che pregiudicano l'autonomia individuale, sulla base di una valutazione obiettiva di tutte le circostanze del caso e delle conoscenze e dei metodi scientifici esistenti, nonché delle informazioni disponibili sull'impatto del sistema sul comportamento delle persone nella vita reale. A tal riguardo, il fatto che un sistema sia in grado di indurre comportamenti che pregiudicano in modo considerevole la capacità delle persone di prendere una decisione informata e che ne compromettono le libere scelte è sufficiente a soddisfare tale condizione e non dipende da considerazioni relative ai "tempi" necessari affinché il danno si concretizzi (ad esempio in caso di comportamenti assimilabili alla dipendenza), purché sia ragionevolmente possibile che esso si verifichi. Ad esempio, un chatbot sul benessere basato sull'IA è destinato dal fornitore ad aiutare e orientare gli utenti nel mantenimento di uno stile di vita sano e a fornire

In tale cornice, l'eventuale responsabilità di soggetti "a valle" (sviluppatori, addetti al *fine-tuning*, etc.) potrà configurarsi – nei casi in cui il sistema non faccia emergere effetti distorsivi tali da determinarne la riconduzione alle pratiche vietate – in relazione ai meccanismi tecnici di monitoraggio e verifica delle interazioni *chatbot*-utente. In particolare, essa potrà sorgere qualora, rispetto allo stato delle conoscenze tecnico-scientifiche e alla prassi applicativa del settore, essi risultino erroneamente configurati o malfunzionanti, impedendo l'emersione o la corretta rappresentazione alla *governance* di criticità rilevanti.

La ricostruzione proposta con riferimento agli operatori non appartenenti al livello di gestione della *governance* si colloca entro le coordinate della responsabilità "per il modo" di produzione: paradigma nel quale, riprendendo le parole di Bricola, «gli eventi derivanti dalle modalità di produzione possono anche essere il frutto di decisioni individuali». L'attività, infatti, finché non risulta che la strutturazione del sistema abbia l'effetto di distorcere il comportamento degli utenti, va qualificata come lecita. Una responsabilità "per il tipo" in capo a operatori non appartenenti alla *governance* potrebbe ipoteticamente configurarsi solo a titolo di concorso doloso, qualora essi contribuiscano consapevolmente alla prosecuzione del servizio nonostante siano emersi i suoi effetti distorsivi sul comportamento degli utenti.

3.1. L'imputazione soggettiva: tra dolo eventuale e principio di precauzione.

Alla conclusione raggiunta è necessario apportare un chiarimento che, in linea di massima, consentirebbe di sciogliere anche la prima questione evocata in tema di responsabilità "per il tipo" (con riferimento ai casi in cui l'impostazione "relazionale" o terapeutica del servizio ne comporti la riconduzione tra le pratiche vietate e, quindi, alle decisioni degli organi di vertice). È quella della configurabilità della colpa penale nello svolgimento di attività vietate dal diritto civile o amministrativo.

Qui il punto sembra risolversi osservando che, nel caso in cui l'agire del *chatbot therapist* faccia emergere schemi di interazione potenzialmente idonei a provocare un danno alla salute degli utenti, il contributo dei profili "a valle" ai fini della prosecuzione del servizio – decisa "a monte" dalla *governance* – potrebbe integrare un concorso doloso in un reato *doloso* dei vertici. Invero, in simili ipotesi la condotta degli organi apicali non solo integrerebbe una deviazione deliberata dallo standard di diligenza che imporrebbe

consulenza personalizzata per esercizi psicologici e fisici. Se però il chatbot sfrutta le vulnerabilità individuali *per indurre ad adottare abitudini poco sane o praticare attività pericolose* (ad esempio praticare troppa attività sportiva senza riposarsi o bere) e si può ragionevolmente prevedere che determinati utenti seguiranno tali consigli, comportandosi diversamente da quanto avrebbero altrimenti fatto, e subiranno danni significativi (ad esempio un attacco cardiaco o altri gravi problemi di salute), tale sistema di IA rientrerebbe nel divieto di cui all'articolo 5, paragrafo 1, lettera a), [nonché, è doveroso aggiungere, nel divieto di cui alla lettera b)] del regolamento sull'IA, anche se tale comportamento e le conseguenze dannose per le persone potrebbero non essere state intenzionali da parte del fornitore. Il semplice fatto che il chatbot sia in grado di pregiudicare in modo considerevole l'autonomia individuale e di distorcere materialmente il comportamento di determinati utenti in modo significativamente dannoso e che il fornitore non abbia adottato misure preventive e di attenuazione adeguate a evitare tali effetti significativamente dannosi è sufficiente per far scattare l'applicazione del divieto» (corsivo di chi scrive).

l'astensione – aspetto rilevante ai soli fini del giudizio normativo-rappresentativo sul dolo, non ancora sufficiente sul piano dell'elemento psicologico della volontarietà –, ma potrebbe far ravvisare un'adesione alla verifica dell'evento lesivo, secondo lo schema del dolo eventuale. La questione dell'apprezzamento della colpa in attività vietata risulta, così, assorbita.

Esiti in parte diversi potrebbero tuttavia prospettarsi ove il divieto fosse interpretato in chiave precauzionale. In tale logica, ai fini dell'applicazione dell'art. 5 AI Act, si potrebbe ritenere necessaria non già l'emersione di segnali di disfunzionamento del *modello* o del *sistema* idonei a produrre effetti distorsivi sul comportamento degli utenti e a provocare danni per la salute, bensì la sola plausibilità che conseguenze di tal genere possano verificarsi¹⁹⁵. Ciò potrebbe avvenire, ad esempio, qualora si ritenesse applicabile il divieto in parola dinanzi a dati di utilizzo che mostrino un aumento della durata delle sessioni di interazione con il *chatbot* (quali il passaggio da un ordine di decine di minuti fino a due ore al giorno), rispetto alle quali la letteratura segnala sì rischi di dipendenza e di isolamento sociale, ma senza individuare la precisa soglia temporale a partire dalla quale tali effetti tendono a manifestarsi.

Se si condivide la tesi, largamente sostenuta in dottrina (v. *supra*, § 2.1.3), secondo cui il principio di precauzione non può fungere da fondamento di responsabilità penale, ne consegue che la violazione del divieto previsto dall'art. 5, par. 1, lett. a) e b), AI Act – ove ritenuto esigibile in ottica precauzionale e, dunque, in assenza di effetti distorsivi del comportamento riconoscibili – non varrebbe, da sola, a giustificare un rimprovero colposo. A fortiori, essa non potrebbe considerarsi sufficiente per l'integrazione del dolo: a venir meno sarebbe – ad avviso di chi scrive – la possibilità di affermare il momento normativo-rappresentativo dell'elemento soggettivo.

In altri termini, ove il divieto di cui all'art. 5, par. 1, lett. a) e b), AI Act venisse inteso in chiave precauzionale, ai fini dell'accertamento della *colpa* del fornitore sarebbe necessario dimostrare la violazione di altre regole a carattere *preventivo*, dirette a governare i rischi tipici connessi all'impiego di tali sistemi, non potendosi invece considerare sufficiente la sola violazione del divieto in parola. Diversamente, si finirebbe per sanzionare sul piano penalistico la mera disobbedienza a un divieto esplicito: opzione censurabile già sotto il profilo dell'offensività.

¹⁹⁵ Quando si ricorre al concetto di “modello” occorre precisare che esso non è univoco nella letteratura tecnico-scientifica. In questa sede si adotta la definizione proposta dall'OECD (2024), p. 8, secondo cui un modello di IA costituisce «una componente centrale di un sistema di IA utilizzata per effettuare inferenze a partire da *inputs* al fine di produrre output» (traduzione di chi scrive). In senso più generale, lo standard ISO/IEC 22989, p. 4, definisce un modello come una «*rappresentazione fisica, matematica o altrimenti logica di un sistema, entità, fenomeno, processo o dato*». In ambito di intelligenza artificiale, il modello costituisce dunque la componente del sistema di IA che trasforma i dati in ingresso in un risultato (ad esempio una previsione o una classificazione) sulla base di una determinata struttura matematica e dei parametri determinati durante la fase di addestramento. Tali strutture possono assumere forme diverse – alberi decisionali, reti neurali o altri modelli statistici – che rappresentano modi differenti di elaborare le informazioni e individuare schemi nei dati. Secondo l'OECD, i modelli di IA possono essere costruiti manualmente da programmatori umani oppure generati automaticamente tramite tecniche di apprendimento automatico, come l'apprendimento supervisionato, non supervisionato o per rinforzo. Quest'ultimo aspetto contribuisce a spiegare perché risultati difficile prevedere con precisione la futura capacità dei sistemi di IA: in alcuni casi, infatti, allo sviluppo delle loro prestazioni contribuisce l'interazione tra gli stessi modelli di IA.

Prima di concludere sulle pratiche vietate, resta da chiarire un ultimo profilo, sinora differito per ragioni espositive e che ora può essere più agevolmente inquadrato. Si tratta di stabilire se l'obbligo di astensione desumibile dall'art. 5, par. 1, lett. a) e b), AI Act configuri una posizione di controllo sulla fonte di pericolo rappresentata dal sistema – potendo divenire esigibile, sul piano penalistico, solo ove sia dato al fornitore conoscere gli effetti distorsivi del *chatbot*, che egli è pertanto tenuto a riconoscere (v. *supra*, § 3) – ovvero se esso si esaurisca sul piano del dovere di diligenza, prevedendo una regola volta a evitare il rischio mediante l'astensione. La seconda soluzione appare preferibile. La prescrizione in parola, infatti, non esalta il disvalore di evento (come le norme che prevedono obblighi di impedimento), bensì un disvalore di condotta¹⁹⁶, consistente nella messa in commercio di sistemi che mostrino l'effetto di distorcere il comportamento degli utenti, anche solo in modo da poter ragionevolmente provocare loro un danno significativo, senza procedere al loro ritiro dal mercato.

4. La responsabilità “per il modo”: i *chatbot therapists* come applicazioni consentite.

Qualora il servizio reso dal *chatbot therapist* non integri le pratiche vietate previste dall'art. 5, par. 1, AI Act, esso può essere lecitamente messo a disposizione degli utenti a condizione che siano osservate le cautele necessarie a impedire o ridurre i rischi tipici connessi al suo impiego.

Già nel solco di questa alternativa – *impedire o contenere* – sorge una questione classica: una volta affermata la rilevanza dell'evitabilità quale *dovere di evitare* delimitato in base ai rischi che si ha il *dovere di riconoscere*, occorre stabilire il parametro per definirne il contenuto.

Non potendo dar conto di tutte le diverse ricostruzioni dottrinali sul punto, si preferisce chiarire le opzioni teoriche cui si farà riferimento.

Anzitutto, si aderisce all'orientamento che individua la regola cautelare doverosa in quella che, nel caso concreto, avrebbe *ridotto le probabilità* di verificazione dell'evento secondo certi nessi eziologici, senza la necessità di accertarne l'efficacia impeditiva¹⁹⁷.

¹⁹⁶ Chiarisce BRUNELLI (2023), p. 114: «Il “momento” giuridico della ricerca della posizione di garanzia esalta il ruolo del disvalore di evento; quello della ricerca del dovere comportamentale esalta il ruolo del disvalore di condotta».

¹⁹⁷ Cfr., per tutti, DONINI (2006), spec. p. 123, ove il chiaro Autore, richiamando ROXIN (1962), p. 171, afferma: «Non è essenziale sapere che cosa sarebbe esattamente successo, a parte il caso della sicura inevitabilità della lesione e dunque dell'impossibilità della cautela di sortire effetti: perché il senso della cautela doverosa è che fosse utile, là dove si sappia che la condotta colposa ha cagionato la lesione e la negligenza ha peggiorato in modo decisivo ($\geq 50\%$), la situazione del bene. Una soluzione di questo tipo è molto più rigorosa di quella che la nostra dottrina tradizionalmente esige, e si armonizza ben anche con le regole cautelari improprie, cioè con le cautele che non mirano a eliminare del tutto un rischio, ma solo a ridurlo (rimanendo una quota di rischio “consentito” ineliminabile), sì che il comportamento alternativo lecito non potrebbe consistere necessariamente (secondo la *ratio* cautelare) in una certezza di escludere l'evento, perché non si ipotizza affatto un obbligo (o un onere) di azzerare il rischio di esso». La distinzione tra *regole cautelari proprie e improprie*, introdotta in Italia da VENEZIANI (2003), pp. 221 ss., sarà ripresa nel prossimo paragrafo in merito all'analisi delle regole cautelari dell'AI Act. In senso più restrittivo, ritenendo rilevante soltanto l'ipotesi in cui l'osservanza della regola cautelare avrebbe evitato l'evento con grado prossimo alla certezza, v. invece MANTOVANI (2013), pp. 347-348.

In secondo luogo, si accoglie la soluzione – oggi maggiormente condivisa in dottrina¹⁹⁸ – secondo cui il parametro ai fini dell’evitabilità va individuato nell’*homo eiusdem professionis et condicionis*¹⁹⁹, inteso come uomo coscienzioso e avveduto appartenente al medesimo ambito professionale e relazionale dell’agente concreto. Beninteso, il riferimento non coincide con la prassi di settore (secondo la logica del “così fan tutti”). Come osserva Marinucci, «la diligenza esigibile si ricava non da quel che si usa fare nello svolgimento di questa o di quella attività, bensì da quel che si poteva pretendere anche in termini di sopportazione dei costi economici, fino alla rinuncia dell’attività dall’*homo eiusdem professionis et condicionis*»²⁰⁰.

Muovendo da queste premesse, la ricostruzione del contenuto della diligenza esigibile dal fornitore deve tener conto che l’esperienza attuale non ha ancora prodotto *best practices* di riferimento. Solo di recente, ad esempio, alcune piattaforme hanno introdotto limiti relativi all’età degli utenti o alla durata dell’utilizzo dei sistemi, nel tentativo di contenere i rischi connessi a forme di attaccamento verso il *chatbot*, a possibili dinamiche di isolamento sociale e di disincentivo al ricorso verso interventi sanitari²⁰¹.

In questo contesto di indagine non ancora assestato, indicazioni utili per la ricostruzione delle cautele esigibili possono comunque trarsi dal caso *Replika* (2025), oggetto di un procedimento istruttorio e di successivi provvedimenti adottati dal Garante per la protezione dei dati personali. Pur collocandosi in un ambito diverso da quello penalistico, tale vicenda consente di osservare, in concreto, tanto il livello delle cautele adottate dai fornitori di *chatbots* per il supporto psicologico, quanto quello delle regole che l’Autorità di controllo ha ritenuto necessarie ai fini della gestione dei rischi connessi al loro uso²⁰².

¹⁹⁸ Cfr., tra i più, MARINUCCI (1965), pp. 154-155; ROMANO (2004), p. 427; PULITANO (2005), p. 388. Propende, invece, per la tesi della miglior scienza ed esperienza come parametro dell’evitabilità, MANTOVANI (2013), pp. 342-343.

¹⁹⁹ Per una critica al criterio dell’*homo eiusdem professionis et condicionis*, ritenuto eccessivamente distante dalla figura dell’agente concreto e, pertanto, afflitto da profili di indeterminazione applicativa, v. GIUNTA (1999), p. 1296; GROTO (2012), pp. 50-51.

²⁰⁰ MARINUCCI (2005), p. 47. Tale impostazione, come rileva CASTRONUOVO (2009), p. 296, è oggi seguita dalla dottrina dominante, alla quale ha fatto seguito la prevalente giurisprudenza. Per un riferimento più aggiornato, v. anche CASTRONUOVO (2021), pp. 200 ss.

²⁰¹ Si vedano, a titolo esemplificativo, le misure progressivamente introdotte dalle piattaforme di “Character.AI” e “Replika” (Luka Inc.). In particolare, “Character.AI” prevede requisiti minimi di età (16 anni nell’UE) e ha introdotto limitazioni alla durata delle conversazioni per gli utenti minorenni (nel momento in cui si scrive, pari a due ore al giorno); analogamente, “Replika” è destinata esclusivamente a utenti maggiorenni (18+). Diversamente, altri *chatbot therapists* – come “Tess” (X2AI) o “Wysa” – riconducibili alla prima delle tre tipologie individuate *supra*, § 1, ossia ai sistemi concepiti per fornire assistenza psicologica di tipo non clinico in senso stretto e presentati espressamente come tali, allo stato non risultano aver introdotto limitazioni alla durata dell’utilizzo. Quanto a “Wysa”, il servizio prevede solo requisiti di età per l’accesso (non inferiore ai 13 anni, con il consenso di un genitore o tutore richiesto per gli utenti minorenni).

²⁰² Sul punto v. [GARANTE PER LA PROTEZIONE DEI DATI PERSONALI, provv. 10 aprile 2025, n. 10130115](#), relativo al servizio di *chatbot* “Replika”. Pur intervenendo nell’esercizio dei poteri attribuiti alle autorità di controllo dal Regolamento generale sulla protezione dei dati, l’atto evidenzia i rischi che l’utilizzo del sistema può comportare per gli utenti, in particolare per i minori: si rileva infatti che la società non aveva adottato «alcuna misura tecnica ed organizzativa atta a garantire il rispetto dei principi generali del Regolamento e la protezione dei diritti e delle libertà dei minori, esponendo in tal guisa i soggetti minorenni ai rischi significativi per la loro persona che la normativa in argomento mira a limitare, tra cui risposte inadeguate rispetto al loro grado di sviluppo psicofisico e di autoconsapevolezza». Nella medesima direzione si colloca anche il comunicato stampa dell’Autorità del 3 febbraio 2023, intitolato: «Intelligenza artificiale, dal Garante privacy stop al *chatbot* ‘Replika’». Troppi i rischi per i minori e le persone emotivamente fragili». Il procedimento risulta concluso con l’adozione di un successivo provvedimento sanzionatorio nei confronti della società: v. il comunicato stampa del [GARANTE PER LA PROTEZIONE DEI DATI PERSONALI](#).

Da alcuni atti del procedimento è possibile ricavare, in particolare, tre ambiti di cautele rilevanti ai fini del discorso.

Un primo gruppo attiene alle misure che la società Luka Inc., fornitore del *chatbot Replika*, ha dichiarato di avere integrato nella progettazione del servizio già prima dell'istruttoria. Tra queste: la limitazione dei dati raccolti in fase di registrazione (nome, indirizzo e-mail, data di nascita ed eventuali dati di log-in di terze parti); l'adozione di procedure di conservazione e cancellazione dei dati; la mancata condivisione delle conversazioni degli utenti con soggetti terzi diversi dai fornitori di servizi essenziali della società, vincolati da obblighi di riservatezza. Trattasi, a ben vedere, di cautele che si collocano prevalentemente sul piano della protezione della privacy e della riservatezza degli utenti, senza incidere direttamente sulla tutela dell'incolumità degli utenti.

In merito a quest'ultimo aspetto assumono rilievo le misure che il Garante ha ritenuto necessarie ai fini della riapertura del servizio all'utenza italiana. In particolare, l'Autorità ha richiesto l'introduzione di meccanismi idonei a impedire l'accesso al servizio da parte dei minori, tra cui: un sistema di age gate nelle pagine di registrazione; la previsione di un periodo di raffreddamento (c.d. *cooling-off period*) volto a evitare l'inserimento reiterato di date di nascita diverse da parte di utenti minorenni; la predisposizione di strumenti di segnalazione dei contenuti inappropriati.

Vengono in considerazione, infine, le iniziative che la società ha dichiarato di aver avviato nel corso del procedimento. In particolare, Luka Inc. ha comunicato di essersi attivata per dare seguito alle richieste formulate dal Garante – anche mediante il coinvolgimento di consulenti ed esperti esterni – e di aver avviato una serie di valutazioni, azioni e processi volti, tra l'altro, all'implementazione di algoritmi e processi adeguati ai fini della moderazione di contenuti inappropriati, secondo il migliore stato dell'arte.

Ulteriori indicazioni, in ordine alle cautele rintracciabili, emergono dall'esperienza applicativa più recente. In particolare, diversi fornitori di *chatbot therapists* – soprattutto riconducibili alle prime due tipologie individuate *supra* (v. *supra*, § 1)²⁰³ – nella presentazione dei propri servizi evidenziano il coinvolgimento di esperti in ambito psicologico o psichiatrico sia nella fase di progettazione sia in quella di funzionamento degli agenti conversazionali. Una circostanza che non è da ritenere rilevante ai fini dell'attribuzione al fornitore di competenze eccedenti quelle dell'agente modello e, come tale, inidonea a incidere sul contenuto della diligenza esigibile. Piuttosto, il ricorso a tali competenze specialistiche va considerato come una caratteristica propria dell'*homo eiusdem professionis et condicionis* operante nel settore. I rischi connessi all'uso di *chatbot therapists* – quali l'instaurazione di dinamiche di attaccamento verso il sistema, l'isolamento sociale e il disincentivo al ricorso a interventi sanitari – devono ritenersi oggetto del *dovere di riconoscere* gravante sui fornitori (v. *supra*,

[19 maggio 2025, 10132048](#). I provvedimenti citati sono ora reperibili sul sito www.garanteprivacy.it.

