

IA: dal marketing alla sanità cosa è permesso e cosa è vietato

Ciccia Messina a pag. 6 e 7

Dalle allucinazioni alla raccolta di voci: le linee guida Ue sulle deroghe alle pratiche proibite

IA manipolativa vietata. Anzi no Ammessi alcuni sistemi persuasivi se non c'è danno rilevante

Pagina a cura di
ANTONIO CICCIA MESSINA

Mano leggera sulle allucinazioni dell'Intelligenza artificiale (IA); ok alle IA che usano tecniche subliminali per curare i disabili mentali, a quelle che raccolgono a strascico sul web la voce delle persone e a quelle che deducono le emozioni dai messaggi vocali in una chat per inviare pubblicità personalizzata: sono tutte eccezioni ai sistemi vietati dall'articolo 5 del regolamento Ue sull'IA n. 2024/1689 (AI Act) e sono tutte descritte dagli orientamenti della Commissione Ue, nella versione finale del 29 luglio 2025. Si tratta di un folto numero di deroghe (si veda *ItaliaOggi* del 17/2/2025 per una prima analisi sulla base dello schema di provvedimento della Commissione), le quali si aggiungono ai casi esclusi dall'ambito di applicazione dell'AI Act (ricerca e sviluppo, attività di persone fisiche di natura personale non professionale e, di norma, sistemi rilasciati con licenza libera e open source).

Le IA, dunque, fanno lo slalom tra i divieti e ciò comporta la necessità per gli utilizzatori (deployer) di andare molto cauti: sebbene "non vietati" i sistemi di IA sono quasi sempre ad alto rischio e, quindi, se si esce dalla zona di liceità, i cui confini sono precari e non esattamente individuati dalla legge europea, possono scattare responsabilità amministrative, per violazione dell'AI Act, e civili per eventuali danni patiti dagli interessati.

Tecniche subliminali accettate. Sono vietati i sistemi di IA che utilizzano tecni-

che subliminali o volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere il comportamento: il divieto scatta, però, solo se provocano o possono ragionevolmente provocare un danno significativo. Allora, se non c'è danno significativo, il sistema subliminale o manipolativo non è vietato.

Applicando l'eccezione, la Commissione precisa che un sistema di IA generativa per il solo fatto di presentare incidentalmente informazioni false o fuorvianti e allucinazioni non può essere considerato un sistema di IA ingannevole. In particolare, continuano le linee guida, ciò può verificarsi quando il fornitore del sistema ha adeguatamente informato gli utenti in merito ai limiti del sistema e ha integrato nel sistema garanzie adeguate a ridurre al minimo tali risultati e, inoltre, a condizione che il sistema non sia destinato a contesti sensibili, né utilizzato in ambiti (per esempio sanità, istruzione, elezioni) in cui è probabile che si verifichino gravi conseguenze dannose. Pertanto, la commissione si dimostra benevola nei confronti degli output allucinati delle IA, che vanno, a questo punto, disinnescati con una costante supervisione umana.

Sempre ragionando sui "danni significativi", la Commissione trova altre deroghe alle IA vietate, disquisendo dei "danni fisici minori": essi, si legge nelle linee guida, possono includere lesioni meno gravi, come contusioni o disagio temporaneo, che non hanno conseguenze significative o durature e non raggiungeranno pertanto la soglia di rilevanza (ai fini del divieto) an-

che se i sistemi possono utilizzare tecniche subliminali, volutamente manipolative o ingannevoli.

L'IA può intrattenere le persone vulnerabili. L'articolo 5 della legge europea sull'IA elenca i sistemi di IA che sfruttano le vulnerabilità dovute a età, disabilità o specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere il comportamento: anche in questo caso, i sistemi sono messi al bando solo se provocano o possono ragionevolmente provocare un danno significativo.

Per esempio, spiega la Commissione Ue, non si dovrebbe automaticamente ritenere che un sistema di IA sfrutti le vulnerabilità socioeconomiche delle persone quando esso presenta delle distorsioni involontarie e incide in modo sproporzionato sulle persone socialmente svantaggiate (discriminazione indiretta) a causa di dati di addestramento distorti, in quanto queste ultime non sono specificamente prese di mira come avviene invece nel caso della discriminazione diretta. La mancanza di dolo diretto ha l'effetto di derubricare risultati (output) dannosi da assolutamente vietati a tollerati. Continua un atteggiamento di clemente indulgenza verso l'IA che sbaglia per negligenza o imperizia.

Sulla stessa scia di benevola deroga ai divieti, si pone l'affermazione delle linee guida, per cui tali divieti non si applicano ai sistemi di IA che



Peso: 1-2%, 6-85%

si rivolgono ai consumatori sulla base di variabili di ampia portata, che non sono correlate tangenzialmente con gruppi vulnerabili in specifiche situazioni sociali o economiche, come la marca e il modello del telefono di una persona, quanto è grande la città in cui vive, quanto e dove viaggia, e così via. Peralto, la cautela è d'obbligo, perché talvolta, le variabili "di ampia portata" sono in sé stesse riferibili a persone o gruppi vulnerabili.