²⁰³ In particolare: (i) i *chatbots* esplicitamente destinati a forme di supporto psicologico non clinico; (ii) gli agenti conversazionali orientati al benessere emotivo, anche quando integrati in piattaforme generaliste di intelligenza artificiale generativa.

§ 2.1.1). Pertanto, il coinvolgimento, nella fase di progettazione e gestione dei sistemi, di figure dotate di qualificate competenze psicologiche o psichiatriche deve ritenersi parte integrante del contenuto del *dovere di evitare*²⁰⁴ – nella misura in cui ciò sia funzionale a prevenire o ridurre tali rischi e non, certamente, a saper sfruttare le vulnerabilità degli utenti. Ciò vale non soltanto per i *chatbots* espressamente progettati per finalità di supporto psicologico o di benessere emotivo, ma anche per gli agenti conversazionali appartenenti alla terza tipologia individuata *supra* al § 1, i quali, pur non essendo concepiti per tali finalità, presentano caratteristiche tali da renderne prevedibile l'utilizzo da parte degli utenti come strumenti di espressione del disagio e di ricerca di sostegno emotivo²⁰⁵. Il fatto che, allo stato delle informazioni pubblicamente disponibili, i fornitori di tali sistemi non risultino generalmente avvalersi del contributo di professionisti dell'area psicologica o psichiatrica non assume, di per sé, rilievo decisivo. Il contenuto del dovere di diligenza non si esaurisce in ciò che usualmente si compie, ma si definisce in relazione a quello che si *deve fare* (v. *supra*).

A tale conclusione può pervenirsi, oltre che in coerenza con l'elaborazione dottrinale maggioritaria, anche alla luce della disciplina dettata dall'art. 9 AI Act, che prevede l'istituzione di un modello di gestione dei rischi volto a dispiegarsi lungo l'intero ciclo di vita del sistema di IA²⁰⁶. Applicata nel contesto dei *chatbot therapists*, la

²⁰⁴ Non si intende affermare che, nel settore dei *chatbot therapists*, l'*homo eiusdem professionis et condicionis* sia strutturalmente interdisciplinare (nel senso di un agente modello che cumuli competenze informatiche e conoscenze proprie delle scienze cognitive, psicologiche o psichiatriche). Piuttosto, si vuole rilevare che il "fornitore modello" del sistema è colui che organizza la propria attività in modo tale da disporre delle risorse e delle competenze necessarie a comprendere e gestire le dinamiche rischiose connesse a tali applicazioni, la cui adeguata valutazione presuppone il ricorso a saperi in ambito psicologico o psichiatrico. Il punto merita tuttavia un'ulteriore precisazione. Ci si potrebbe infatti chiedere se le eventuali conoscenze superiori alla media dei consulenti di cui il fornitore si avvale giustifichino un innalzamento dello standard di diligenza in capo a quest'ultimo. La dottrina osserva, in proposito, come sia «piuttosto comune ammettere che, una volta fissato lo *standard* della condotta richiesta per il tramite dell'*homo eiusdem*, quello stesso *standard* possa poi essere 'innalzato' in funzione delle maggiori e specifiche conoscenze e capacità del soggetto» – GROTTO (2012), p. 193. Ritenere diversamente, prosegue l'Autore, «significherebbe mandare esente da pena un agente gravemente sospettato di responsabilità: l'agente dotato di conoscenze superiori, infatti, potrebbe astenersi dal fare quanto gli è facilissimo fare (con un elevato grado di rimproverabilità soggettiva) per il solo fatto che la media degli uomini non possiede quelle conoscenze che egli, invece, possiede» – *ivi*, p. 194. Salve le questioni relative al luogo sistematico in cui tale profilo debba essere apprezzato – se sul terreno della tipicità o su quello della colpevolezza – in questa sede preme evidenziare che le eventuali conoscenze superiori dei consulenti del fornitore, di cui sovente le stesse piattaforme danno evidenza nella presentazione dei propri servizi, non dovrebbero rilevare ai fini dell'innalzamento dello standard di diligenza esigibile da quest'ultimo. In tal caso, infatti, si finirebbe per ammettere una forma di attribuzione di responsabilità penale per fatto altrui. Tali conoscenze potranno, piuttosto, rilevare ai fini della valutazione della diligenza esigibile dai consulenti medesimi, qualora ne ricorrano i presupposti per un concorso nel fatto con il fornitore. All'inverso, qualora le competenze di tali soggetti risultino palesemente inferiori alla media e inidonee a fornire le conoscenze necessarie per assolvere al *dovere di riconoscere* e di *evitare* il rischio, al fornitore potrà, al più, imputarsi una *culpa in eligendo*.

²⁰⁵ Tale ricostruzione risulta avvalorata dalla circostanza che i suicidi segnalati hanno riguardato prevalentemente utenti di agenti conversazionali non progettati per finalità di supporto psicologico, riconducibili alla terza tipologia individuata *supra*, § 1, e cioè agenti conversazionali non progettati né presentati come strumenti finalizzati al supporto psicologico di utenti con disturbi psichici, ma in effetti utilizzati da questi per tale scopo, in ragione di specifiche architetture del sistema (quando, ad esempio, consentano di attribuire al personaggio tratti riconducibili a figure affettivamente significative) e del target di riferimento.

²⁰⁶ L'art. 9 del Regolamento stabilisce che, in relazione ai "sistemi di IA ad alto rischio", debba essere istituito, attuato, documentato e mantenuto un "sistema di gestione dei rischi" – inteso come processo iterativo continuo lungo l'intero ciclo di vita del sistema – soggetto a costante riesame; esso comprende l'identificazione e l'analisi dei rischi noti e ragionevolmente prevedibili, la stima e valutazione dei rischi anche in condizioni di uso improprio ragionevolmente prevedibile, la valutazione di ulteriori rischi derivanti dal monitoraggio successivo all'immissione sul mercato e l'adozione di misure di gestione dei rischi opportune e mirate. Non si tratta, beninteso, di una norma volta a delineare

natura dei rischi implica che tale analisi non possa essere correttamente svolta se non mediante il ricorso a competenze specialistiche nell'ambito della salute mentale. Vero è che la disciplina di cui all'art. 9 del Regolamento è formalmente riferita ai soli fornitori di sistemi "ad alto rischio", tra i quali – come si è rilevato – non rientrano, ad oggi, i *chatbot therapists*. Tuttavia, nella prospettiva che si è già delineata *supra*, § 2.1.1, l'elenco formale dei destinatari delle *regole cautelari*, ossia del contenuto della diligenza (e non dei *doveri di diligenza*! V. *infra*, § 4.1), non vale ancora a restringere i rischi doverosi da riconoscere, né, di conseguenza, a determinare le cautele da adottare, nei termini che ora si precisano²⁰⁷.

4.1. Le cautele previste ma non imposte dall'AI Act. Il criterio del rischio tipico nel giudizio di evitabilità.

Dopo aver sondato il referente sociale degli standard cautelari in materia, si tratta ora di esaminare il contributo che, sul piano dell'evitabilità, può offrire la disciplina dettata dall'AI Act. In proposito, viene in rilievo la classica distinzione tra *regole cautelari proprie* – idonee ad azzerare o quasi i rischi di concretizzazione dell'evento – e *regole cautelari improprie*, le quali, pur non eliminando il rischio, mirano a contenerlo²⁰⁸.

Le prescrizioni dettate dal Regolamento con riferimento alle applicazioni consentite dei sistemi di intelligenza artificiale si riconducono alla seconda categoria. Tale inquadramento comporta almeno due conseguenze sul piano della determinazione del contenuto del *dovere di evitare* dei fornitori di *chatbot therapists*.

In primo luogo, seguendo l'orientamento dottrinale più accreditato, in caso di violazione di regole cautelari improprie, l'efficacia salvifica del comportamento alternativo diligente può accertarsi tramite il criterio del "più probabile che non" (e non, invece, dell'infallibilità del decorso alternativo salvifico)²⁰⁹.

Secondariamente, poiché le regole cautelari improprie non garantiscono una prevenzione assoluta, viene da escludere che la loro osservanza possa, *a priori*, esaurire il contenuto del dovere di diligenza. La disponibilità di nuove o diverse cautele idonee a ridurre ulteriormente il rischio²¹⁰ – già riconoscibile – rende non più giustificabile la probabilità di offesa residua derivante dalla loro trasgressione. Applicando tali coordinate nel contesto in esame, il rispetto degli obblighi di trasparenza – a cui, ai sensi dell'art. 50 del Regolamento, sono (per ora) soggetti i fornitori di *chatbot therapists* – non

una posizione di garanzia (nel qual caso non sarebbe il rischio tipico a determinarne l'ambito di operatività, v. *infra*); essa, piuttosto, mira a delineare un protocollo cautelare, suscettibile di rilevare rispetto al contenuto del dovere di diligenza sul versante organizzativo.

²⁰⁷ In ogni caso, anche a non voler aderire alla ricostruzione che segue, il ricorso a competenze specialistiche nell'ambito della salute mentale conserva rilievo rispetto a un'ipotetica *colpa per assunzione* del fornitore: un criterio che, com'è noto, valorizza il contenuto delle conoscenze e competenze che l'agente è tenuto a possedere (o ad acquisire) in ragione dell'attività rischiosa intrapresa; sul punto, v. *infra*, § 4.2.

²⁰⁸ Su tale distinzione, v. VENEZIANI (2003), pp. 15 ss.

²⁰⁹ PIERGALLINI (2017), p. 243. In senso analogo, anche DONINI (2006), p. 123, già richiamato *supra*, § 4, nota n. 197.

²¹⁰ Giova precisare che la "disponibilità" della regola cautelare va «intesa in senso "normativo" (e non, invece, in senso assoluto od ontologico), e viene a coincidere non con la sua semplice "esistenza", bensì con il momento in cui la stessa divenga "patrimonio cognitivo consolidato e diffuso"» – CASTRONUOVO (2011), pp. 1626-1627.

sembra esaurire il contenuto della diligenza gravante su tali operatori. Nell'economia dell'AI Act, infatti, possono individuarsi regole cautelari diverse che, pur non formalmente prescritte ai fornitori di *chatbots* per il supporto psicologico (essendo sancite per i "sistemi di IA ad alto rischio") coprono rischi tipici coincidenti con quelli di tali agenti conversazionali.

Si tratta, allora, di stabilire se i fornitori di IA – e, più in generale, gli esercenti attività pericolose – siano tenuti al rispetto di *disposizioni* idonee a governare i rischi connessi al loro operato, pur *non appartenendo* formalmente al novero di soggetti cui esse si rivolgono.

Orbene, se si assume che le classi di rischi riconoscibili (nel senso: *doverosi da riconoscere*) mantengano tale qualificazione indipendentemente dalle opzioni contingenti del legislatore circa i destinatari delle regole cautelari a esse rivolte (v. *supra*, § 2.1.1), ne consegue che anche quelli *doverosi da evitare* seguano la stessa logica. Di pari passo, se si ritiene che, sul piano del diritto positivo, i rischi *da riconoscere* siano determinabili in base al rischio tipico delle regole di diligenza, ne deriva che gli stessi – una volta così individuati – sono anche quelli la cui concretizzazione è *da evitare*. Risulta, allora, che il criterio del rischio tipico orienta non solo la selezione dei rischi da riconoscere, ma anche le cautele da adottare²¹¹. In altri termini, la doverosità delle regole cautelari si radica nella loro efficacia preventiva rispetto a rischi tipici²¹².

In coerenza con tale approdo, viene in rilievo lo sviluppo scientifico – elaborato in dottrina – secondo cui «Le regole cautelari [...] sono *sempre* valide – indipendentemente dalla fonte che le recepisce e dalla sua eventuale abrogazione – sulla base della loro efficacia preventiva, oltre che della pertinenza al caso concreto»²¹³.

Provando ora a rispondere al quesito appena posto: se la regola cautelare è valida a prescindere dalle vicende che investono l'atto legislativo che le contempla, potrebbe allora ritenersi che essa conservi tale validità anche oltre il novero dei soggetti cui il provvedimento formalmente si rivolge (restando distinta l'ipotesi – che si esaminerà a breve – in cui le prescrizioni siano dirette a individuare i destinatari del dovere di diligenza il cui contenuto è da regolare)²¹⁴.

Questa soluzione potrebbe, nondimeno, apparire problematica alla luce del principio di legalità. In merito, risulta fondamentale la distinzione – offerta dalla dottrina italiana – tra "dovere di diligenza" e "diligenza (doverosa)" intesa come "contenuto del dovere" (espressione già richiamata e che può ora essere chiarita)²¹⁵. In particolare, afferma Castronuovo, «Per l'identificazione del primo, il *dovere* di diligenza, [...] valgono

²¹¹ Sul necessario raccordo tra il rischio concretizzatosi e lo spettro preventivo della regola cautelare violata, nonché sulle implicazioni che ne derivano quando una medesima regola incida su più rischi e aree di tutela, v. Cass. Pen., Sez. IV, 6 settembre 2021, n. 32899, cit., e gli studi di CONTRI (2025) sul caso del disastro di Viareggio.

²¹² CASTRONUOVO (2009), spec. p. 289. Si tratta in fondo, come si rileva nel testo citato, di un esito imposto dalla punibilità, nell'ordinamento italiano, della colpa generica (*ivi*, p. 188).

²¹³ *Ivi*, p. 286.

²¹⁴ In questo senso, si deve distinguere tra *destinatari del dovere di diligenza* e *destinatari del contenuto del dovere di diligenza*: i primi sono individuati dalla *norma* e non sono suscettibili di estensione oltre i limiti imposti dal principio di legalità; i secondi si identificano, invece, in coloro che esercitano un'attività pericolosa i cui rischi sono tipici rispetto a una regola cautelare, la quale, dunque, definisce (o concorre a definire) il contenuto del loro dovere di diligenza.

²¹⁵ *Ibidem*.

in maniera stringente i principi costituzionali in materia penale, e tra questi quello di tipicità e quello di legalità in tutte le sue articolazioni», mentre «Per l'individuazione del [...] second[o], trattandosi di “regole” strumentali, non necessariamente giuridiche, ossia di schemi di comportamento finalizzati alla riduzione o alla eliminazione del rischio di verificazione di una certa classe di eventi, quei principî valgono in misura diversa, condizionati come sono dalla presenza di una “clausola generale” come quella della responsabilità “per colpa”»²¹⁶. Ne deriva che la *legalità* attiene alla *previsione* della *responsabilità* colposa (e, semmai, della posizione di garanzia, dell'obbligo di evitare l'evento, etc.), mentre il contenuto del dovere di diligenza può essere determinato sia da doveri cautelari positivizzati, sia da regole “generiche” di diligenza, la cui inosservanza rileva proprio perché è il legislatore ad ammettere che la colpa possa consistervi (in Italia secondo il paradigma dell'art. 43 c.p., che si riferisce alla colpa generica e a quella specifica).

Si può dunque (ri)affermare che, nell'ambito della responsabilità *penale* dei fornitori di *chatbot therapists*, l'individuazione delle cautele doverose non dipende dalla qualificazione formale di questi agenti all'interno delle categorie previste dall'AI Act (“sistemi di IA ad alto rischio”, “modelli di IA per finalità generali”, “sistemi di IA destinati a interagire direttamente con le persone fisiche”)²¹⁷, bensì dall'idoneità delle stesse a contenere i rischi che l'impiego di tali *chatbots* è suscettibile di concretizzare – ciò salvo che sia lo stesso fatto tipico a selezionare, a monte, il dovere di diligenza (ad esempio, prevedendo espressamente la responsabilità colposa solo in caso di mancato rispetto di specifiche cautele riferite a specifici soggetti).

Fuori da quest'ultima ipotesi normativa – di cui si dirà a breve – se, come si è messo in luce, le regole cautelari rilevanti sul piano dell'evitabilità coincidono con quelle che definiscono il contenuto del *dovere di riconoscere* – essendo identificate tramite lo stesso parametro: il rischio tipico – si può ritenere che, nell'attuale cornice dell'AI Act, le cautele doverose per i fornitori di *chatbot therapists* siano quelle evocate, in via schematica, al § 2.1.1, dedicato alla riconoscibilità dei rischi connessi a tali agenti. In particolare, potrebbero rientrare nel novero non solo gli obblighi di trasparenza previsti dall'art. 50 per i sistemi che interagiscono direttamente con le persone fisiche (categoria nella quale, come si è già evidenziato, rientrano i *chatbot therapists*), ma anche le principali misure contenute nel capo III, tra cui l'art. 9 (gestione del rischio lungo l'intero ciclo di vita del sistema), l'art. 13 (funzionamento sufficientemente trasparente e interpretabilità dell'*output* ai fini di un uso consapevole), l'art. 14 (*human oversight*) e l'art. 15 (affidabilità tecnica del sistema), benché – si insiste – riferite ai “sistemi di IA ad alto rischio”²¹⁸.

²¹⁶ *Ibidem*.

²¹⁷ È, invece, sul terreno del regime sanzionatorio delineato dall'AI Act (artt. 99 ss.) che tale classificazione dispiega la propria rilevanza.

²¹⁸ *Supra*, § 2.1.1, si è già argomentato nel senso di escludere che il legislatore unionale abbia deliberatamente sottratto dal novero dei rischi tipici delle prescrizioni in esame quelli connessi ai *chatbot therapists*. Da un lato, in quanto tali applicativi rappresentano un caso paradigmatico dei rischi per la salute, anche psicologica, cui i considerando nn. 5, 28 e 29 AI Act si riferiscono e verso cui si dirigono le regole in esame. Dall'altro, la mancata inclusione dei *chatbot therapists* tra i “sistemi di IA ad alto rischio” appare più verosimilmente riconducibile alla fisiologica asimmetria temporale tra evoluzione tecnico-applicativa dell'IA e regolazione, messa particolarmente in tensione nella fase

Nel confronto con l'attuale dato positivo, tuttavia, il quadro sin qui ricostruito – valevole sul piano delle categorie generali – va misurato con due aspetti specifici.

Il primo attiene alla funzione che deve riconoscersi a questo secondo gruppo di prescrizioni – quelle contenute nel capo III del Regolamento –: se sia volto a delimitare i destinatari del dovere di diligenza e, dunque, non sia estensibile ai fornitori di *chatbot therapists* per il vincolo di legalità, ovvero esprima il contenuto regolativo del suddetto dovere. In altri termini, se si tratti di norme giuridiche ovvero di regole cautelari²¹⁹. Un argomento nel senso della seconda opzione può trarsi dalla nomenclatura delle sezioni del capo III del Regolamento. Nello specifico, mentre la sezione II – in cui sono inseriti gli artt. 9, 13, 14 e 15 – è intitolata “Requisiti per i sistemi di IA ad alto rischio”, la sezione III è rubricata “Obblighi dei fornitori e dei deployer dei sistemi di IA ad alto rischio e di altre parti”. Tale distinzione lessicale, considerata unitamente al tenore complessivo delle disposizioni, è quantomeno compatibile con l'idea che le prescrizioni contenute nella sezione II non siano dirette a individuare i destinatari del *dovere di evitare*, ma concorrano piuttosto a precisare il suo *contenuto*²²⁰.

Il secondo aspetto con cui misurarsi riguarda la formulazione dell'art. 437-bis c.p.; qui, per chiarezza delle osservazioni che seguono, conviene riportarne i commi 1 e 3: «Chiunque omette di adottare le misure tecniche di sicurezza, previste per la progettazione, l'addestramento, la produzione, l'immissione sul mercato di sistemi di intelligenza artificiale ad alto rischio, idonee a prevenire malfunzionamenti o alterazioni del funzionamento dei sistemi ovvero omette di adottare misure di sorveglianza umana è punito con la reclusione da uno a cinque anni, quando da tali omissioni derivi pericolo per la vita o l'incolumità pubblica o individuale. Qualora dal fatto derivi un pericolo per la sicurezza dello Stato, la pena è della reclusione da due a otto anni»; «Se taluno dei fatti previsti dal comma primo è commesso per colpa grave, la pena è ridotta da un terzo a un sesto». Alla luce di tale costruzione normativa, e in considerazione del rinvio alle definizioni dell'AI Act operato dal D. Lgs n. 160 del 2026 nell'introdurre la disposizione, quanto sin qui ricostruito sul piano del rischio tipico quale criterio di individuazione delle cautele doverose va declinato entro i limiti posti dal fatto tipico sopra esposto. Il legislatore, infatti, con riguardo all'omissione delle *misure tecniche di sicurezza* – tra cui rientrano quelle previste dagli artt. 8-15, dall'AI Act (v. *supra*) – sembra averne ristretto la rilevanza penale in ordine ai soli sistemi di IA ad alto rischio; si tratterebbe, dunque, di un caso in cui è lo stesso legislatore a selezionare, a monte, il *dovere di diligenza*, riferendolo, per questo segmento della fattispecie, ai soggetti coinvolti nelle fasi di progettazione, addestramento, produzione e immissione sul mercato di sistemi formalmente etichettati come ad alto rischio. Conseguenza di questa esegesi è che, ai

conclusiva del processo di adozione del Regolamento per effetto della rapida diffusione dell'IA generativa.

²¹⁹ Insegna la distinzione terminologica tra “norme giuridiche” e “regole cautelari” sempre CASTRONUOVO (2009), pp. 284 ss. Egli chiarisce: «le regole cautelari non hanno una natura necessariamente “giuridica”, ma si fondano solitamente sul *paradigma etiologico o nomologico* [...] oppure si fondano sull'esperienza» (*ivi*, p. 284) laddove, invece, «il dovere di diligenza è posto dalle *norme*» (*ivi*, p. 287).

²²⁰ *Dovere di evitare* (e dunque di diligenza) che, giova ripetere, trova fondamento in caso di previsione della responsabilità in ordine a determinati fatti, dannosi o pericolosi; *contenuto del dovere di evitare* (o diligenza doverosa) che, invece, attiene agli «schemi di comportamento [precisati nel dato positivo e non] finalizzati alla riduzione o alla eliminazione del rischio di verificazione di una certa classe di eventi» – CASTRONUOVO (2009), p. 286.

sensi dell'art. 437-*bis*, c.p., la mancata osservanza delle misure tecniche di sicurezza previste dall'AI Act, anche qualora dirette a governare rischi tipici connessi ai *chatbots therapists*, non avrebbe rilevanza penale per i fornitori di tali sistemi – non essendo etichettati come “ad alto rischio” – per espressa opzione legislativa; la loro omissione potrebbe venire in rilievo solo nell'ambito di una diversa fattispecie colposa – ragionevolmente di pericolo, dunque, *de lege ferenda* – nella quale tale delimitazione non operi.

L'alternativa, non priva di asperità, sarebbe ritenere che la formulazione dell'art. 437-*bis*, comma 1, c.p. non assuma come termine di selezione i sistemi di IA ad alto rischio in quanto tali (con i relativi fornitori, *deployer*, etc.), bensì le regole di sicurezza «previste per» i sistemi ad alto rischio; del resto, la disposizione prosegue precisando che le regole di sicurezza previste per i sistemi di IA ad alto rischio devono essere «idonee a prevenire malfunzionamenti o alterazioni del funzionamento *dei sistemi*», e non “*di tali sistemi*” riferito ai soli sistemi ad alto rischio (corsivo nostro). A sostegno di questa lettura potrebbe porsi la rubrica della norma – «Omessa adozione di misure tecniche di sicurezza nei sistemi di intelligenza artificiale e alterazione illecita dei sistemi» – che si riferisce genericamente ai «sistemi», senza ulteriori specificazioni. Nello stesso senso potrebbe essere letto il comma 4, il quale, quando limita la responsabilità ai soggetti coinvolti nell'agire dei soli sistemi di IA ad alto rischio, lo afferma espressamente con riguardo all'«utilizzatore professionale di sistemi di intelligenza artificiale ad alto rischio».

Problematico è stabilire se la selezione rispetto ai soli fornitori di sistemi di IA ad alto rischio non sia stata operata in relazione all'omissione di *misure di sorveglianza umana*, che l'art. 14 dell'AI Act rende obbligatorie per i soli sistemi ad alto rischio, ma che risultano funzionali a contenere anche rischi propri dei *chatbot therapists*, e, in particolare, la tendenza dell'utente «a fare automaticamente affidamento o a fare eccessivo affidamento sul sistema» (v. *supra*, § 2.1.1.). Infatti, nella formulazione dell'art. 437-*bis*, comma 1, c.p. le *misure di sorveglianza umana*, diversamente dalle misure tecniche di sicurezza, potrebbero finanche leggersi come svincolate dal rapporto con i sistemi qualificati come “ad alto rischio” (non comparando nemmeno tra le definizioni di cui all'art. 3 dell'AI Act, alle quali rinvia l'art. 11 del D. Lgs. n. 160 del 2026).

In questa prospettiva, seguendo la logica del rischio tipico, l'omessa adozione di misure di sorveglianza umana da parte dei fornitori di *chatbot therapists*, qualora ne derivi un pericolo concreto per la vita o l'incolumità individuale, potrebbe assumere rilevanza ai sensi dell'art. 437-*bis* c.p.

Il rilievo non si arresta, tuttavia, alla colpa grave (specifica). Se, entro la ricostruzione dogmatica posta a fondamento della metodologia qui seguita (v. *supra*, § 2.2.2, spec. nota n. 172), la diligenza, quantomeno sul versante del *dovere di evitare*, si colloca su un piano oggettivo comune non soltanto alla colpa, ma anche al dolo, la medesima omissione può assumere rilievo anche a titolo doloso. Ne consegue che l'omessa adozione delle misure di sorveglianza umana ex art. 14 AI Act da parte del fornitore di *chatbot therapists* potrebbe ricadere già nel comma 1 dell'art. 437-*bis* c.p., ove ne ricorrano i presupposti del dolo eventuale rispetto all'evento di pericolo.

Resta, tuttavia, aperto un interrogativo rispetto agli obblighi di trasparenza di cui all'art. 50 AI Act, già direttamente gravanti sui fornitori di *chatbot therapists* ai sensi del Regolamento: quanto alla loro rilevanza ai sensi dell'art. 437-bis c.p., essi non paiono riconducibili né alle misure tecniche di sicurezza previste per i sistemi ad alto rischio, né a quelle di sorveglianza umana.

A questo punto, merita di essere considerato altresì il profilo dell'effettiva operatività di tali prescrizioni e, in particolare, dell'art. 14, avuto riguardo al regime di applicazione temporale delineato dall'AI Act. Sebbene quest'ultimo sia formalmente entrato in vigore il 1° agosto 2024, non tutte le sue disposizioni sono ancora obbligatorie ai sensi del Regolamento: in particolare, le regole di trasparenza di cui all'art. 50 sono efficaci dal 2 agosto 2026, mentre, a seguito delle modifiche apportate dal Regolamento (UE) 2026/1744, per i requisiti previsti dalla sezione II del capo III – tra cui gli artt. 9, 13, 14 e 15 – è attualmente previsto un regime di applicazione differita, decorrente, a seconda dei casi, dal 2 dicembre 2027 e dal 2 agosto 2028.

Si pone, pertanto, la questione se tali misure possano già oggi assumere rilievo sul piano della diligenza penale, pur in assenza di una loro immediata cogenza sul versante amministrativo.

Orbene, posta l'autonomia del diritto penale, il tema si presta a essere illuminato dalla dottrina sopra richiamata, che chiarisce: «Se possono concorrere a integrare la fattispecie colposa anche regole “generiche” di diligenza (come stabilisce l'art. 43 c.p.), deve ammettersi che, qualora tali regole siano state “specificate” mediante la previsione in un testo formale di normazione [...] le stesse risultino pur sempre applicabili anche se l'atto normativo che le recepisce non sia ancora entrato in vigore oppure –laddove si tratti di una legge – sia stato dichiarato incostituzionale»²²¹. Ed è qui che essa fornisce il passaggio decisivo per la questione posta, rimarcando che il discorso si estende «anche alle ipotesi di leggi non ancora vigenti o non più vigenti»²²².

In tale prospettiva, il differimento dell'applicazione delle disposizioni dell'AI Act non risulta decisivo ai fini della loro eventuale rilevanza sul piano della diligenza penale, ove esse risultino idonee a esprimere il contenuto del dovere di tale diligenza.

4.2. L'espansione della diligenza: tra dinamiche di ingerenza e profili di colpa per assunzione.

Se, nei paragrafi precedenti, l'attenzione si è concentrata sulla soglia primaria del *dovere di riconoscere* i rischi e di evitare l'evento – il pericolo per la vita e l'incolumità

²²¹ *Ivi*, p. 186.

²²² *Ivi*, p. 187. Nell'opera si offre una penetrante ricostruzione sul punto: «Il principio di irretroattività [...] assume, con riferimento alle regole cautelari (contenuto del dovere di diligenza), un significato differenziato. La regola cautelare è applicabile se soddisfa le condizioni seguenti: occorre che sia “vigente” in senso sostanziale, ossia che definisca uno schema di comportamento valido ed efficace rispetto allo scopo preventivo cui è orientata; occorre, inoltre, che la regola cautelare pre-esista *ex ante* (al momento della condotta inosservante), non potendo “nascere” solo *ex post*, ossia né nell'imputazione-accusa, né in sentenza. Per contro, è solo con riferimento al *dovere di diligenza* che il principio di irretroattività assume il suo significato più pieno: ossia, quello richiesto, nel diritto penale, in funzione di garanzia per il cittadino, a partire dall'Illuminismo» (*ivi*, p. 190).

individuale di utenti con disturbi psichici –, è ora rispetto alla sua possibile espansione che si volge l'indagine.

Seguendo la tripartizione proposta *supra* al § 1, con riferimento ai *chatbots* espressamente concepiti e presentati come strumenti di supporto terapeutico psicologico non clinico (i) e, benché in misura minore, a quelli progettati e pubblicizzati per finalità di benessere emotivo (ii) si affaccia la questione se le modalità di offerta del servizio non valgano, sul piano penalistico, quale forma di *assunzione volontaria di compiti di tutela*²²³. Una siffatta scelta di intervento, alla luce delle più recenti derive giurisprudenziali in tema di ingerenza, potrebbe fondare un obbligo giuridico di *protezione* ai sensi dell'art. 40, comma 2, c.p., con il conseguente innalzamento dello standard di diligenza esigibile. Questo verrebbe a coincidere con l'adozione di *tutte* le misure utili a impedire l'offesa, comprensiva dell'evento morte dell'utente²²⁴.

L'attenzione è attratta dalle modalità con cui le piattaforme presentano il servizio, nei limiti di quanto emerge da una ricognizione dei rispettivi siti *web*. Le applicazioni dell'IA qui considerate risultano frequentemente accompagnate da enunciati promozionali che enfatizzano l'efficacia dello strumento – talora mediante il richiamo a percentuali di riduzione dei disturbi psichici degli utenti –, la costante accessibilità del servizio e la possibilità di fruire di uno spazio di interazione asseritamente privo di giudizio²²⁵. Altresì, merita considerazione l'accento posto sull'assenza dei tempi di attesa tipici dei percorsi tradizionali²²⁶.

In simile contesto, a venire in rilievo quale fonte di una posizione di protezione “da ingerenza” non è la stipulazione di un formale contratto o il ricorso all'art. 2028 c.c.

²²³ Il tema della posizione di garanzia si colloca, come noto, entro un quadro teorico tutt'altro che assestato, nel quale convergono istanze di protezione e spinte espansive della responsabilità penale. Il crescente e criticato rilievo riconosciuto all'ingerenza ha contribuito ad ampliare, in via pretoria, l'area dei soggetti gravati da obblighi impeditivi, attraverso il riconoscimento di forme di responsabilità fondate sull'assunzione – anche di fatto – di compiti di protezione. Il tutto, peraltro, non senza diffuse e condivisibili riserve in dottrina. La complessità e l'ampiezza del dibattito impediscono, in questa sede, una trattazione analitica, per la quale si rinvia – anche per ulteriori riferimenti – a CARRIERO (2023).

²²⁴ In proposito, va subito segnalato un profilo problematico: la fisiologica imprevedibilità del gesto suicidario – da non confondere con la prevedibilità ed evitabilità dell'incremento del rischio (v. *supra*, § 1) – pone un margine di limitatezza dei poteri impeditivi dell'evento, come evidenziato dalla dottrina in tema di posizione di garanzia dello psichiatra – cfr. CUPELLI (2016), *Id.* (2017), pp. 370 ss.

²²⁵ A fini meramente esemplificativi, si riportano alcune enunciazioni promozionali tratte da siti di *chatbots* destinati al supporto psicologico (traduzioni di chi scrive): «Chatting with Tess led to reduced symptoms of depression (-28%) and anxiety (-18%)» (Conversare con Tess ha comportato una riduzione dei sintomi di depressione [-28%] e di ansia [-18%]), reperita all'indirizzo: <https://www.x2ai.com/uprisehealth> (ultima consultazione: 30 settembre 2026); «Completely anonymous. No stigma. No limits» (Completamente anonimo. Nessuno stigma. Nessun limite), «People everywhere are turning to AI. Wysa is the safe, evidence-based alternative. It is always available and designed to integrate with your care model» (Sempre più persone si rivolgono all'IA. Wysa rappresenta un'alternativa sicura e basata su evidenze. È sempre disponibile ed è progettata per integrarsi con il tuo modello di cura), «Scale of impact: 11 million lives covered; 105 countries; 45+ peer-reviewed publications; 10x access to mental health support; 91% of users find Wysa helpful; 6 million people helped; 33% reduction in lost-time days; 1 billion AI conversations» (Scala dell'impatto: 11 milioni di persone raggiunte; 105 paesi; oltre 45 pubblicazioni sottoposte a revisione tra pari; accesso al supporto per la salute mentale aumentato di 10 volte; il 91% degli utenti ritiene Wysa utile; 6 milioni di persone aiutate; riduzione del 33% dei giorni di lavoro persi; 1 miliardo di conversazioni con IA), reperita all'indirizzo: <https://www.wysa.com> (ultima consultazione: 30 settembre 2026).

²²⁶ Emblematica, sotto questo profilo, la seguente formulazione «No matter the time of day, Wysa is ready to help: no appointments, no waiting, just instant access to care» (A prescindere dall'ora del giorno, Wysa è pronta ad aiutare: nessun appuntamento, nessuna attesa, accesso immediato all'assistenza, traduzione di chi scrive), reperita all'indirizzo: <https://www.wysa.com> (ultima consultazione: 30 settembre 2026).

– quest’ultima norma, come condivisibilmente affermato in dottrina, riguarda i soli beni patrimoniali²²⁷ – quanto, potrebbe sembrare, l’obbligo nascente dal c.d. contatto sociale o, più verosimilmente, la *volontaria assunzione della cura di un terzo*²²⁸.

Come evidenziato dalla scienza penalistica, tuttavia, il comportamento concludente, anche quando comporta aumento del rischio per il bene tutelato, non è sufficiente a determinare una posizione di garanzia: necessario è, infatti, un *presupposto ulteriore*, che dimostri sia la prevedibilità dell’evento e il potere di impedirlo, sia la *volontà* del soggetto agente di assumere su di sé specifici obblighi e poteri di tutela²²⁹.

Sennonché – e salvi gli aspetti problematici afferenti al potere *impeditivo* a fronte del suicidio²³⁰ – la dimostrazione di una siffatta volontà si rivela tutt’altro che agevole ove si considerino, ancora una volta, i contenuti *online* delle piattaforme di riferimento. All’enfasi sull’efficacia, la maggiore accessibilità, le tempistiche e i costi ridotti, l’analisi empirica – a cui questa parte dello studio non può rinunciare – mostra la successiva precisazione, da parte delle principali piattaforme, che lo strumento non è finalizzato a sostituire l’intervento dei sanitari né a gestire crisi suicidarie. In tali casi, si verifica una comunicazione a tratti ambivalente, idonea: da un lato, a ingenerare affidamento; dall’altro, a circoscrivere la portata della responsabilità che l’agente si assume.

Se una posizione di garanzia in capo al fornitore sembra, allo stato, da escludersi, la specificità degli interventi sopra delineati, in un ambito quale quello della salute mentale, lascia residuare un interrogativo sul versante della colpa per assunzione. Come si è espresso in dottrina, «Parlare di colpa per assunzione sembra quasi un’assunzione di colpa, ove in essa si ravvisi il fondamento di imputazioni anticipate»²³¹: nondimeno, proprio nella prospettiva di scongiurare esiti ricostruttivi in contrasto con il principio del *nulla poena sine culpa*, il tema merita una (breve) considerazione. La figura in esame si delinea come violazione di un obbligo di adeguata informazione (e, dunque, di *facere*), che si attualizza dinanzi all’esercizio di un’attività cui tale informazione è funzionalmente riferita²³². In questa sede, si tratta dunque di verificare se l’offerta di servizi volti a incidere sulla sfera della salute mentale di utenti anche con disturbi psichici presupponga il possesso, o quantomeno l’acquisizione, di conoscenze specialistiche in ambito psichiatrico/psicologico.

²²⁷ Cfr. MANTOVANI (2001), spec. par. 7 – riferimento tratto da CARRIERO (2023), p. 968.

²²⁸ Ai fini del riconoscimento della volontaria assunzione della cura di un terzo, la giurisprudenza ha individuato i seguenti criteri: «Occorre in primo luogo valutare “la particolare vulnerabilità del bene, a causa dell’incapacità del titolare di proteggerlo. [...] la tutela del bene [deve essere inoltre] affidata, anteriormente alla verifica della situazione di pericolo, a un terzo, che è per l’appunto il garante. Infine [...] il garante [deve essere] in grado di intervenire sullo svolgimento dei decorsi causali, in modo da impedire la verifica dell’evento tipico; [...] è il comportamento del garante ad aver determinato un aumento del rischio per il bene da proteggere”» – CARRIERO (2023), p. 972, la quale riprende Cass. Pen., Sez. V, 19 luglio 2021, n. 27905 (Vannini-*bis*).

²²⁹ *Ivi*, p. 973.

²³⁰ V. CUPELLI (2016) e, per un primo richiamo, *supra*, nota n. 224.

²³¹ PIVA (2020), p. 1.

²³² *Ivi*. Detto altrimenti, si versa in colpa per assunzione «quando le regole cautelari violate concernono le capacità fisico-intellettuali e le conoscenze tecniche necessarie all’espletamento di un’attività che, posta in essere in assenza di esse, ha conseguentemente condotto alla realizzazione di un evento offensivo; salvo, ai fini di un’imputazione (autenticamente) colpevole del fatto, l’accertamento della circostanza che l’agente conosceva o avrebbe comunque dovuto conoscere la sua inadeguatezza rispetto allo svolgimento dell’attività da intraprendere» – MEZZETTI (2010), pp. 513-514.

All'interrogativo, in realtà, sembra doversi fornire risposta positiva anzitutto sul piano della diligenza obiettiva del fornitore. Con riferimento al contenuto del *dovere di riconoscere* (oltretutto *di evitare*), si dovrebbe valorizzare l'art. 9 dell'AI Act²³³, ritenendolo operante nel contesto dei *chatbot therapists* in virtù del criterio del rischio tipico, (v. *supra*, § 2.1.1 sulla riconoscibilità e § 4.1 per l'evitabilità, nei limiti delle opzioni esegetiche del dato normativo vigente *ivi* riportate). Ma anche considerando l'esperienza applicativa più recente, come si è riportato (v. *supra*, § 4), il ricorso a figure dotate di competenze psichiatriche si presenta come elemento strutturale nelle fasi di progettazione e di funzionamento dei *chatbot therapists*, necessario a garantire un servizio *clinically safe*.

Nondimeno – venendo a quanto qui rileva – il ricorso a tali competenze è ciò che permette al fornitore non solo di *riconoscere* i rischi ed *evitare* il loro realizzarsi, ma, più nello specifico, di intendere le cautele da osservare nel trattamento dei disturbi psichici e, dunque, di *verificare* la possibilità di rispettarle nello svolgimento del servizio. In altri termini, solo il possesso (o l'acquisizione tramite profili esperti) di una preparazione nel campo della salute mentale consente di valutare la possibilità di conformarsi alle *leges artis* (di generalizzata validità) che legittimano l'esercizio dell'attività pericolosa – nel contesto in esame, l'intervento terapeutico sulla salute mentale altrui. Da quest'angolo visuale, qualora, nonostante la *conoscibilità* della sua inidoneità – ottenibile tramite il ricorso a suddette competenze –, il fornitore di *chatbot therapists* decida di intraprendere l'attività, un'assunzione colposa di un'attività potrebbe ravvisarsi, sempreché da ciò derivi un'offesa²³⁴.

Non si tratta, è bene chiarirlo, di un innalzamento del livello di diligenza (che potrebbe aversi, invece, qualora venisse ravvisata una posizione di garanzia in capo al fornitore). Nell'ipotesi di colpa per assunzione, la violazione concerne regole cautelari – deducibili da fonti formali e dalle *leges artis* – consolidate tra gli appartenenti al settore di attività che si intraprende²³⁵. Nel caso dei *chatbot therapists*, l'ambito di riferimento non può essere ridotto a quello della sola messa a disposizione del pubblico di sistemi di IA, ma andrebbe più correttamente individuato nell'intervento – sia pure mediato tecnologicamente – nella sfera della salute mentale degli utenti (a prescindere dalla volontà del fornitore di assumere obblighi di cura, valevole – come si è esposto – sul distinto piano degli obblighi di impedimento). Del resto, «versare in colpa per

²³³ Si riprende, per chiarezza, l'art. 9, par. 2, lett. a), AI Act, nella parte in cui, con riguardo alla fase iniziale del sistema di gestione del rischio, richiede l'«identificazione e analisi dei rischi noti e ragionevolmente prevedibili che il sistema di IA ad alto rischio può porre per la salute, la sicurezza e i diritti fondamentali, quando è utilizzato conformemente alla sua finalità prevista». L'individuazione dei rischi per la salute mentale derivanti dall'interazione con i *chatbot therapists* richiede, ontologicamente, il ricorso a conoscenze in ambito psichiatrico/psicologico.