Proseguendo nella delimitazione del divieto, la Commissione approfondisce il concetto di "persuasione lecita" e ne rinviene una applicazione, per esempio, nel caso di un sistema di IA che, analizzando le emozioni dei clienti (soggetti vulnerabili, poiché in posizione di debolezza contrattuale rispetto alle imprese), al fine di fornire loro assistenza, mette in atto un'azione persuasiva e si allinea ai loro interessi. È facile constatare, però, che un venditore non è mosso da intenti altruistici e che lo stesso venditore ricostruisce l'interesse dal cliente dal suo punto di vista, ciò dal punto di vista di chi parte del presupposto che il cliente "deve" comprare.

Passando ad altro tipo di vulnerabilità, consistente in uno stato patologico, le linee

guida della commissione scrivono che le tecniche subliminali basate sull'IA possono essere utilizzate nel trattamento psicologico di una malattia mentale o nella riabilitazione fisica, quando tali pratiche sono svolte conformemente al diritto e alle norme in ambito medico, tra cui l'ottenimento del consenso esplicito della persona fisica o dei suoi rappresentanti legali. Lo sdoganamento delle IA subliminali per il trattamento delle patologie mentali è basato sul consenso informato, che è un finito consenso. Il consenso presuppone la libertà di esprimerlo e il malato si trova per definizione in uno stato di costrizione determinato dalla malattia.

La Commissione elenca, infine, una lunga lista di casi, riportati nella tabella pubblicata in pagina, in cui le IA possono essere usate in relazione a soggetti vulnerabili. Anche per questi sistemi emerge il fondato dubbio che possano avere effetti manipolativi e, pertanto, occorre che le persone, tenute alla cura e alla vigilanza delle persone vulnerabili, sorvegliano le interazioni di queste ultime con le IA.

Stop allo scraping di immagini, ma non delle voci. L'AI act proibisce il web scraping di immagini facciali, ma, come notano gli orientamenti

in esame, non quello non mirato di dati biometrici diversi dalle immagini facciali, come i campioni vocali.

L'interdizione, poi, non si applica ai sistemi di IA che raccolgono grandi quantità di immagini facciali da internet per costruire modelli di IA, che generano nuove immagini di persone fittizie, perché tali sistemi non darebbero luogo al riconoscimento di persone reali.

Intercettazioni legali di emozioni. Sono preclusi i sistemi di IA che inferiscono le emozioni sul luogo di lavoro o negli istituti di istruzione, ma fanno eccezione se usati per motivi medici o di sicurezza.

Interpretando il regolamento, la Commissione nota che i sistemi di riconoscimento delle emozioni utilizzati in un contesto commerciale, per rivolgersi ai clienti, non rientrano nel divieto, a prescindere che siano basati o meno su dati biometrici. Pertanto, per esempio, non rientrano nel divieto i sistemi di IA che consentono il riconoscimento delle emozioni sulla base della battitura su tastiera o dei messaggi vocali dei clienti (per esempio messaggi chat, uso di assistenti vocali virtuali), utilizzati nel marketing online per applicazioni per la visualizzazione di messaggi personalizzati e a fini publi-

citari, anche in ambienti smart (cartelloni intelligenti). Tali pratiche possono tuttavia rientrare nei divieti di manipolazione e sfruttamento dannosi, ma, anche qui, se sono soddisfatte tutte le condizioni per la loro applicazione.

Punteggi sociali ammessi. L'AI act sbarrà la strada ai sistemi di attribuzione di punteggio sociale, sulla base del comportamento sociale o delle caratteristiche personali, ma non quando il trattamento è giustificato e proporzionato, come nel caso delle piattaforme online che, per motivi di sicurezza, profilano degli utenti sulla base di dati pertinenti per il contesto.

Non rientrano nel divieto i sistemi di IA che consentono il riconoscimento delle emozioni sulla base della battitura su tastiera o dei messaggi vocali dei clienti, utilizzati nel marketing online

I sistemi non vietati

Non sono vietati i sistemi di IA che

Aiutano i bambini nell'apprendimento e nei giochi

Aiutano gli anziani nella vita quotidiana e migliorano la loro salute e le cure mediche, come assistenti personali o robot assistivi, o rafforzano le loro competenze digitali

Sostengono l'integrazione economica e di altro tipo delle persone socialmente svantaggiate nella società, migliorano le loro competenze

Assistono le persone con disabilità visive o uditive o forniscono un apprendimento adattato e personalizzato

Generano soluzioni accessibili eliminando gli ostacoli all'uso di prodotti e servizi da parte delle persone con disabilità

Aiutano le persone con disabilità nella loro vita quotidiana e consentono loro di integrarsi e di partecipare pienamente alla società



Peso: 1-2%, 6-85%