²³⁴ Beninteso: la colpa per assunzione non deriva *in re ipsa* dalla sola incapacità dell'agente di conformarsi alle regole d'azione: necessario è che accertare che egli, prima di intraprendere l'attività, *potesse* già conoscere tale incapacità – cfr., MEZZETTI (2010), p. 515-516. Puntualizzazione parimenti opportuna: «il giudizio di colpa [per assunzione] non può arretrare oltre l'inizio dell'attività rischiosa» – PISANI (2012), p. 150, al quale si rinvia anche per le critiche alla ricostruzione della colpa per assunzione come violazione di un dovere di astensione allorché lo svolgimento dell'attività rischiosa richieda un livello di conoscenze superiore a quello possedute dall'agente (*ivi*, pp. 106 ss.). A questo proposito, l'Autore pone l'esempio del datore di lavoro, il quale, gravato da una posizione di garanzia, non potrebbe comunque astenersi dall'agire. La tesi della colpa per assunzione quale violazione di un dovere di astensione dinanzi ad attività eccedenti il patrimonio di competenze dell'agente è, invece, sostenuta, tra gli altri, da FORTI (1990), pp. 291 ss.; MANTOVANI (1997), p. 144; MEZZETTI (2010), p. 514.

²³⁵ MEZZETTI (2010), p. 528.

assunzione significa tradire l'aspettativa dei consociati che ci si adegui alle regole precauzionali note all'agente modello dello specifico campo d'azione»²³⁶: una tutela dell'affidamento che verrebbe meno solo qualora l'utente fosse messo in condizione di presumere che l'agente – il fornitore – si discosterà dalla condotta diligente²³⁷. Condizione che difficilmente ricorrerebbe nel caso di fornitori che, fin dalla pagina di apertura del servizio, enfatizzino l'efficacia dello strumento, ne rivendichino la validazione scientifica e ne prospettino l'idoneità a incidere positivamente su stati quali l'ansia e la depressione: l'utente non potrebbe che essere indotto a confidare nella conformità dell'attività alle regole cautelari proprie del settore trattamentale della salute mentale.

Il tema dell'affidamento dell'utente, evocato in partenza per illuminare i profili di rischio dei *chatbot therapists* (v. *supra*, § 1), e a cui ora ha ricondotto il percorso di indagine, si rivela snodo decisivo anche ai fini della sua conclusione. Esso segna il punto di convergenza di due questioni, qui affrontate per ultime non in ordine di importanza, ma per la loro attitudine a orientare e a essere orientate nel futuro. Rispettivamente: l'autodeterminazione dell'utente quale limite alla tutela penale approntata in suo favore; la possibilità per lo stesso di accorgersi del pericolo a cui si autoespone, quale limite alla diligenza dell'agente.

5. Il principio di affidamento e l'autoesposizione al pericolo dell'utente. Brevi spunti di impronta vittimologica e vittimodogmatica.

In tema di principio di affidamento – quale criterio di delimitazione della diligenza esigibile nei casi di interazione tra soggetti – la dottrina ha ben evidenziato come esso operi in senso “biunivoco”: da un lato, ed è il profilo che si è appena esaminato, «come affidamento del consumatore sul rispetto degli standard di sicurezza da parte del produttore e sulla correttezza delle istruzioni e informazioni fornite»²³⁸; dall'altro, ed è l'aspetto che si prenderà ora in considerazione, «come affidamento del produttore su di un “uso normale o ragionevolmente prevedibile” del prodotto da parte del consumatore»²³⁹.

Rispetto al secondo profilo, emerge l'interrogativo – a tratti inedito – su quale possa dirsi l'uso “ragionevolmente prevedibile” di sistemi che, secondo la letteratura di settore esaminata, risultano idonei a ingenerare dinamiche di attaccamento emotivo, particolarmente in soggetti caratterizzati da vulnerabilità psichiche (v. *supra*, § 1). Così, guardando ai doveri di informazione²⁴⁰: può dirsi sufficiente (oltreché necessario) – ai

²³⁶ *Ivi*, p. 532.

²³⁷ In questo senso, su un piano più in generale rispetto al contesto in esame, v. ancora MEZZETTI (2010), p. 533.

²³⁸ CASTRONUOVO (2009), p. 328. Nella dottrina italiana sul principio di affidamento, v. per tutti MANTOVANI (1997); *Id.* (2009), pp. 536 ss.; MANTOVANI (2021), pp. 1 ss.

²³⁹ *Ibidem*.

²⁴⁰ I doveri di informazione sono enucleati da PIERGALLINI (2004), pp. 46-49, in relazione al terzo dei possibili difetti del prodotto, che l'Autore individua in: (a) *difetti di costruzione*, (b) *difetti di fabbricazione*, (c) difetti da informazione e (d) difetti da rischio di sviluppo – (*ivi*, p. 46). In particolare, «I difetti da informazione consistono nel mettere in circolazione un prodotto senza accompagnarlo con adeguate istruzioni relative al “modo di impiego”. Il prodotto è

fini dell'affidamento del fornitore sull'utente – l'avvertimento “da foglietto informativo” sull'eventualità che il sistema generi forme di affidamento eccedenti il controllo riflessivo dell'utente, sulla sua inidoneità alla gestione di crisi suicidarie, nonché sul rischio di progressiva sostituzione delle relazioni interpersonali? Oppure, alla luce della riconosciuta propensione umana all'affidamento sugli *outputs* dell'IA, devono ritenersi necessarie ulteriori misure (ad esempio, messaggi grafico-testuali reiterati nel corso dell'interazione sui limiti del servizio, ovvero indicatori costanti della natura artificiale dell'interlocutore)? In assenza di indicazioni provenienti dalla prassi (che avrebbero forse reso più agile la risposta), dovrebbero tali misure ritenersi doverose al pari di quelle delineate nei paragrafi precedenti in tema di evitabilità?

Si tratta di quesiti i cui risposte si collocano nel dominio delle scienze cognitive applicate, ma i cui effetti si spiegano anche sul piano della responsabilità penale²⁴¹. La scienza penalistica non è dunque deputata a fornirle, bensì dovrà rintracciarle²⁴².

Prima di concludere, resta imprescindibile confrontarsi con la questione che si colloca alle *fondamenta* della materia criminale: la tutela penale nei casi di lesioni o messa in pericolo di sé.

La scelta dell'utente di non conformarsi alle indicazioni, ove impartite, del fornitore e di sostituire l'interlocutore artificiale al personale medico/psicologo – con possibile aggravamento del rischio suicidario, anche in ragione delle dinamiche di isolamento che tendono a instaurarsi, dell'assenza di un monitoraggio sanitario dei “segnali di allarme” e, talvolta, della perdita di motivi esistenziali a favore della proiezione del sé in una realtà altra (virtuale)²⁴³ – non integra un'ipotesi di negligenza,

fabbricato a regola d'arte, ma appare difettoso perché omette di informare, o informa in modo incompleto, sul modo di impiego e sui pericoli connessi all'uso. In via di prima approssimazione, l'informazione può ritenersi corretta quando: 1) con riferimento alla forma, risulta “chiara”; 2) con riguardo al contenuto, appare “completa” (cioè indica i componenti del prodotto e informa degli eventuali rischi connessi all'uso), “intensa” (vale a dire proporzionata al rischio [...]) ed “evidente” (nel senso che l'avviso deve restituire *icto oculi* l'immagine della sfera di rischio contro la quale si intende mettere in guardia, anche tenuto conto della sua gravità)» (*ivi*, pp. 47-48).

²⁴¹ La soluzione a tali interrogativi, rimessa ad apporti scientifici ulteriori, implica uno studio centrato sulla figura della “vittima”. Si tratta, pertanto, di uno studio apprezzabile, almeno in parte, in ottica vittimologica: non già per la finalità – anch'essa propria di tale ambito – di individuare forme di riparazione del pregiudizio subito (estranea alla presente trattazione), bensì per la rilevanza che l'attenzione sulla persona offesa può assumere sul piano della prevenzione delle offese considerate. Per i profili connessi all'apporto e allo statuto della vittimologia, v. lo studio sistematico di TORDINI-CAGLI (2009), spec. pp. 5 ss., e l'opera di VENTUROLI (2015), pp. 13 ss.

²⁴² Non tanto, ovviamente, per volersi rifare al superato insegnamento dell'indirizzo tecnico-giuridico – che, con Arturo ROCCO, indica di mantenersi «religiosamente e scrupolosamente» ancorati allo studio del diritto vigente, nel quadro della «divisione del lavoro scientifico» (ROCCO, 2020, ed. originale 1910, pp. 504-505) –, ma piuttosto alla luce delle acquisizioni più recenti, che escludono esiti di chiusura del diritto penale e valorizzano la necessità di un confronto con saperi extragiuridici (già non estranea, per la verità, allo stesso ROCCO, il quale avvertiva: «Distinzione, non è separazione e tanto meno divorzio scientifico»: *ivi*, p. 518). Per tali approdi, v., anche per i riferimenti, DONINI (2010).

²⁴³ Il rilievo esige di essere sottratto a letture riduttive. Numerosi studi sul suicidio – per come riportati da GUIDUCCI (2025), pp. 5 ss. – fanno emergere la tendenza del soggetto a rappresentarsi una forma di esistenza successiva alla morte. In particolare, afferma l'Autore, «Il potenziale suicida, *vivendosi* come *già morto*, può [...] entrare in uno stato di *delirio di onnipotenza* vedendosi finalmente un uomo socialmente riconosciuto, integrato e gratificato *dopo* la morte. La vicina morte reale diventa, così, il ponte fra la morte immaginaria presente e una vita immaginaria futura dopo la morte reale. Si potrebbe, dunque, dire che *il suicida è colui che si toglie la vita per averla, che la perde per trovarla*» (*ivi*, p. 45). In questa prospettiva, si osserva come molti soggetti suicidi si rappresentino la morte quale mezzo per ottenere una vita percepita come mancante o gravemente incompleta: «Il suicida è, quindi, fondamentalmente un escluso che cerca di affermare la propria esistenza. Nel momento in cui si uccide, immagina il proprio funerale, tende a suscitare negli altri quella attenzione che non era riuscito ad ottenere da vivo, riappropriandosi di quell'oggetto d'amore che non aveva potuto raggiungere o aveva perduto» (*ivi*, p. 30). Una conferma di tali dinamiche viene rinvenuta

ma si configura come una forma di *autoesposizione al pericolo (consapevole?)*²⁴⁴. In simili casi, afferma la dottrina, «L'evento finale [in questo caso la morte o l'aggravamento della psicopatologia] alla cui causazione abbia decisamente contribuito la libera condotta della vittima, e che dunque venga a configurarsi come sotto-evento nella catena causale messa in moto dal "primo" agente, non potrà essere imputato a quest'ultimo, che con la sua azione *non ha aumentato il rischio* di verificazione di esso *rispetto alla misura del consentito* e quindi, normalmente, nella prospettiva *ex ante* di costruzione del dovere di diligenza, non avrebbe avuto il *dovere* di riconoscerne il possibile intervento» (corsivo di chi scrive)²⁴⁵. Da tale impostazione si ricava, in via speculare, che l'agente potrà invece essere ritenuto responsabile qualora l'evento – realizzato dalla vittima – costituisca la concretizzazione del rischio da lui originariamente creato.

In questa prospettiva, l'autoesposizione al pericolo dell'utente non vale a interrompere il processo di imputazione nei confronti del fornitore di *chatbot therapists* che non si sia conformato – nei termini sin qui ricostruiti – al contenuto della diligenza esigibile. La successiva condotta dell'utente – che può atteggiarsi in forme di progressiva dipendenza emotiva dai messaggi del *chatbot*, di allontanamento dai circuiti terapeutici tradizionali, di ritiro dal contesto sociale, sino a un prevalente investimento nella dimensione virtuale²⁴⁶ –, pur libera, non si presenta come fattore eccentrico rispetto alla sfera di rischio originariamente aperta dal fornitore (ossia la messa a disposizione, nei confronti di soggetti con disturbi psichici, di sistemi idonei a innescare dinamiche di antropomorfizzazione, di omofilia e di iper-personalizzazione dell'interlocutore artificiale) e da lui non contenuta entro la misura del lecito tramite le pertinenti regole cautelari. Tra queste, ad esempio, nei termini della ricostruzione proposta in tema di evitabilità (v. *supra*, § 4.1): l'adozione di un sistema di gestione dei rischi lungo l'intero ciclo di vita dell'IA, con il coinvolgimento di competenze specialistiche in ambito psicologico e/o psichiatrico; la predisposizione di adeguate misure di *human oversight*; l'assicurazione di un livello di trasparenza e comprensibilità dell'interazione tale da

in epoca antica: Plutarco riferisce dell'epidemia di suicidi tra alcune giovani di Mileto, interrottasi quando fu introdotta la pratica di esporne pubblicamente i corpi nudi. Come chiarisce FORNARI (1970), la psicoanalisi interpreta l'episodio con l'idea che le giovani, rappresentandosi esposte nude dopo la morte, provassero vergogna. Il dato appare significativo nella misura in cui conferma come, anche dinanzi alla scelta suicidaria, il soggetto si rappresenti una qualche forma di esistenza successiva (*ivi*, p. 31). Le dinamiche richiamate possono venire in rilievo anche con riguardo ai suicidi che si sono verificati in concomitanza con l'uso pervasivo dei *chatbots* terapeutici (v. *supra*, § 1). In simili circostanze, possono riscontrarsi situazioni in cui soggetti trovano nell'interazione con tali interlocutori artificiali forme di sostegno affettivo e di riconoscimento che non rinvergono altrove. Possono svilupparsi dinamiche di proiezione che si innestano sulle rappresentazioni post morte sopra richiamate, orientando soggetti non dotati di piena capacità critica – o attraversati dalla crisi suicidaria, ove le facoltà cognitive, come si è riportato *supra*, § 1, sono ridotte – verso l'idea di poter conseguire, in una dimensione successiva, quella condizione esistenziale avvertita come carente o irraggiungibile nella vita reale. Può venire in considerazione l'esempio riportato *supra*, § 1, alla nota n. 32.

²⁴⁴ La formula "autoesposizione al pericolo della vittima" – comunemente ricondotta alla teorizzazione di ROXIN sull'imputazione oggettiva dell'evento (in merito a quest'ultima elaborazione, v. *supra*, § 2.1.1, nota n. 161) – si ricollega alla riflessione sviluppata nell'ambito della *vittimodogmatica* – cfr., TORDINI-CAGLI (2008), spec. pp. 19-20. Tale corrente, muovendo dagli apporti della vittimologia, propone una rilettura della teoria del reato che valorizza il comportamento della vittima quale parametro rilevante ai fini della meritevolezza e del bisogno di tutela, sino a condizionare la legittimazione dell'intervento penale alla possibilità di autotutela del titolare del bene – cfr., *ivi*, pp. 9 ss.

²⁴⁵ FORTI (1990), p. 611.

²⁴⁶ Per la rilevanza di tali fattori rispetto all'aumento del rischio di suicidio, sia consentito un rimando agli studi scientifici in materia riportati *supra*, § 1.

rendere l'utente consapevole della natura e dei limiti del sistema; l'introduzione di limiti alla continuità dell'interazione; l'implementazione di meccanismi di verifica dell'età e di segnalazione dei contenuti inappropriati.

6. Considerazioni finali sulla tutela della vita e dell'incolumità di utenti caratterizzati da disturbi mentali: per un dialogo tra scienze.

Se si riavvolgono i fili dell'indagine, il problema che ne ha segnato l'avvio è se, dinanzi ai rischi posti dall'intelligenza artificiale, il diritto penale possa offrire una tutela effettiva, senza ridursi a risposta simbolica all'inquietudine. Nel caso dei *chatbot therapists*, uno spazio per tale intervento sembra delinearsi: i rischi appaiono, per quanto è stato possibile ricostruire, *riconoscibili*; gli eventi, almeno in parte, *evitabili*.

La conclusione acquista particolare evidenza ove si abbia come riferimento il pericolo concreto per la vita e l'incolumità individuale di utenti con disturbi mentali.

Nel confronto con la legislazione nazionale vigente in materia di IA – e in particolare con l'art. 437-bis, commi 1 e 3, c.p. –, ai fini di un'eventuale responsabilità penale del fornitore di *chatbots* per il supporto psicologico potrebbe venire rilievo, solo all'esito di una lettura audace (ma non preclusa dal dato positivo), l'omessa adozione delle misure di sorveglianza umana: previste dall'AI Act per i soli sistemi di IA ad alto rischio, ma preordinate al governo di rischi tipici che ricorrono anche nell'impiego dei *chatbots* qui considerati. Entro tale esegesi, le risultanze emerse sul piano della diligenza obiettiva consentono di precisare il titolo della responsabilità penale: il *dovere di evitare*, a differenza del *dovere di riconoscere*, si colloca su un piano comune alla colpa e al dolo. La medesima omissione potrebbe pertanto rilevare, ricorrendone i presupposti, a titolo di dolo eventuale ai sensi del comma 1, oltre che di colpa grave ai sensi del comma 3. Del resto, limitando lo sguardo agli agenti conversazionali in esame, la riconoscibilità dei rischi cui tali misure sono rivolte ne confermerebbe la funzione preventiva. Resta aperto, peraltro, il problema degli obblighi di trasparenza di cui all'art. 50 AI Act, già direttamente gravanti sui fornitori di *chatbot therapists*, ma non immediatamente riconducibili né alle misure tecniche di sicurezza né a quelle di sorveglianza umana di cui all'art. 437-bis c.p.

Sul versante dell'incolumità pubblica, invece, la questione della propagazione incontrollabile del pericolo connessa all'impiego crescente, da parte di soggetti vulnerabili, di *chatbot therapists* in sostituzione della terapia psichiatrica rimane consegnata a un orizzonte ancora indefinito.

Resta, nondimeno, che l'affermarsi del *fenomeno* in esame – se sono fondate le premesse da cui si è partiti – si sottrae alla contingenza; mutevoli possono essere, invece, gli interventi normativi volti a ordinarlo. Da qui l'importanza dell'approccio dogmatico, oltre che esegetico, che si è cercato di praticare. Per ciò stesso, lo scenario reclamava e

reclama un’elaborazione teorica all’altezza della sua portata, che non può che affidarsi a voci più mature, espressione di un dibattito cui questo lavoro è ampiamente debitore²⁴⁷.

Permane, tuttavia, in controluce una realtà che il diritto, tutto autonomamente, non è in grado di presidiare: colui che, non trovando nel consorzio umano la cura per il proprio male, muore suicida è «un escluso che cerca di affermare la propria esistenza»²⁴⁸. Una *mancanza* – per riprendere il concetto evocato in apertura – che nemmeno l’IA si è mostrata capace di dominare a vantaggio del bene giuridico della vita.

²⁴⁷ Limitandosi al ricchissimo contesto italiano, sul rapporto tra diritto penale e intelligenza artificiale, cfr.: [MAGRO](#) (2014), pp. 477 ss. (ingl.), pp. 499 ss. (it.); [RIONDATO](#) (2014), pp. 589 ss. (ingl.), pp. 599 ss. (it.); PAGALLO (2017), pp. 643 ss.; [CAPPELLINI](#) (2018-2019), pp. 499 ss.; CONSULICH (2018), pp. 195 ss.; [BASILE](#) (2019), pp. 1 ss.; [BORSARI](#) (2019), pp. 262 ss.; BURCHARD (2019), pp. 1909 ss.; [CAPPELLINI](#) (2019), pp. 325 ss.; MAGRO (2019), pp. 1179 ss.; [PALMISANO](#) (2019), pp. 129 ss.; [LA VATTIATA](#) (2020), pp. 22 ss.; LAGIOIA e SARTOR (2020), pp. 433 ss.; [MAGRO](#) (2020); [MANES](#) (2020); [PIERGALLINI](#) (2020), pp. 1745 ss.; [PREZIOSI](#) (2020), pp. 173 ss.; [SEVERINO](#) (2020), pp. 531 ss.; [CATERINI](#) (2021), pp. 73 ss.; [GIANNINI](#) (2021); [LA VATTIATA](#) (2021), pp. 140 ss.; LO MONTE (2021), pp. 41 ss.; [MAZZACUVA](#) (2021), pp. 143 ss.; [MINELLI](#) (2021), pp. 57 ss.; [MORAITI](#) (2021), pp. 109 ss.; [PANATTONI](#) (2021a), pp. 317 ss.; [PANATTONI](#) (2021b), pp. 125 ss.; [PICOTTI](#) (2021), pp. 813 ss.; [SALVADORI](#) (2021), pp. 83 ss.; [BORGOGNO](#) (2022), pp. 727 ss.; [CONSULICH](#) (2022), pp. 1015 ss.; [CORRIDORI](#) (2022); [D’AMICO](#) (2022), pp. 593 ss.; [GIUCA](#) (2022); [LA VATTIATA](#) (2022), pp. 795 ss.; [MASSI](#) (2022), pp. 665 ss.; [MINELLI](#) (2022), pp. 50 ss.; [PICOTTI](#) (2022a), pp. 67 ss.; [PICOTTI](#) (2022b), pp. 1477 ss.; [PICOTTI](#) (2022c); [PICOTTI](#) (2022d), pp. 111 ss.; [PIVA](#) (2022), pp. 681 ss.; [PREZIOSI](#) (2022), pp. 713 ss.; [SEVERINO](#) (2022), pp. 83 ss.; [TRIPODI](#) (2022), pp. 745 ss.; [TRONCONE](#) (2022), pp. 3287 ss.; [AMORE](#) (2023), pp. 101 ss.; [BADALAMENTI](#) (2023), pp. 1423 ss.; [BARRESI](#) (2023), pp. 269 ss.; [BASILE](#) (2023), pp. 1 ss.; [BASILE e FRAGASSO](#) (2023), pp. 752 ss.; [CAPPELLINI](#) (2023), pp. 19 ss.; [COLACURCI](#) (2023a), pp. 119 ss.; [COLACURCI](#) (2023b), pp. 353 ss.; [CONSULICH](#) (2023), pp. 721 ss.; [CONSULICH et al.](#) (2023); [CORRIDORI](#) (2023), pp. 1394 ss.; [CUPELLI](#) (2023); [D’AMICO](#) (2023); [DE SIMONE](#) (2023), pp. 51 ss.; [ESPOSITO](#) (2023), pp. 37 ss.; [FLOR](#) (2023), p. 788; [FRAGASSO](#) (2023), pp. 30 ss.; [GIANNINI](#) (2023a); [GIANNINI](#) (2023b), pp. 43 ss.; [GIANNINI](#) (2023c), pp. 1437; [GIUCA](#) (2023), pp. 46 ss.; [GIUSTI](#) (2023), pp. 1356 ss.; [GRANDI](#) (2023), pp. 67 ss.; [LANZI](#) (2023a); [LANZI](#) (2023b), pp. 1379 ss.; [MAGRO](#) (2023), pp. 1232; [MASSARO](#) (2023), pp. 27 ss.; [MONGILLO](#) (2023), pp. 77 ss.; [PANATTONI](#) (2023), pp. 269 ss.; [PAPA e TAVELLA](#) (2023); [PICOTTI](#) (2023a), pp. 11 ss.; [PICOTTI](#) (2023b), pp. 255 ss.; [ROMANÒ](#) (2023); [SALCUNI e DI FLORIO](#) (2023), pp. 1403 ss.; [TERRIZZI](#) (2023); [TRIPODI](#) (2023); [VADALÀ](#) (2023); [BARONE](#) (2024), pp. 1047 ss.; [CANATO](#) (2024); [CARRIERO](#) (2024); [COMPOSTELLA](#) (2024); [CONSULICH](#) (2024), pp. 237 ss.; [DE FLAMMINEIS](#) (2024), pp. 5 ss.; [FLORIO](#) (2024), pp. 5 ss.; [FRAGASSO](#) (2024), pp. 69 ss.; [LOMBARDI](#) (2024), pp. 57 ss.; [MANGIONE](#) (2024); [MONGILLO](#) (2024); [PANATTONI](#) (2024), pp. 525 ss.; [PICOTTI](#) (2024a), pp. 83 ss.; [PICOTTI](#) (2024b); [PICOTTI](#) (2024c), pp. 293 ss.; [ROMANO](#) (2024), pp. 403 ss.; [CANATO](#) (2025), pp. 61 ss.; [DA ROLD](#) (2025), pp. 351 ss.; [D’AMICO](#) (2025); [D’ANGELO](#) (2025); [FEDERICI](#) (2025), pp. 116 ss.; [FRAGASSO](#) (2025a), pp. 157 ss.; [FRAGASSO](#) (2025b); [LANZI](#) (2025); [MATTARELLA](#) (2025), pp. 250 ss.; [MUCCIARELLI](#) (2025); [PANATTONI](#) (2025); [PANEBIANCO](#) (2025), pp. 395 ss.; [PICOTTI](#) (2025); [PIETRELLA](#) (2025), pp. 101 ss.; [SALVADORI](#) (2025); [CALETTI](#) (2026), pp. 199 ss.; [GIANNINI](#) (2026), pp. 1005 ss.; [ONOFRI](#) (2026); [PAPA](#) (2026), pp. 1029 ss.

²⁴⁸ GUIDUCCI (2025), p. 30.

Una frattura che sollecita l'insieme e l'incontro dei saperi, nonché gli stessi consociati che ne sono portatori ²⁴⁹; «Ma questa, per l'appunto, è una storia in gran parte ancora da scrivere»²⁵⁰.

Bibliografia

AA. VV., in CADOPPI, Alberto, CANESTRARI, Stefano, CONTI, Carlotta, MANNA, Adelmo, PAPA, Michele (eds.) (2026): *Tutela penale delle persone vulnerabili* (Milano, Giuffrè)

ABRAMS, Zara (2025): "Using generic AI chatbots for mental health support: A dangerous trend", *American Psychological Association Services Inc.*, 12 marzo 2025, disponibile su: www.apaservices.org (URL verificato il 30 settembre 2026)

AGENZIA PER L'ITALIA DIGITALE (2018): *Libro bianco sull'intelligenza artificiale al servizio del cittadino* (Roma, AgID).

AMERICAN ASSOCIATION OF SUICIDOLOGY (2023): *Know the signs: how to tell if someone might be suicidal*, disponibile su <https://suicidology.org> (URL verificato il 30 settembre 2026)

²⁴⁹ Questa risulta la visione delineata anche dalla letteratura più accreditata in ambito psichiatrico: «La nuova prospettiva, orientata alla prevenzione del suicidio, richiede uno sforzo alla società intera, affinché i membri che la compongono sostengano in maniera attiva gli individui in crisi» – POMPILI (2022), p. 22. Non senza un'eco – che sembra restituirne intatta l'attualità – degli studi di DURKHEIM. Solo un breve passaggio di *Le Suicide*: «Esiste tutta una serie di funzioni che interessano solo l'individuo; sono quelle necessarie a mantenere la vita fisica. Poiché sono destinate unicamente a questo scopo, una volta che l'hanno raggiunto non debbono far di più. Di conseguenza, per tutto quel che le riguarda, l'uomo può agire ragionevolmente senza che abbia a proporsi fini che lo superano [...] Questo non è però il caso dell'uomo civilizzato che ha raggiunto l'età adulta. In lui esiste una quantità di idee, di sentimenti, di esperienze che nulla hanno a che vedere con le necessità organiche. L'arte, la morale, la religione, la fede politica, la scienza stessa non hanno il compito di rimediare all'usura degli organi né di mantenerne il buon funzionamento. Né sono le sollecitazioni dell'ambiente cosmico a svegliare e sviluppare tutta questa vita extra-fisica, ma quelle dell'ambiente sociale. È l'azione della società che ha suscitato in noi questi sentimenti di simpatia e di solidarietà che ci predispongono verso gli altri; è lei che, modellandoci a sua immagine, ci ha impregnati di quelle credenze religiose, politiche, morali che formano la nostra condotta. È per poter svolgere il nostro ruolo sociale che abbiamo lavorato a coltivare la nostra intelligenza ed è sempre la Società che, trasmettendoci la scienza di cui è depositaria, ci ha fornito gli strumenti di questo sviluppo. Se essa invece si disgrega, se non si fa più sentire viva e attiva attorno e sopra di noi, quanto vi è di sociale in noi si trova sprovvisto di ogni fondamento obiettivo. Ne risulta che le ragioni di vivere ci mancano, perché l'unica vita per la quale possiamo aver interesse non corrisponde più alla realtà e l'unica che sia ancora fondata sul reale non risponde più ai nostri bisogni. *Iniziati ad una esistenza più elevata, non ci può più soddisfare quella di cui si accontentano i bambini e gli animali, e allora ci sfugge anche quella, lasciandoci sgomenti*» – DURKHEIM (ed. it. 2025), pp. 303-305 (corsivo di chi scrive).

²⁵⁰ L'espressione è di CASTRONUOVO (2011), p. 1645, a conclusione di una profonda riflessione sulle trasformazioni della colpa penale nella realtà tecnico-sociale del rischio. Lì si segnala come tali sviluppi rendano necessario il confronto tra dottrina, giurisprudenza e saperi extragiuridici. Un'esigenza che, nella sua portata di fondo, si lascia cogliere, giunti alle conclusioni, anche rispetto al tema che qui si è cercato di offrire in esame.

AMERICAN PSYCHOLOGICAL ASSOCIATION SERVICES INC. (2025): *APA Health Advisory on the Use of Generative AI Chatbots and Wellness Applications for Mental Health*, disponibile su www.apa.org (URL verificato il 30 settembre 2026)

AMORE, Nicolò (2023): "Parte seconda: l'effetto della robotica e dell'IA nell'imputazione giuridica degli eventi infausti", in AMORE, Nicolò e ROSSERO, Eleonora, *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Un'indagine sociologica e giuridico-penale* (Bologna, Il Mulino), pp. 101 ss.

AMORE, Nicolò e ROSSERO, Eleonora (2023): *Robotica e intelligenza artificiale nell'attività medica. Organizzazione, autonomia, responsabilità. Un'indagine sociologica e giuridico-penale* (Bologna, Il Mulino)

ANDLER, Daniel (2023): *Intelligence artificielle, intelligence humaine: la double énigme* (Parigi, Gallimard)

ANDREAS, Michael (2016): *La colpa in attività illecita. Tesi di dottorato di ricerca in scienze giuridiche. Tutore Prof. Vallini Antonio* (Firenze, Università degli Studi di Firenze)

ANTOLISEI, Francesco (1948): "La colpa per inosservanza di leggi", *Giustizia Penale*, II, p. 8

BADALAMENTI, Valentina (2023): "La responsabilità penale del sanitario alla luce dell'evoluzione tecnologica", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed., (Milano, UTET), pp. 1423 ss.

BARLOW, David H. (2010): "Negative Effects From Psychological Treatments", *American Psychologist*, vol. 65, n. 1, pp. 13 ss.

BARONE, Giulia (2024): "La regolamentazione dell'Intelligenza Artificiale: è 'corsa agli armamenti'", *Diritto penale e processo*, n. 8, pp. 991 ss.

BARRESI, Olimpia (2023): "Italian Report on Traditional Criminal Law Categories and AI", in PICOTTI, Lorenzo e PANATTONI, Beatrice (eds.), *Traditional Criminal Law and AI: Crisis or Palingenesis, International Colloquium Section I, Siracusa, 15-16 September 2022, Revue Internationale de Droit Pénal*, 94, 1, pp. 269 ss.

BARTOLI, Roberto (2010): "'Colpa' in attività illecita: un discorso ancora da sviluppare!", *Diritto Penale e Processo*, n. 16, p. 1049

BASILE, Fabio (2005): *La colpa in attività illecita. Un'indagine di diritto comparato sul superamento della responsabilità penale* (Milano, Giuffrè)

BASILE, Fabio (2011): "L'alternativa tra responsabilità oggettiva e colpa in attività illecita per l'imputazione della conseguenza ulteriore non voluta, alla luce della sentenza Ronci delle Sezioni Unite sull'art. 586 c.p.", *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 938-939

BASILE, Fabio (2019): "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine", in *Diritto Penale e Uomo*, fasc. 10, pp. 1 ss.

BASILE, Fabio (2023): “Intelligenza artificiale e diritto penale: qualche aggiornamento e qualche nuova riflessione”, in BALBI, Giuliano, DE SIMONE, Federica, ESPOSITO, Andrea, MANACORDA, Stefano (eds.), *Diritto penale e intelligenza artificiale. “Nuovi Scenari”*, (Milano, Giappichelli), pp. 1 ss.

BASILE, Fabio e FRAGASSO, Beatrice (2023): “Intelligenza artificiale e diritto penale: prove tecniche di convivenza”, in *Diritto di Difesa*, fasc. 3-4, pp. 752 ss.

BATHAEE, Yavar (2018): “The Artificial Intelligence Black Box and the Failure of Intent and Causation”, in *Harvard Journal of Law & Technology*, 31, 2, pp. 907-908

BAUMAN, Zygmunt (2006, ed. it. 2009): *Paura liquida*, traduzione di Marco Cupellaro (Roma-Bari, Laterza), p. 69

BECCAGUTTI, Guido (2025): *Audizione informale, XII Commissione Affari Sociali, Camera dei Deputati. Proposte di legge A.C.1208 Loizzo et al. e A.C.2095 Quartini et al. “Disposizioni in materia di terapie digitali”*, 21 gennaio, disponibile su: <https://documenti.camera.it> (URL verificato il 30 settembre 2026)

BECK, Susanne (2016): “Intelligent agents and criminal law-Negligence, diffusion of liability and electronic personhood”, *Robotics and Autonomous Systems*, 86, pp. 138 ss.

BECK, Ulrich (1986, ed. it. 2000): *La società del rischio. Verso una seconda modernità* (trad. Walter Privitera, Carlo Sandrelli, Roma, Carocci)

BÉNÉJAT, Murielle (2019): “Le droit pénal de la route face aux nouveaux modes de transport”, *Actualité juridique Pénal*, vol. 9, pp. 428 ss.

BERNARDELLI, Luca (2025): “Isolamento sociale e adolescenza: rischi psicosociali delle interazioni con i chatbot”, in *Rivista di Pediatria*, n. 11-12, pp. 22-23

BERNARDI, Alessandro (2003): “La responsabilità da prodotto nel sistema italiano: profili sanzionatori”, in *Rivista trimestrale di diritto penale dell'economia*, pp. 1 ss.

BLANCO CORDERO, Isidoro (2025): “Vehículos autónomos y responsabilidad penal: retos para la imputación de resultados lesivos”, *Revista General de Derecho Penal*, n. 44

BORGOGNO, Roberto (2022): “La responsabilità penale nei processi ad elevata automazione”, in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (eds.), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), pp. 727 ss.

BORSARI, Riccardo (2019): “Intelligenza Artificiale e responsabilità penale: prime considerazioni”, *MediaLaws. Rivista di diritto dei media*, n. 3, pp. 262 ss.

BOUCHER, Eliane M., HARAKE, Nicole R., WARD, Haley E., STOECKL, Sarah Elizabeth, VARGAS, Junielly, MINKEL, Jared, PARKS, Acacia C. e ZILCA, Ran (2021): “Artificially intelligent chatbots in digital mental health interventions: a review”, *Expert Review of Medical Devices*, vol. 18, pp. 37-49

BRAUNER, Philipp, GLAWE, Felix, LIEHNER, Gian Luca, VERVIER, Luisa, ZIEFLE, Martina (2026): “Charting cross-cultural AI perception: An exploratory study of

expectations, risk–benefit trade-offs, and values in Germany and China”, *Acta Psychologica*, vol. 268

BRICOLA, Franco (1978): “Responsabilità penale per il tipo e per il modo di produzione”, in AA. VV., *La responsabilità dell’impresa per i danni all’ambiente e ai consumatori* (Milano, Giuffrè), pp. 75 ss.

BRUNELLI, David (2023): *Riflessioni sulla condotta nel reato omissivo improprio* (Pisa, Pisa University Press)

BURCHARD, Christoph (2019): “L’intelligenza artificiale come fine del diritto penale? Sulla trasformazione algoritmica della società”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 4, pp. 1909 ss.

BUSSOLATI, Nicolò (2026): “La vulnerabilità nella realtà virtuale”, in CADOPPI, Alberto, CANESTRARI, Stefano, CONTI, Carlotta, MANNA, Adelmo, PAPA, Michele (eds.), *Tutela penale delle persone vulnerabili* (Milano, Giuffrè), pp. 993 ss.

CALAMANDREI, Piero (1939): “Il giudice e lo storico”, *Rivista di diritto processuale civile*, vol. XVI – Parte I, p. 110

CALETTI, Gian Marco (2026): “Diffusione illecita di immagini o video sessualmente espliciti (art. 612-ter c.p.) e di contenuti generati o alterati con sistemi di intelligenza artificiale (art. 612-quater c.p.)”, in CADOPPI, Alberto, CANESTRARI, Stefano, CONTI, Carlotta, MANNA, Adelmo, PAPA, Michele (eds.), *Tutela penale delle persone vulnerabili* (Milano, Giuffrè), pp. 199 ss.

CANATO, Maria Carla (2024): “Verso il superamento del ‘legal risk’ europeo: intelligenza artificiale e approccio proporzionale al rischio”, *La Legislazione Penale*, 31 luglio

CANATO, Maria Carla (2025): “La responsabilità da reato degli enti di fronte al ‘rischio da intelligenza artificiale’: responsibility gap e ‘rinnovata’ corporate liability”, in LAMBRINI, Paola (direzione), *Quaderni del dottorato in giurisprudenza dell’Università di Padova 2024*, pp. 61 ss.

CANESTRARI, Stefano (2023): “Osservazioni su colpa in attività illecita, omicidio preterintenzionale e art. 586 c.p.”, in *La riforma dei delitti contro la persona. Proposte dei gruppi di lavoro dell’AIPDP* (Milano, DiPLaP), pp. 89 ss.

CAPPELLINI, Alberto (2018): “Machina delinquere non potest? Brevi appunti su intelligenza artificiale e responsabilità penale”, *Criminalia. Annuario di scienze penalistiche*, pp. 499 ss.; anche su *DisCrimen*, 27 marzo 2019

CAPPELLINI, Alberto (2019): “Profili penalistici delle self-driving cars”, *Diritto Penale Contemporaneo – Rivista trimestrale*, 2, pp. 325 ss.

CAPPELLINI, Alberto (2023): “Reati colposi e tecnologie dell’intelligenza artificiale”, in BALBI, Giuliano, DE SIMONE, Federica, ESPOSITO, Andrea, MANACORDA, Stefano (eds.),

Diritto penale e intelligenza artificiale. “Nuovi Scenari” (Milano, Giappichelli), pp. 19 ss.; anticipato in *Archivio Penale*, 2022, fasc. 3

CARMONA, Angelo (2001): “Il *versari in re illicita* ‘colposo’. Un breve percorso tra pratiche giurisprudenziali e suggestioni dogmatiche, pensando alla riforma del codice penale”, *Indice penale*, pp. 237-238

CARRIERO, Maria Federica (2023): “L’ingerenz-theorie alla luce di nuove (e più sofisticate) derive giurisprudenziali”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 957 ss.

CARRIERO, Maria Federica (2024): “Crisi della legalità e intelligenza artificiale: uno sguardo (comparato) d’insieme”, *Archivio Penale*, fasc. 2

CASTALDO, Andrea R. (1989): *L’imputazione oggettiva nel delitto colposo d’evento* (Napoli, Jovene), pp. 59 ss.

CASTRONOVO, Carlo (2024): *Dogmatica del diritto* (Milano, Giuffrè)

CASTRONOVO, Carlo (2026): *Dogmatica del diritto – Lectio Magistralis prof. Castronovo*, Università degli Studi di Trento, 12 marzo, online: <https://www.youtube.com/watch?v=6kM0qIQfY70> (min. 13:12 ss.)

CASTRONUOVO, Donato (2005): “Responsabilità da prodotto e struttura del fatto colposo”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 1, pp. 301 ss.

CASTRONUOVO, Donato (2009): *La colpa penale* (Milano, Giuffrè)

CASTRONUOVO, Donato (2011): “L’evoluzione teorica della colpa penale tra dottrina e giurisprudenza”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 4, pp. 1594 ss.

CASTRONUOVO, Donato (2012): *Principio di precauzione e diritto penale. Paradigmi dell’incertezza nella struttura del reato* (Roma, Aracne)

CASTRONUOVO, Donato (2021): “Colpa penale”, in DONINI, Massimo (ed.), *Enciclopedia del diritto – I Tematici: Reato colposo* (Milano, Giuffrè), pp. 200 ss.

CASTRONUOVO, Donato (2024): “La tutela dell’incolumità pubblica. Profili storici, sistematici e critici”, *La Legislazione Penale*, 27 novembre

CASTRONUOVO, Donato (2025): “Gestione “organizzata” del rischio di calamità. Colpa ‘distale’ e ‘prossimale’ alla luce della sentenza della Cassazione sul disastro di Rigopiano”, *Archivio Penale*, fasc. 3

CATERINI, Mario (2021): “Alla ricerca di una certezza ‘artificiale’ del sistema penale”, in BASILE, Fabio, CATERINI, Mario, ROMANO, Sabato, *Il sistema penale ai confini delle hard sciences* (Pisa, Pacini Giuridica), pp. 73 ss.

CHAUDHRY, Beenish Moalla e DEBI, Happy Rani (2024): “User perceptions and experiences of an AI-driven conversational agent for mental health support”, *mHealth*, vol. 10, art. 22

CIVELLO, Gabriele (2021): “Prevedibilità e reato colposo”, in DONINI, Massimo (ed.), *Enciclopedia del diritto – I Tematici: Reato colposo* (Milano, Giuffrè), pp. 1004 ss.

CIVELLO, Gabriele (2023): “Dolo eventuale senza accettazione dell’evento: per la Corte di cassazione è sufficiente la prevedibilità secondo il ‘normale bagaglio di conoscenze dell’uomo medio?’”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 1097 ss.

COECKELBERGH, Mark (2020): “Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability”, *Science and Engineering Ethics*, 26, p. 2052

COGLAN, Simon, LEINS, Kobi, SHELDRICK, Susie, CHEONG, Marc, GOODING, Piers, D’ALFONSO, Simon (2023): “To chat or bot to chat: Ethical issues with using chatbots in mental health”, *Digital Health*, vol. 9

COLACURCI, Marco (2023a): “Riconoscimento facciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini”, in BALBI, Giuliano, DE SIMONE, Federica, ESPOSITO, Andrea, MANACORDA, Stefano (eds.), *Diritto penale e intelligenza artificiale. “Nuovi Scenari”* (Milano, Giappichelli), pp. 119 ss.

COLACURCI, Marco (2023b): “Quale diritto penale dell’IA? Alcune riflessioni a partire dalla proposta di regolamento dell’Unione Europea”, *JUS*, 3, pp. 353 ss.

COMMISSIONE EUROPEA (2021): *Commission Staff Working Document – Impact Assessment Accompanying the Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act)*, SWD (2021) 84 final (Bruxelles), 21 aprile (spec. Annex 2)

COMMISSIONE EUROPEA (2025): *Analysis of stakeholder feedback on AI definitions and prohibited practices* (Bruxelles), 12 maggio

COMMISSIONE EUROPEA (2025): *Guidelines on Prohibited Practices under the Artificial Intelligence Act* (Bruxelles), 4 febbraio

COMPOSTELLA, Roberto (2024): *Auto a guida autonoma e diritto penale. Profili di responsabilità individuale e collettiva* (Napoli, Editoriale Scientifica)

CONTRI, Francesco (2025): “Nesso di rischio ‘prevalente’ e delitti contro l’incolumità pubblica. Un’analisi critica del diritto vivente”, *La Legislazione Penale*, 19 aprile

CONSORTE, Francesca (2013): *Tutela penale e principio di precauzione. Profili attuali, problematicità, possibili sviluppi* (Torino, Giappichelli)

CONSULICH, Federico (2018): “Il nastro di Möbius. Intelligenza artificiale e imputazione penale nelle nuove forme di abuso del mercato”, *Banca, borsa e titoli di credito*, fasc. 2, pp. 195 ss.

CONSULICH, Federico (2022): “Flash offenders. Le prospettive di accountability penale nel contrasto alle intelligenze artificiali devianti”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 1015 ss.

CONSULICH, Federico (2023): “Il diritto penale al tempo dell’intelligenza artificiale. Prospettive punitive nazionali dopo l’AI Act”, *Diritto di Difesa*, fasc. 3-4, pp. 721 ss.

CONSULICH, Federico (2024): “Intelligenza artificiale e reati finanziari”, in ID. (ed.), *Reati in materia bancaria e finanziaria*, in PALAZZO, Francesco, PALIERO, Carlo Enrico, PELLISSERO, Marco, *Trattato teorico-pratico di diritto penale* (Torino, Giappichelli), pp. 237 ss.

CONSULICH, Federico, MAUGERI, Marco, MILIA, Carlo, POLI, Tommaso Nicola, TROVATORE, Gianfranco (2023): “AI e abusi di mercato: le leggi della robotica si applicano alle operazioni finanziarie?”, *Quaderni giuridici CONSOB*, 29 maggio

COONS, Christian e WEBER, Michael (2014): “Introduction: Investigating the Core Concept and Its Moral Status”, in COONS, Christian e WEBER, Michael (eds.), *Manipulation. Theory and Practice* (New York, Oxford University Press), p. 14

CORN, Emanuele (2013): *Il principio di precauzione nel diritto penale. Studio sui limiti all’anticipazione della tutela penale* (Torino, Giappichelli)

CORRIDORI, Costanza (2022): “Machina delinquere non potest”, *Giustizia Insieme*, 19 maggio

CORRIDORI, Costanza (2023): “L’intelligenza artificiale come vittima del reato”, in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed. (Milano, UTET), pp. 1394 ss.

CRISTEA, Ioana Alina, SUCALA, Madalina, DAVID, Daniel (2013): “Can you tell the difference? Comparing face-to-face versus computer-based interventions. The ‘Eliza’ effect in psychotherapy”, *Journal of Cognitive and Behavioral Psychotherapies*, vol. 13, n. 2, pp. 291-298

CUPELLI, Cristiano (2016): “La colpa dello psichiatra. Rischi e responsabilità tra poteri impositivi, regole cautelari e linee guida”, *Diritto Penale Contemporaneo*, 21 marzo

CUPELLI, Cristiano (2017): “La responsabilità penale dello psichiatra: nuovi spunti, diverse prospettive, timide aperture”, *Diritto penale e processo*, n. 3, pp. 370 ss.

CUPELLI, Cristiano (2023): “La sfida dell’intelligenza artificiale al diritto penale”, *Sistema penale*, 21 aprile

D’ALESSANDRO, Francesco (2021): “Precauzione (principio di)”, in DONINI, Massimo (ed.), *Enciclopedia del diritto – I Tematici: Reato colposo* (Milano, Giuffrè), pp. 975 ss.

D’ALFONSO, Simon (2020): “AI in mental health”, *Current Opinion in Psychology*, vol. 36, pp. 112 ss.

D’AMICO, Laura (2022): “Intelligenza artificiale e auto a guida autonoma. Tra prevenzione primaria, colpa penale e rischio consentito”, *Rivista Italiana di Medicina Legale e del Diritto in Campo Sanitario*, pp. 593 ss.

D’AMICO, Laura (2023): “Colpa, precauzione e rischio. Le tensioni penalistiche nella moderna era tecnologica”, *La Legislazione Penale*, 21 ottobre

D’AMICO, Laura (2025): *La misura della (im)prevedibilità. Modelli d’imputazione della responsabilità al tempo dell’intelligenza artificiale* (Napoli, Edizioni Scientifiche Italiane)

D'ANGELO, Marina Poggi (2025): "Condotte manipolative, High Frequency Trading e intelligenza artificiale: possibili rimedi all'ossimoro della circostanza aggravante 'attenuante'", *DisCrimen*, 27 ottobre

DA ROLD, Matthias (2025): "Diritto penale e intelligenza artificiale: il problema della colpa", in AMOROSO, Daniele e DEFFENU, Andrea (eds.), *Intelligenza artificiale, dati e diritto* (Bari, Carucci Editore), pp. 351 ss.

DE FLAMMINEIS, Siro (2024): "Fattispecie penali nel contesto dell'intelligenza artificiale. Lo spunto del D.D.L. 1146/2024", *Sistema Penale*, 9, pp. 5 ss.

DE FREITAS, Julian, UĞURALP, Ahmet K., UĞURALP, Zeliha O., PUNTONI, Stefano (2024): "AI Companions Reduce Loneliness", *Harvard Business School Working Paper*, n. 24-078

DE FREITAS, Julian, UĞURALP, Zeliha O., UĞURALP, Ahmet K. (2025): "Emotional Manipulation by AI Companions", *Harvard Business School Working Paper*, n. 26-005, pp. 3 ss.

DE LEONARDIS, Francesco (2005): *Il principio di precauzione nell'amministrazione del rischio* (Milano, Giuffrè)

DE SIMONE, Federica (2023): "L'implementazione delle nuove tecnologie nelle politiche anticorruzione", in BALBI, Giuliano, DE SIMONE, Federica, ESPOSITO, Andrea, MANACORDA, Stefano (eds.), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"* (Milano, Giappichelli), pp. 51 ss.

DONG, Mengchen, CONWAY, Jane Rebecca, BONNEFON, Jean-François, SHARIFF, Azim, RAHWAN, Iyad (2026): "Fears About Artificial Intelligence Across 20 Countries and Six Domains of Application", *American Psychologist*, vol. 81, n. 1, pp. 53 ss.

DONINI, Massimo (1991): *Illecito e colpevolezza nell'imputazione del reato* (Milano, Giuffrè)

DONINI, Massimo (2004): *Il volto attuale dell'illecito penale. La democrazia penale tra differenziazione e sussidiarietà* (Milano, Giuffrè), p. 121

DONINI, Massimo (2006): *Imputazione oggettiva dell'evento. «Nesso di rischio» e responsabilità per fatto proprio* (Torino, Giappichelli)

DONINI, Massimo (2010): "Tecnicismo giuridico e scienza penale cent'anni dopo. La Prolusione di Arturo Rocco (1910) nell'età dell'europeismo giudiziario", *Criminalia. Annuario di scienze penalistiche*, p. 167

DONINI, Massimo (2013): "L'elemento soggettivo della colpa. Garanzie e sistematica", *Rivista Italiana di Diritto e Procedura Penale*, fasc. 1, pp. 127 ss.

DOUGLAS, Mary (1992, ed. it. 1996): *Rischio e colpa*, traduzione di Giovanna Bettini (Bologna, Il Mulino)

DOUGLAS, Thomas (2008): "Moral Enhancement", *Journal of Applied Philosophy*, vol. 25, n. 3, pp. 228 ss.

- DOUMA, Frank e PALODICHUK, Sarah Aue (2012): "Criminal Liability Issues Created by Autonomous Vehicles", *Santa Clara Law Review*, 52(4), pp. 1157 ss.
- DURKHEIM, Émile (1893, ed. fr. 2004): *De la division du travail social* (Paris, PUF), p. 76
- DURKHEIM, Émile (1897, ed. it. 2025): *Il suicidio*, traduzione di Rosantonietta Scramaglia (Milano, Rizzoli)
- ELLYSON, Laura (2018): "La responsabilité criminelle et l'intelligence artificielle : quelques pistes de réflexion", *Les Cahiers de propriété intellectuelle*, vol. 30, n. 3, pp. 879 ss.
- ENGISCH, Karl (1930): *Untersuchungen über Vorsatz und Fahrlässigkeit im Strafrecht* (Berlin, Scientia), pp. 271-273
- ESPOSITO, Andreana (2023): "Note sparse sull'intelligenza artificiale", in BALBI, Giuliano, DE SIMONE, Federica, ESPOSITO, Andrea, MANACORDA, Stefano (eds.), *Diritto penale e intelligenza artificiale. "Nuovi Scenari"*, (Milano, Giappichelli), pp. 37 ss.
- EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (2021): *Preparare un giusto futuro. L'intelligenza artificiale e i diritti fondamentali – Sintesi* (Lussemburgo, Ufficio delle pubblicazioni dell'Unione europea)
- EUSEBI, Luciano (1993): *Il dolo come volontà* (Brescia, Morcelliana)
- FEDERICI, Dalila (2025): "La tutela delle funzioni dell'Autorità di vigilanza e di notifica nel quadro normativo delineato dal DDL sull'intelligenza artificiale", *Nullum Crimen*, fasc. 2, pp. 116 ss.
- FIANDACA, Giovanni e MUSCO, Enzo (2023): *Diritto penale. Parte generale*, 8ª ed. (Bologna, Zanichelli)
- FLOR, Roberto (2023): "Cybersecurity for Artificial Intelligence e diritto penale: prime riflessioni nel prisma del diritto europeo", in LUCHTMAN, Michiel, DE JONG, Ferry, KRISTEN, François, LIGETI, Katalin, LINDEMAN, Joep, TOSZA, Stanisław, WIDDERSHOVEN, Rob, ZAITCH, Damián (eds.), *Of swords and shields: due process and crime control in times of globalization – Liber amicorum prof. dr. J.A.E. Vervaele* (The Hague – L'AIA, Eleven International Publishing), pp. 785 ss.
- FLORIDI, Luciano (2015): *The Onlife Manifesto. Being Human in a Hyperconnected Era* (London, Springer)
- FLORIDI, Luciano (2022): *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide* (Milano, Raffaello Cortina)
- FLORIO, Marco Edgardo (2024): "Il dibattito sulla responsabilità penale diretta delle IA: 'molto rumore per nulla'?", *Sistema Penale*, 2, pp. 5 ss.
- FORNARI, Franco (1970): "Nota sulla psicoanalisi del suicidio", in *Nuovi orientamenti nella psicoanalisi* (Milano, Feltrinelli)
- FORNASARI, Gabriele (1990): *Il principio di inesigibilità nel diritto penale* (Padova, Cedam), pp. 334 ss.

- FORNASARI, Gabriele (1993): *I principi del diritto penale tedesco* (Padova, Cedam), pp. 220 ss.
- FORTI, Gabrio (1990): *Colpa ed evento nel diritto penale* (Milano, Giuffrè)
- FORTI, Gabrio (2006): “Per una discussione sui limiti morali del diritto penale, tra visioni ‘liberali’ e paternalismi giuridici”, in DOLCINI, Emilio, PALIERO, Carlo Enrico (eds.), *Studi in onore di Giorgio Marinucci*, I (Milano, Giuffrè), p. 331.
- FOUCAULT, Michel (1975, ed. it. 2014): *Sorvegliare e punire. Nascita della prigione* (Torino, Einaudi)
- FRAGASSO, Beatrice (2023): “La responsabilità penale del produttore di sistemi di intelligenza artificiale”, *Diritto Penale Contemporaneo – Rivista trimestrale*, 1, pp. 26 ss.
- FRAGASSO, Beatrice (2024): “Crimes of the future: Assessing causation in AI-related crimes”, *Revista Portuguesa de Ciência Criminal*, vol. 34, 1, pp. 69 ss.
- FRAGASSO, Beatrice (2025a): “Profili penalistici della legge sull’intelligenza artificiale: osservazioni a prima lettura”, *Sistema Penale*, 10, pp. 157 ss.
- FRAGASSO, Beatrice (2025b): *Intelligenza artificiale e responsabilità penale. Principi e categorie alla prova di una tecnologia “imprevedibile”* (Torino, Giappichelli)
- FRAGASSO, Beatrice (2026): “Disposizioni attuative in materia di intelligenza artificiale: il reato di omessa adozione di misure di sicurezza nei sistemi di I.A. e altre novità in materia di autorità di vigilanza, identificazione biometrica, responsabilità da reato degli enti e responsabilità civile”, *Sistema Penale*, 12 giugno
- FRANCES, Allen (2026): “Warning: AI chatbots will soon dominate psychotherapy”, *The British Journal of Psychiatry*, pp. 474 ss.
- FUSELLI, Stefano (2021): “Neurocorrection. On the use of neurodevices for criminals”, in *Teoria e Critica della Regolazione Sociale*, 1, (Milano, Mimeis), pp. 209 ss.
- GALIMBERTI, Umberto (1999): *Psiche e techne. L’uomo nell’età della tecnica* (Milano, Feltrinelli)
- GARANTE PER LA PROTEZIONE DEI DATI PERSONALI (2025): *Potere e responsabilità. La cultura della protezione dei dati. Relazione annuale 2024*, Roma, Camera dei deputati, 15 luglio 2025, disponibile su www.garanteprivacy.it (URL verificato il 30 settembre 2026)
- GARGANI, Alberto (2020): “Delitti di pericolo personale e individuale. Osservazioni in prospettiva di riforma”, *La Legislazione Penale*, 9 settembre
- GARGANI, Alberto, ZIRULIA, Stefano, CASTRONUOVO, Donato (2023): “Tutela della vita e della salute (nei settori della sicurezza del lavoro, degli alimenti, dei farmaci, etc.)”, in *Associazione Italiana dei Professori di Diritto Penale e Laboratorio Permanente di Diritto e Procedura Penale* (eds.), *La riforma dei delitti contro la persona. Proposte dei gruppi di lavoro dell’AIPDP. Atti dei seminari di discussione in collaborazione con il DiPLaP*, (Milano, DiPLaP Editor), pp. 801 ss.

- GARRAUD, René (1924): *Traité Théorique et pratique du Droit Pénal Français* (Paris, Sirey), p. 405
- GEHLEN, Arnold (1984, ed. it. 2003): *L'uomo nell'era della tecnica*, a cura di Maria Teresa Pansera, (Roma, Armando), pp. 31 ss.
- GIANNINI, Alice (2021): "Intelligenza artificiale, *Human Oversight* e responsabilità penale: prove d'impatto a livello europeo", *Criminalia. Annuario di scienze penalistiche*, pp. 249 ss.
- GIANNINI, Alice (2023a): *Criminal Behavior and Accountability of Artificial Intelligence Systems* (The Hague, Eleven International Publishing)
- GIANNINI, Alice (2023b): "Negligence failures and negligence fixes. A comparative analysis of criminal regulation of AI and autonomous vehicles", *Criminal Law Forum*, vol. 34, pp. 43 ss.
- GIANNINI, Alice (2023c): "Responsabilità da reato degli enti e intelligenza artificiale", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed. (Milano, UTET), pp. 1437 ss.
- GIANNINI, Alice (2026): "L'intelligenza artificiale come entità vulnerabile", in CADOPPI, Alberto, CANESTRARI, Stefano, CONTI, Carlotta, MANNA, Adelmo, PAPA, Michele (eds.), *Tutela penale delle persone vulnerabili* (Milano, Giuffrè), pp. 1005 ss.
- GIUCA, Marta (2022): "Disciplina dell'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma", *Archivio Penale*, fasc. 2, 20 maggio
- GIUCA, Marta (2023): "AI and Criminal Liability. Algorithmic Error and Human Negligence in the Context of the European Regulation", *Diritto Penale Contemporaneo – Rivista trimestrale*, 1, pp. 46 ss.
- GIUNTA, Fausto B. (1993): *Illiceità e colpevolezza nella responsabilità colposa* (Padova, Cedam)
- GIUNTA, Fausto B. (1999): "I tormentati rapporti tra colpa e regola cautelare. A proposito di recenti fatti di cronaca", *Diritto Penale e Processo*, n. 10, p. 1296
- GIUNTA, Fausto B. (2006): "Il diritto penale e le suggestioni del principio di precauzione", in *Criminalia. Annuario di scienze penalistiche*, pp. 241-242
- GIUNTA, Fausto B. (2009): "La legalità della colpa", *Criminalia. Annuario di scienze penalistiche*, pp. 150 ss.
- GIUSTI, Patrizia (2023): "La responsabilità penale dell'intelligenza artificiale: i termini generali del problema, con particolare riguardo alle auto senza conducente", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime* (Milano, UTET), pp. 1356 ss.
- GRANDI, Ciro (2016): *Neuroscienze e responsabilità penale. Nuove soluzioni per problemi antichi?* (Torino, Giappichelli)

GRANDI, Ciro (2023): “Positive Obligations (*Garentestellung*) Grounding Criminal Responsibility for not Having Avoided an Illegal Result Connected to the AI Functioning”, in PICOTTI, Lorenzo e PANATTONI, Beatrice (eds.), *Traditional Criminal Law and AI: Crisis or Palingenesis, International Colloquium Section I, Siracusa, 15-16 September 2022, Revue Internationale de Droit Pénal*, 94, 1, pp. 67 ss.

GROSSI, Paolo (2003): *Prima lezione di diritto* (Roma-Bari, Laterza), p. 15

GROTTO, Marco (2012): *Principio di colpevolezza, rimproverabilità soggettiva e colpa specifica* (Torino, Giappichelli)

GRUPPO DI ESPERTI AD ALTO LIVELLO SULL’INTELLIGENZA ARTIFICIALE (2019): *Orientamenti etici per un’IA affidabile* (Bruxelles, Commissione europea)

GUIDUCCI, Roberto (1987): “L’interpretazione del suicidio da Durkheim a oggi”, introduzione a DURKHEIM, Émile (1897, ed. it. 2025), *Il suicidio* (Milano, Rizzoli)

HALLEVY, Gabriel (2010): “The criminal liability of artificial intelligence entities – From science fiction to legal social control”, *Akron Intellectual Property Journal*, vol. 4, n. 2, pp. 191 ss.

HALLEVY, Gabriel (2013): *When Robots Kill: Artificial Intelligence under Criminal Law* (Boston, Northeastern University Press)

HAMON, Ronan, JUNKLEWITZ, Henrik, SANCHEZ, Ignacio (2020): *Robustness and Explainability of Artificial Intelligence – From Technical to Legal and Ethical Considerations* (Luxembourg, Publications Office of the European Union)

HAQUE, M.D. Romael e RUBYA, Sabirat (2023): “An Overview of Chatbot-Based Mobile Mental Health Apps: Insights From App Description and User Reviews”, *JMIR mHealth and uHealth*, vol. 11, e44838

HEIDEGGER, Martin (1954, ed. it. 1976): *Saggi e discorsi*, traduzione di Gianni Vattimo (Milano, Mursia)

HOFSTADTER, Douglas R. (1995): *Fluid Concepts & Creative Analogies. Computer Models of the Fundamental Mechanisms of Thought* (New York, Basic Books), pp. 157-158

HUSSAIN, Shafquat, SIANAKI, Omid Ameri, ABABNEH, Nedat (2019): “A survey on conversational agents/chatbots classification and design techniques”, in BAROLLI, Leonard, TAKIZAWA, Makoto, XHAFA, Fatos, ENOKIDO, Tomoya (eds.), *Web, Artificial Intelligence and Network Applications. Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications* (Cham, Springer), vol. 927, pp. 946 ss.

HUSSERL, Edmund (2025): *Il mondo della vita. Analisi del mondo pre-dato e della sua costituzione*, a cura di Sonia Lisco (Milano–Udine, Mimesis), p. 15

JONAS, Hans (1979, ed. it. 2009): *Il principio di responsabilità. Un’etica per la civiltà tecnologica*, a cura di Pier Paolo Portinaro (Torino, Einaudi)

KANT, Immanuel (1781, ed. it. 1967): *Critica della ragion pura*, a cura di Pietro Chiodi (Torino, U tet), pp. 143 ss.

KAUFMANN, Armin (1964): *Das fahrlässige Delikt*, in *ZfRV*, p. 51

KELLY, Morgan (2025): *First Therapy Chatbot Trial Yields Mental Health Benefits*, in *Trustees of Dartmouth College*, 27 marzo 2025, disponibile su: <https://home.dartmouth.edu> (URL verificato il 30 settembre 2026)

KIRK, Hannah Rose, GABRIEL, Iason, SUMMERFIELD, Chris, VIDGEN, Bertie, HALE, A. Scott (2025): "Why human-AI relationships need socioaffective alignment", *Humanities and Social Sciences Communications*, vol. 12, art. 728

KOULOURI, Theodora, MACREDIE, Robert D., OLAKITAN, David (2022): "Chatbots to Support Young Adults' Mental Health: An Exploratory Study of Acceptability", *ACM Transactions on Interactive Intelligent Systems*, vol. 12, n. 2, pp. 1 ss.

KRAUSS, Detlef (1964): "Erfolgsunwert und Handlungsunwert im Unrecht", *ZStW*, Bd. 76, pp. 47 ss.

KRETZSCHMAR, Kira, TYROLL, Holly, PAVARINI, Gabriela, MANZINI, Arianna, SINGH, Iliana, NEUROX YOUNG PEOPLE'S ADVISORY GROUP (2019): "Can Your Phone Be Your Therapist? Young People's Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support", *Biomedical Informatics Insights*, vol. 11, pp. 1-9

LA VATTIATA, Federico Carmelo (2020), "Artificial Intelligence in Healthcare: Risk Assessment and Criminal Law", *Diritto Penale e Uomo*, fasc. 12, pp. 23 ss., prima pubblicato in *Special Collection on Artificial Intelligence* (Torino, UNICRI), pp. 48 ss.

LA VATTIATA, Federico Carmelo (2021): "Brevi note "a caldo" sulla recente Proposta di Regolamento UE in tema di intelligenza artificiale", *Diritto Penale e Uomo*, fasc. 6, pp. 140 ss.

LA VATTIATA, Federico Carmelo (2022): "La responsabilità penale per danni da intelligenza artificiale alla prova del processo", in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (eds.), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), p. 701

LAESTADIUS, Linnea, BISHOP, Andrea, GONZALEZ, Michael, ILLENČÍK, Diana, CAMPOS-CASTILLO, Celeste (2024): "Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika", *New Media & Society*, vol. 26, n. 10, pp. 5926-5935

LAFAVE, Wayner R. e SCOTT, Austin Wakeman (1986): *Substantive Criminal Law, I*, (St. Paul, Minn., West Publishing Company), pp. 339 ss.

LAGIOIA, Francesca e SARTOR, Giovanni (2020): "AI Systems Under Criminal Law: a Legal Analysis and a Regulatory Perspective", *Philosophy & Technology*, 33, pp. 433 ss.

LANZI, Massimiliano (2023a): *Self-driving cars e responsabilità penale. La gestione del "rischio stradale" nell'era dell'intelligenza artificiale* (Torino, Giappichelli)

LANZI, Massimiliano (2023b): "La responsabilità penale per le auto a guida autonoma", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed., (Milano, UTET), pp. 1373 ss.

LANZI, Massimiliano (2025): "Uso di strumenti a intelligenza artificiale e imputazione della responsabilità penale: difficoltà ricostruttive e prospettive di intervento", *Diritto di Difesa*, 24 gennaio

LAURETTA, Francesco (2025): *Intelligenza artificiale, supporto emotivo e dipendenza affettiva*, in IPSICO. Istituto di Psicologia e Psicoterapia Comportamentale e Cognitiva – Blog (s.d.), disponibile su <https://ipsico.it> (URL verificato il 30 settembre 2026)

LO MONTE, Elio (2021): "Intelligenza artificiale e diritto penale: le categorie dommatiche alla prova del futuribile" in BASILE, Fabio, CATERINI, Mario e ROMANO, Sabato, *Il sistema penale ai confini delle hard sciences* (Pisa, Pacini Giuridica), pp. 41 ss.

LOMBARDI, Oriana (2024): "Responsabilità penale dell'uomo per il danno cagionato attraverso condotte dolose e colpose nell'impiego dei sistemi di intelligenza artificiale", *Sistema Penale*, 12, pp. 57 ss.

LUHMANN, Niklas (1991, ed. it. 1996): *Sociologia del rischio* (Milano, Mondadori Bruno)

MAESTRI, Enrico (2024): *Introduzione e note di sintesi al Regolamento (UE) 2024/1689 sull'intelligenza artificiale (testo ufficiale e integrale)* (Napoli, Jovene), p. 15

MAGRO, Maria Beatrice (2014): "Biorobotica, robotica e diritto penale", in PROVOLO, Debora, RIONDATO, Silvio, YENISEY, Feridun (eds.), *Genetics, Robotics, Law, Punishment* (Padova, Padova University Press), pp. 477 (ingl.) e pp. 499 (it.)

MAGRO, Maria Beatrice (2019): "Robot, cyborg e intelligenze artificiali", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime* (Milano, UTET), pp. 1179 ss.

MAGRO, Maria Beatrice (2020): "Decisione umana e decisione robotica. Un'ipotesi di responsabilità da procreazione robotica", *La Legislazione Penale*, 10 maggio

MAGRO, Maria Beatrice (2023): "Il problema della responsabilità per l'uso di intelligenze artificiali", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed., (Milano, UTET), pp. 1232 ss.

MAIWALD, Manfred (1984): "Zurechnungsprobleme im Rahmen Erfolgsqualifizierter Delikte", *JuS*, p. 444

MANES, Vittorio (2020): "L'oracolo algoritmico e la giustizia penale: al bivio tra tecnologia e tecnocrazia", in RUFFOLO, Ugo (ed.), *Intelligenza Artificiale – Il diritto, i diritti, l'etica* (Milano, Giuffrè), pp. 547 ss.; anticipato in *DisCrimen*, 15 maggio

MANGIONE, Angelo (2024): *Intelligenza artificiale, attività d'impresa e diritto penale. La «funzione di garanzia» nell'organizzazione e dell'organizzazione per la «sorveglianza dell'AI»* (Torino, Giappichelli)

MANTOVANI, Ferrando (1988): “voce Colpa”, *Digesto discipline penalistiche*, vol. II (Torino, UTET), p. 205

MANTOVANI, Ferrando (1997): *Il principio di affidamento nella teoria del reato colposo* (Milano, Giuffrè)

MANTOVANI, Ferrando (2001): “L'obbligo di garanzia ricostruito alla luce dei principi di legalità, di solidarietà, di libertà e di responsabilità personale”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 2, pp. 337 ss.

MANTOVANI, Ferrando (2009): “Il principio di affidamento nel diritto penale”, *Rivista Italiana di Diritto e Procedura penale*, fasc. 2, pp. 536 ss.

MANTOVANI, Ferrando (2007): *Diritto penale. Parte generale*, 5ª ed. (Padova, Cedam), pp. 345 ss.

MANTOVANI, Ferrando (2013): *Diritto penale. Parte generale*, 8ª ed. (Padova, Cedam), pp. 342 ss.

MANTOVANI, Marco Orlando (2021): “Affidamento (principio di)”, in DONINI, Massimo (ed.), *Enciclopedia del diritto – I Tematici: Reato colposo* (Milano, Giuffrè), pp. 1 ss.

MARINUCCI, Giorgio (1965): *La colpa per inosservanza di leggi* (Milano, Giuffrè)

MARINUCCI, Giorgio (1991): “Non c'è dolo senza colpa. Morte dell'«imputazione oggettiva dell'evento» e trasfigurazione nella colpevolezza?”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 1, pp. 3 ss.

MARINUCCI, Giorgio (2005): “Innovazioni tecnologiche e scoperte scientifiche: costi e tempi di adeguamento delle regole di diligenza”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 1, pp. 46 ss.

MASSARO, Antonella (2023): “Intelligenza artificiale e neuroscienze: l'eterno ritorno del diritto penale?”, in MASSARO, Antonella, RIERA, Jaime Miguel Peris, *Derecho penal, inteligencia artificial y neurociencias*, pp. 27 ss.

MASSI, Silvia (2022): “Affidamento sull'intelligenza artificiale e “disimpegno morale” nella definizione dei presupposti della responsabilità penale”, in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (eds.), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), pp. 665 ss.

MASULLO, Maria Novella (2012): *Colpa penale e precauzione nel segno della complessità. Teoria e prassi nella responsabilità dell'individuo e dell'ente* (Napoli, Edizioni Scientifiche Italiane)

MASULLO, Maria Novella (2013): *La connivenza. Uno studio sui confini con la complicità* (Padova, Cedam)

MATTARELLA, Andrea (2025): “Diritto penale e nuove tecnologie: dalla Convenzione Onu contro i reati informatici alle sfide dell’intelligenza artificiale”, *Diritto penale e processo*, n. 2, pp. 250 ss.

MAZZACUVA, Federico (2021): “The impact of ai on corporate criminal liability: Algorithmic misconduct in the prism of derivative and holistic theories”, in VERMEULEN, Gert, PENSAK, Nina, RECCHIA, Nicola (eds.), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice*, pp. 143 ss.

MCNEIL, Ian (1990): *An Encyclopedia of the History of Technology* (London, Routledge), pp. 5 ss.

MCPHERSON, Miller, SMITH-LOVIN, Lynn, COOK, James M. (2001): “Birds of a feather: Homophily in social networks”, *Annual Review of Sociology*, vol. 27, pp. 415-444

MERLE, Roger e VITU, André (1997): *Traité de Droit Criminel. Tome I. Septième édition* (Paris, Éditions Cujas), p. 757

MEZZETTI, Enrico (2010): “Colpa per assunzione”, in VINCIGUERRA, Sergio e DASSANO, Francesco (eds.), *Scritti in memoria di Giuliano Marini* (Napoli, Edizioni Scientifiche Italiane)

MILITELLO, Vincenzo (1988): *Rischio e responsabilità penale* (Milano, Giuffrè), p. 140

MINELLI, Camilla (2022): “La responsabilità ‘penale’ tra persona fisica e corporation alla luce della Proposta di Regolamento sull’Intelligenza Artificiale”, *Diritto Penale Contemporaneo – Rivista trimestrale*, 2, pp. 50 ss.

MINELLI, Giulia Rizzo (2021): “Quando l’autore del reato è un robot: tra vecchi modelli imputativi e nuovi possibili paradigmi di responsabilità penale”, in BASILE, Fabio, CATERINI, Mario e ROMANO, Sabato, *Il sistema penale ai confini delle hard sciences* (Pisa, Pacini Giuridica), pp. 57 ss.

MINISTERO DELLO SVILUPPO ECONOMICO (2019): *Strategia italiana per l’intelligenza artificiale* (Roma, MISE), 31 luglio

MITTELSTADT, Brent, ALLO, Patrick, TADDEO, Mariarosaria, WACHTER, Sandra, FLORIDI, Luciano (2016): “The Ethics of Algorithms: Mapping the Debate”, *Big Data & Society*, vol. 3. n. 2, pp. 1 ss.

MOBILIO, Giuseppe (2020): “L’intelligenza artificiale e i rischi di una “disruption” della regolamentazione giuridica”, *BioLaw Journal – Rivista di BioDiritto*, n. 2, pp. 401 ss.

MONGILLO, Vincenzo (2023): “Corporate Criminal Liability for AI-Related Crimes: Possible Legal Techniques and Obstacles”, in PICOTTI, Lorenzo e PANATTONI, Beatrice (eds.), *Traditional Criminal Law and AI: Crisis or Palingenesis, International Colloquium Section I, Siracusa, 15-16 September 2022*, in *Revue Internationale de Droit Pénal*, 94, 1, pp. 77 ss.

MONGILLO, Vincenzo (2024): “Responsabilità da reato degli enti e crimini connessi all’intelligenza artificiale: tecniche giuridiche di intervento e principali ostacoli”, *Archivio Penale*, fasc. 2

MORAITI, Anna (2021): “AI crimes and misdemeanors: debating the boundaries of criminal liability and imputation”, in VERMEULEN, Gert, PENSACK, Nina, RECCHIA, Nicola (eds.), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice*, pp. 143 ss.

MORONE, Giovanni, DE ANGELIS, Luigi, CINNERA, Alex Martino, CARBONETTI, Riccardo, BISIRRI, Alessio, CIANCARELLI, Irene, IOSA, Marco, NEGRINI, Stefano, KEIKENS, Carlottes, NEGRINI, Francesco (2025): “Artificial intelligence in clinical medicine: a state-of-the-art overview of systematic reviews with methodological recommendations for improved reporting”, *Frontiers in Digital Health*, vol. 7, art. 1550731

MUCCIARELLI, Francesco (2025): “Intelligenza artificiale e condotte manipolative su strumenti finanziari: una nuova aggravante con effetto attenuante”, *Sistema Penale*, 17 ottobre

NISCO, Attilio (2024): “Le neurotecnologie assistite dall’intelligenza artificiale nell’ottica del diritto penale”, *La Legislazione Penale*, 13 settembre

NOGUÉRO, David e VINGIANO-VIRICEL, Iolande (2026): “Intelligence artificielle et véhicules autonomes”, in LOISEAU, Grégoire e BENSAMOUN, Alexandra (eds.), *Droit de l’intelligence artificielle*, 3^e ed. (Parigi, Lextenso-LGDJ), pp. 101 ss.

OECD (2024): *Explanatory memorandum on the updated OECD definition of an AI system*, *OECD Artificial Intelligence Papers*, n. 8

OLTEDAL, Sigve, MOEN, Bjørg-Elin, KLEMPE, Hroar, RUNDMO, Torbjørn (2004): “Explaining risk perception. An evaluation of cultural theory”, *Rotunde*, n. 85 (Trondheim, Torbjørn Rundmo, NTNU)

ONOFRI, Beatrice (2026): “Intelligenza artificiale e crimmigration: profili di discriminazione razziale del migrante alla luce del nuovo AI Act europeo”, *Archivio Penale*, fasc. 1

ORDISH, Johan, MURFET, Hannah, HALL, Alison (2019): *Algorithms as Medical Devices* (Cambridge, PHG Foundation)

ORSINA, Amalia (2015): *Rischio da incertezza scientifica e modelli di tutela penale. Il TUSL come laboratorio di soluzioni al problema dell’esposizione professionale ad agenti patogeni* (Torino, Giappichelli)

PADOVANI, Tullio (1969): “Il grado della colpa”, *Rivista Italiana di Diritto e Procedura Penale*, 1969, pp. 877 ss.

PAGALLO, Ugo (2010): “Saggio sui robot e il diritto penale”, in VINCIGUERRA, Sergio e DASSANO, Francesco (eds.), *Scritti in memoria di Giuliano Marini* (Napoli, E.S.I.), pp. 595 ss.

PAGALLO, Ugo (2017): "AI and bad robots: The criminology of automation", in MCGUIRE, M.R. e HOLT, Thomas J. (eds.), *The Routledge Handbook of Technology, Crime and Justice* (New York, Routledge), pp. 643 ss.

PAGLIARO, Antonio (1985): "Fatto, condotta illecita e responsabilità obiettiva nella teoria del reato", in *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 646-647

PALIERO, Carlo Enrico (1992): "Consenso sociale e diritto penale", *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, pp. 849 ss.

PALIERO, Carlo Enrico (1994): "L'autunno del patriarca. Rinnovamento e trasmutazione del diritto penale dei codici?", *Rivista Italiana di Diritto e Procedura Penale*, fasc. 4, pp. 1240 ss.

PALMISANO, Marta (2019): "L'abuso di mercato nell'era delle nuove tecnologie", *Diritto Penale Contemporaneo – Rivista trimestrale*, 2, pp. 129 ss.

PANATTONI, Beatrice (2021a): "Intelligenza artificiale: le sfide per il diritto penale nel passaggio dall'automazione tecnologica all'autonomia artificiale", *Diritto dell'informazione e dell'informatica*, n. 2, pp. 317 ss.

PANATTONI, Beatrice (2021b): "AI and criminal law: the myth of 'control' in a data-driven society", in VERMEULEN, Gert, PENSAK, Nina, RECCHIA, Nicola (eds.), *Artificial Intelligence, Big Data and Automated Decision-Making in Criminal Justice*, pp. 125 ss.

PANATTONI, Beatrice (2023): "Profili penali dell'interazione uomo-macchina nell'ambito dell'attività medica", in PICOTTI, Lorenzo (ed.), *Automazione, Diritto e Responsabilità* (Napoli, Edizioni Scientifiche Italiane), pp. 269 ss.

PANATTONI, Beatrice (2024): "Condizionamenti e manipolazioni nell'era digitale. Il divieto europeo di pratiche di Intelligenza Artificiale 'manipolative'", *Diritto Penale e Processo*, n. 4, pp. 525 ss.

PANATTONI, Beatrice (2025): "Generative AI and criminal law", *Cambridge Forum on AI: Law and Governance*, I, e9

PANEBIANCO, Giuseppina (2025): "I droni quali "veicoli" di responsabilità penale", *Rivista di Diritto dell'Economia, dei Trasporti e dell'Ambiente*, vol. XXIII, pp. 395 ss.

PANSERA, Maria Teresa (2003): "Introduzione. Il mondo della tecnica e l'uomo artificiale", in GEHLEN, Arnold, *L'uomo nell'era della tecnica* (1984, ed. it. 2003), a cura di Maria Teresa Pansera (Roma, Armando)

PAPA, Michele (2026): "Nuove tecnologie e "accesso alla giustizia" per le persone vulnerabili", in CADOPPI, Alberto, CANESTRARI, Stefano, CONTI, Carlotta, MANNA, Adelmo, PAPA, Michele (eds.), *Tutela penale delle persone vulnerabili* (Milano, Giuffrè), pp. 1005 ss.

PAPA, Michele e TAVELLA, Giulia (2023): "Intelligenza artificiale e legislazione penale", in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed. (Milano, UTET), pp. 1483 ss.; anticipato in *DisCrimen*, 23 giugno

PARABIAGHI, Alberto e MONTE, Marianna (2025): *Chatbot, IA ed empatia: la simulazione comunicativa dell'intelligenza artificiale*, IRCCS, 21 ottobre 2025, disponibile su www.marionegri.it (URL verificato il 30 settembre 2026)

PASCUZZI, Giovanni (2022): "La definizione del problema nella ricerca interdisciplinare", in BELLANTUONO, Giuseppe e IZZO, Umberto (eds.), *Il rapporto tra diritto, economia e altri saperi: la rivincita del diritto. Atti della Lectio Magistralis di Guido Calabresi in occasione della chiusura dell'anno accademico del Dottorato in Studi Giuridici Comparati ed Europei. Facoltà di Giurisprudenza. Trento, 24 ottobre 2019* (Trento, Università di Trento), pp. 55 ss.

PENTINA, Iryna, HANCOCK, Tyler, XIE, Tianling (2023): "Exploring relationship development with social chatbots: A mixed-method study of Replika", *Computers in Human Behavior*, vol. 140, art. 107600

PERKINS, Rollin M. e BOYCE, Ronald N. (1982): *Criminal Law* (New York, Foundation Press), pp. 840 ss.

PICHOWICZ, Wojciech, KOTAS, Marek, PIOTROWSKY, Patryk (2025): "Performance of mental health chatbot agents in detecting and managing suicidal ideation", *Scientific Reports*, vol. 15, art. 31652

PICOTTI, Lorenzo (2021): "Profili di responsabilità penale per la circolazione di veicoli a guida autonoma", in CATENACCI, Mauro, D'ASCOLA, Vincenzo Nico e RAMPIONI, Roberto, *Studi in onore di Antonio Fiorella*, vol. 1, (Roma, RomaTrePress), pp. 813 ss.

PICOTTI, Lorenzo (2022a): "Cybercrimes and criminal relevance of artificial intelligence", in TOLEDO, Francisco J. Castro (Coord.), *La transformación algorítmica del sistema de justicia penal*, p. 67 ss.

PICOTTI, Lorenzo (2022b): "Ente", in MANNOZZI, Grazia, PERINI, Chiara, SCOLETTA, Marco Maria, SOTIS, Carlo, TAVERRINI, Sara Bianca, CONSULICH, Federico, *Studi in onore di Carlo Enrico Paliero* (Milano, Giuffrè), pp. 1477 ss.

PICOTTI, Lorenzo (2022c): "«Nuovi» crimini cibernetici e possibile rilevanza dell'intelligenza artificiale. Gli Atti Digitali degli Stati Generali del Diritto di Internet", *Diritto di Internet*, suppl. fasc. 1

PICOTTI, Lorenzo (2022d): "Spunti di riflessione per il penalista di fronte all'utilizzo della robotica nella medicina", *Responsabilità sanitaria, rischio clinico e valore della persona*, 1, pp. 111 ss.

PICOTTI, Lorenzo (2023a): "General report", in PICOTTI, Lorenzo e PANATTONI, Beatrice (eds.), *Traditional Criminal Law and AI: Crisis or Palingenesis, International Colloquium Section I, Siracusa, 15-16 September 2022*, in *Revue Internationale de Droit Pénal*, 94, 1, pp. 11 ss.

PICOTTI, Lorenzo (2023b): “Veicoli a guida autonoma e responsabilità penale”, in CASSANO, Giuseppe e PICOTTI, Lorenzo (eds.), *Veicoli a guida autonoma. Veicoli a impatto zero. Regole, intelligenza artificiale, responsabilità* (Pisa, Pacini Giuridica), pp. 255 ss.

PICOTTI, Lorenzo (2024a): “Diritto penale e nuove tecnologie: una sfida per la prassi e per la teoria”, in BONDI, Alessandro, MARRA, Gabriele, PALAVERA, Rosa, *Diritto penale tra teoria e prassi* (Urbino, Urbino University Press), pp. 83 ss.

PICOTTI, Lorenzo (2024b): “Categorie tradizionali del diritto penale e intelligenza artificiale: crisi o palingenesi? Le raccomandazioni dell’Association Internationale de Droit Pénal e la rilevanza del recente regolamento europeo sull’intelligenza artificiale”, *Sistema Penale*, 31 luglio

PICOTTI, Lorenzo (2024c): “Intelligenza artificiale e diritto penale: le sfide ad alcune categorie tradizionali”, *Diritto penale e processo*, n. 3, pp. 293 ss.

PICOTTI, Lorenzo (2025): “I contenuti penali della legge sull’intelligenza artificiale”, *Sistema Penale*, 2 dicembre

PIERGALLINI, Carlo (2004): *Danno da prodotto e responsabilità penale. Profili dogmatici e politico-criminali* (Milano, Giuffrè)

PIERGALLINI, Carlo (2017): “voce Colpa (Diritto penale)”, in AA. VV., *Enciclopedia del Diritto*, Annali, vol. X (Milano, Giuffrè), p. 243

PIERGALLINI, Carlo (2020): “Intelligenza artificiale: da “mezzo” ad “autore” del reato?”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 4, pp. 1743 ss.

PIETRELLA, Tommaso (2025): “Intelligenza artificiale: rischio di manipolazione ed esigenze di criminalizzazione”, in GAMBARDELLA, Marco (ed.), *Transizioni e tutela penale. Le sfide della modernità in materia di immigrazione, ambiente e intelligenza artificiale* (Milano, Giappichelli), pp. 101 ss.

PISANI, Nicola (2012): *La “colpa per assunzione” nel diritto penale del lavoro. Tra aggiornamento scientifico e innovazioni tecnologiche* (Napoli, Jovene)

PIVA, Daniele (2020): “Spunti per una riscoperta della colpa per assunzione”, *Discrimen*, 9 settembre, p. 1

PIVA, Daniele (2022): “Machina discere, (deinde) delinquere et puniri potest”, in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (eds.), *Il diritto nell’era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), pp. 681 ss.

POLI, Pier Francesco (2021): *La colpa grave. I gradi della colpa tra esigenze di extrema ratio ed effettività della tutela penale* (Milano, Giuffrè)

POMPILI, Maurizio (2022): *Il rischio di suicidio. Valutazione e gestione* (Milano, Raffaello Cortina Editore)

POMPILI, Maurizio (2024): “Gestire il rischio di suicidio oggi”, intervento al Congresso Nazionale della SIP-Società Italiana di Psichiatria, Verona, 29 maggio – 1° giugno 2024, disponibile su: www.youtube.com (URL verificato il 30 settembre 2026)

PORTINARO, Pier Paolo (2009): “Introduzione”, in JONAS, Hans, *Il principio di responsabilità. Un’etica per la civiltà tecnologica* (1979, ed. it. 2009), a cura di Pier Paolo Portinaro (Torino, Einaudi), p. XXIV

PREZIOSI, Stefano (2020): “La responsabilità da reato degli enti e intelligenza artificiale”, *La responsabilità amministrativa delle società e degli enti*, n. 4, pp. 173 ss.

PREZIOSI, Stefano (2022): “La responsabilità penale per eventi generati da sistemi di IA o da processi automatizzati”, in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (eds.), *Il diritto nell’era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), pp. 713 ss.

PULITANÒ, Domenico M. (2005): *Diritto penale* (Torino, Giappichelli), p. 388

PULITANÒ, Domenico M. (2006): “Il diritto penale tra vincoli di realtà e sapere scientifico”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. 3, p. 779

PULITANÒ, Domenico M. (2020): “Il penale tra teoria e politica”, *Sistema Penale*, 9 novembre

PULITANÒ, Domenico M. (2023): *Diritto penale*, X ed. (Torino, Giappichelli), p. 168

RIONDATO, Silvio (2014): “Robotica e diritto penale”, in PROVOLO, Debora, RIONDATO, Silvio, YENISEY, Feridun (eds.), *Genetics, Robotics, Law, Punishment* (Padova, Padova University Press), pp. 589 (ingl.) e pp. 599 (it.)

ROCCO, Arturo (1910): “Il Problema e il metodo della scienza del diritto penale”, *Rivista Italiana di Diritto e Procedura Penale*, fasc. I, pp. 504-505; anche in SEMINARA, Sergio (2020), *Prefazione ad Arturo Rocco. Il problema e il metodo della scienza del diritto penale* (Napoli, Editoriale Scientifica)

ROMANO, Bartolomeo (2004): *Commentario sistematico del Codice penale* (Milano, Giuffrè), p. 427

ROMANO, Bartolomeo (2024): “Il ddl in materia di IA: l’utilizzo nell’attività giudiziaria e in ambito medico”, *Rivista Italiana di Medicina Legale e del Diritto in Campo Sanitario*, 1-2, pp. 403 ss.

ROMANÒ, Leonardo (2023): “La responsabilità penale al tempo di ChatGPT: Prospettive de iure condendo in tema di gestione del rischio da intelligenza artificiale generativa”, *Sistema Penale*, 17 maggio

ROXIN, Claus (1962): “Pflichtwidrigkeit und Erfolg bei fahrlässigen Delikten”, *ZStW*, p. 172

ROXIN, Claus (1979): *Strafrecht. Allgemeiner Teil. Band I: Grundlagen. Der Aufbau der Verbrechenslehre* (München, C.H. Beck), p. 133 ss.

ROYAL SOCIETY (1992): *Risk Analysis, Perception and Management. Report of a Royal Society Study Group* (London, The Royal Society)

RUGA RIVA, Carlo (2006): "Principio di precauzione e diritto penale. Genesi e contenuto della colpa in contesti di incertezza scientifica", in DOLCINI, Emilio, PALIERO, Carlo Enrico (eds.), *Studi in onore di Giorgio Marinucci*, vol. II (Milano, Giuffrè), pp. 1759 ss.

SALCUNI, Giandomenico e DI FLORIO, Mattia (2023): "Intelligenza artificiale e abusi di mercato" in CADOPPI, Alberto, CANESTRARI, Stefano, MANNA, Adelmo, PAPA, Michele (eds.), *Cybercrime*, II ed., (Milano, UTET), pp. 1394 ss.

SALVADORI, Ivan (2021): "Agenti artificiali, opacità tecnologica e distribuzione della responsabilità penale", *Rivista Italiana di Diritto e Procedura Penale*, fasc. 1, pp. 83 ss.

SALVADORI, Ivan (2025): "Minacce alla *automotive cybersecurity* e tutela penale della sicurezza dei trasporti", *Archivio Penale*, fasc. 1

SCHEUTZ, Matthias (2012): "The inherent dangers of Unidirectional Emotional Bonds between Humans and Social Robots", in LIN, Patrik, ABNEY, Keith, BEKEY, A. George (eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (Cambridge, MIT press), pp. 205 ss.

SCHMITT, Carl (1932, ed. it. 2018): *Legalità e legittimità*, a cura di Carlo Galli, traduzione di Giovanni Zanotti (Bologna, Il Mulino)

SCHROEDER, Friedrich-Christian (1989): "Die Fahrlässigkeit als Erkennbarkeit der Tatbestandsverwirklichung", *JZ*, p. 776

SEVERINO, Emanuele (1988): *La tendenza fondamentale del nostro tempo* (Milano, Adelphi), pp. 67 ss.

SEVERINO, Paola (2020): "Intelligenza artificiale e diritto penale", in RUFFOLO, Ugo (ed.), *Intelligenza Artificiale – Il diritto, i diritti, l'etica* (Milano, Giuffrè), pp. 531 ss.

SEVERINO, Paola (2022): "Le implicazioni dell'intelligenza artificiale nel campo del diritto con particolare riferimento al diritto penale", in SEVERINO, Paola (ed.), *Intelligenza artificiale. Politica, economia, diritto, tecnologia* (Roma, Luiss University Press), pp. 76 ss.

STANZIONE, Pasquale (2022): *Diritto e potere nell'era dell'intelligenza artificiale. Lezione nell'ambito del TrAIL Jean Monnet Project – Convegno finale*, Università degli Studi di Trento, 10 marzo 2022, online: <http://www.youtube.com/> (min. 1:02:13)

STELLA, Federico (2003): *Giustizia e modernità. La protezione dell'innocente e la tutela delle vittime* (Milano, Giuffrè), p. 552

STORTONI, Luigi (2004): "Angoscia tecnologica ed esorcismo penale", *Rivista Italiana di Diritto e Procedura Penale*, fasc. I, pp. 71 ss.

SUNSTEIN, Cass Robert (2005, ed. it. 2010): *Il diritto della paura. Oltre il principio di precauzione* (Bologna, Il Mulino)

- SURDEN, Harry e WILLIAMS, Mary-Anne (2016): *Technological opacity, predictability, and Self-driving cars*, in *Cardozo Law Review*, vol. 38, n.1, p. 157
- SWITZKY, Lawrence (2020): "Eliza Effects. Pygmalion and the Early Development of Artificial Intelligence", *Shaw. The Annual of Bernard Shaw Studies*, vol. 40, n. 1, pp. 50-68
- TALLACCHINI, Maria Teresa (2006): "Principio di precauzione: epistemologia e diritto", *Quaderni della Segreteria Generale della CEI*, vol. X, pp. 125-140
- TCHAKHOTINE, Serge (1938, ed. fr. 1973): *Le viol des foules par la propagande politique* (Parigi, Gallimard)
- TERRIZZI, Laura Anna (2023): *Medical devices e diritto penale. Profili di responsabilità del produttore e dell'utilizzatore* (Milano, Giuffrè)
- TORDINI-CAGLI, Silvia (2008): *Principio di autodeterminazione e consenso dell'avente diritto* (Bologna, Bononia University Press)
- TRIFFTERER, Otto (1979): "Die „objektive Voraussehbarkeit“ (des Erfolges und des Kausalverlaufs) – unverzichtbares Elementi im Begriff der Fahrlässigkeit oder allgemeiner Verbrechenselement aller Erfolgsdelikte?", *Festschrift für Bocklmann*, pp. 201 ss.
- TRIPODI, Andrea Francesco (2022): "Abusi di mercato e trading algoritmico", in GIORDANO, Rosaria, PANZAROLA, Andrea, POLICE, Aristide, PREZIOSI, Stefano, PROTO, Massimo (ed.), *Il diritto nell'era digitale. Persona, mercato, amministrazione, giustizia* (Milano, Giuffrè), pp. 745 ss.
- TRIPODI, Andrea Francesco (2023): "Uomo, societas, machina", *La Legislazione Penale*, 10 maggio
- TRONCONE, Pasquale (2022): "Il sistema di intelligenza artificiale nella trama grammaticale del diritto penale. Dalla responsabilità umana alla responsabilità delle macchine pensanti: un inatteso *return trip effect*", *Cassazione Penale*, n. 9, pp. 3287 ss.
- TURKLE, Sherry (1995, ed. it. 1997): *La vita sullo schermo. Nuove identità e relazioni sociali nell'epoca di Internet*, traduzione di Bernardo Parrella (Milano, Apogeo), pp. 116 ss.
- UBERTIS, Giulio (2020): "Intelligenza artificiale, giustizia penale e controllo umano significativo", *Diritto Penale Contemporaneo – Rivista trimestrale*, 4, pp. 78-79
- VADALÀ, Rosa Maria (2023): "La questione penale delle auto a guida autonoma in prospettiva comparata", *La Legislazione Penale*, 13 novembre
- VAIDYAM, Aditya Nrusimha, WISNIEWSKI, Hannah, HALAMKA, John David, KASHAVAN, Matcheri S., TOROUS, John Blake (2019): "Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape", *The Canadian Journal of Psychiatry*, vol. 64, n. 7, pp. 456 ss.
- VENEZIANI, Paolo (2003): *Regole cautelari proprie ed improprie nella prospettiva delle fattispecie colpose causalmente orientate* (Padova, Cedam)

VENTUROLI, Marco (2015): *La vittima nel sistema penale. Dall'oblio al protagonismo?* (Napoli, Jovene)

VIDAL, Georges (1949): *Cours de Droit criminel et de science pénitentiaire. Tome I* (Paris, Rousseau & C^{ie}, Editeurs), pp. 193-194

WEIZENBAUM, Joseph (1966): "ELIZA – A Computer Program For the Study of Natural Language Communication Between Man And Machine", *Communications of the ACM*, vol. 9, n. 1, pp. 36-45

WELZEL, Hans (1961): *Fahrlässigkeit und Verkehrsdelikte. Zur Dogmatik der fahrlässigen Delikte* (Karlsruhe, Müller), pp. 14 ss.

WILLIAMS, Glanville (1983): *Textbook of Criminal Law* (London, Stevens), pp. 96 ss. e pp. 388 ss.

XIE, Tianling e PENTINA, Iryna (2022): "Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika", *Proceedings of the 55th Hawaii International Conference on System Sciences* (Honolulu, University of Hawaii at Mānoa), pp. 2046-2055

ZIRULIA, Stefano (2018): *Esposizione a sostanze tossiche e responsabilità penale* (Milano, Giuffrè)

Media e cronaca

BERETTA, Ilaria (2025): *Se il chatbot è diventato psicologo, amico e compagno degli adolescenti*, in *avvenire.it*, 16 luglio, disponibile al link www.avvenire.it (URL verificato il 30 settembre 2026)

BELOT, Laure (2026): *Generative AI, psychiatry and the risks of self-service therapy*, in *lemonde.fr*, 24 gennaio, disponibile (per gli abbonati) al link www.lemonde.fr (URL verificato il 30 settembre 2026)

COLUMBRO, Donata (2023): *Le macchine e l'inganno banale, ma necessario, dell'intelligenza artificiale*, in *La Stampa* (Rubriche), 31 marzo 2023, disponibile su www.lastampa.it (URL verificato il 30 settembre 2026)

DA ROLD, Cristina (2025): *Che cosa serve all'IA per potersi occupare di salute mentale?*, in *infodata.ilsole24ore.com*, 6 ottobre, disponibile al link: www.infodata.ilsole24ore.com (URL verificato il 30 settembre 2026)

GUARRERA, Stefano (2025): *Character.AI vieta le chat ai minorenni e su ChatGpt oltre un milione di utenti parla di suicidio*, in *corriere.it*, 29 ottobre, disponibile al link www.corriere.it (URL verificato il 30 settembre 2026)

KANG, Cecilia, MAC, Ryan, TAN, Eli (2026): *Meta and YouTube Found Negligent in Landmark Social Media Addiction Case*, in *nytimes.com*, 25 marzo, disponibile al link www.nytimes.com (URL verificato il 30 settembre 2026)

KUZNIA, Rob, GORDON, Allison, LAVANDERA, Ed (2025): *'You're not rushing. You're just ready:' Parents say ChatGPT encouraged son to kill himself*, in *edition.cnn.com*, 20 novembre, disponibile al link <https://edition.cnn.com> (URL verificato il 30 settembre 2026)

MACCARTHY, Mark (2025): *AI Companies Should be Liable for the Illegal Conduct of AI Chatbots*, in *techpolicy.press*, 20 agosto, disponibile al link <https://www.techpolicy.press> (URL verificato il 30 settembre 2026)

MONTGOMERY, Blake (2024): *Mother says AI chatbot led her son to kill himself in lawsuit against its maker*, in *theguardian.com*, 23 ottobre, disponibile al link www.theguardian.com (URL verificato il 30 settembre 2026)

MCALLEN, Jess (2025): *Il terapeuta artificiale*, in *internazionale.it*, 31 gennaio, disponibile al link www.internazionale.it (URL verificato il 30 settembre 2026)

MCLENNAN, April (2025): *AI chatbots could be accelerating the loneliness epidemic, according to new research*, in *abc.net.au*, 21 ottobre, disponibile al link www.abc.net.au (URL verificato il 30 settembre 2026)

MCBAIN, Ryan K. (2025): *Teens Are Using Chatbots as Therapists. That's Alarming*, in *nytimes.com*, 25 agosto, disponibile al link: www.nytimes.com (URL verificato il 30 settembre 2026)

PIERINI, Simone (2025): *Sophie, la 29enne suicida negli Usa dopo la "terapia" virtuale con ChatGpt. «L'IA l'ha aiutata a scrivere la lettera d'addio ai genitori»*, in *ilmessaggero.it*, 21 agosto, disponibile al link www.ilmessaggero.it (URL verificato il 30 settembre 2026)

REDAZIONE IL MESSAGGERO (2025): *Ragazzi, intelligenza artificiale psicologo chatbot e salute mentale*, in *ilmessaggero.it*, 9 dicembre, disponibile al link www.ilmessaggero.it (URL verificato il 30 settembre 2026)

REDAZIONE LE FIGARO (2023): *Un Belge se suicide après avoir trouvé refuge auprès d'un robot conversationnel*, in *lefigaro.fr*, 29 marzo, disponibile al link www.lefigaro.fr (URL verificato il 30 settembre 2026)

REDAZIONE THE ECONOMIST (2025): *Millions are turning to AI for therapy. But is the technology ready?*, in *economist.com*, 11 novembre, disponibile (per gli abbonati) al link www.economist.com (URL verificato il 30 settembre 2026)

REDAZIONE EURONEWS (2025): *Caso adolescente suicida a causa di un chatbot, giudice Usa: libertà d'espressione non è diritto dell'AI*, in *euronews.com*, 22 maggio, disponibile al link it.euronews.com (URL verificato il 30 settembre 2026)

SAVE THE CHILDREN (2025): *Italy: about 42% of teenagers turn to ai when they feel sad, anxious, or for life decisions*, 17 novembre 2025, disponibile su: www.savethechildren.net (URL verificato il 30 settembre 2026)

SCHULZE, Elizabeth e LEIB, Mason (2026): *Jury decides YouTube, Meta should pay \$6 million in damages after finding them negligent in landmark trial*, in *goodmorningamerica.com*, 25 marzo, disponibile al link www.goodmorningamerica.com (URL verificato il 30 settembre 2026)

SOCIETÀ ITALIANA DI PSICHIATRIA (2025): *AI e adolescenti in crisi: stregati dal chatbot*, in *psichiatria.it*, 3 settembre, disponibile al link psichiatria.it (URL verificato il 30 settembre 2026)

TIDY, Joe (2024): *Character.ai: Young people turning to AI therapist bots*, in *bbc.com*, 5 gennaio, disponibile al link www.bbc.com (URL verificato il 30 settembre 2026)

TIKU, Nitasha (2025): *A teen contemplating suicide turned to a chatbot. Is it liable for her death?*, in *washingtonpost.com*, 16 settembre, disponibile (per gli abbonati) al link www.washingtonpost.com (URL verificato il 30 settembre 2026)

WOEBOT HEALTH (2021): *"New survey shows paradigm shift in American attitudes and opinions about mental health technology"*, *Woebot Health*, 18 ottobre 2021, disponibile su: woebothealth.com (URL verificato il 30 settembre 2026)