Digeat N.3 - 19 Settembre 2024

Sistemi di IA generativa e fiducia nell’ecosistema dell’informazione: fake news e deep fake

Di Giovanni Fiorino

Abstract

La fiducia è definita come l’atteggiamento, verso altri o verso sé stessi, che risulta da una valutazione positiva di fatti, circostanze, relazioni, per cui si confida nelle altrui o proprie possibilità, e che generalmente produce un sentimento di sicurezza e tranquillità: essa, dunque, è alla base del circolo virtuoso che consente la crescita ed ha, quale presupposto, la “valutazione positiva di fatti, circostanze, relazioni” e, quale risultato, il “sentimento di sicurezza e tranquillità”. Com’è di tutta evidenza la fiducia assume particolare rilievo nell’ambito dei rapporti che nascono e si sviluppano in un contesto, quello telematico, caratterizzato dalla distanza e, in genere, dalla assenza di contatti fisici idonei a consentire di valutare l’affidabilità dell’interlocutore. Segue una disamina del quadro normativo volto a regolamentare il cd. “ecosistema dell’informazione”.

Indice

La fiducia nella disciplina normativa europea

In materia di incidenza sul mercato e sulla sicurezza delle transazioni e dei servizi commerciali,

il considerando n. 1 del **Regolamento n. 910/2014 del Parlamento Europeo e del Consiglio del 23 luglio 2014 – c.d. EIDAS** – reca testualmente: “Instaurare la fiducia negli ambienti online è fondamentale per lo sviluppo economico e sociale” cosicché, prosegue, “La mancanza di fiducia, dovuta in particolare a una percepita assenza di certezza giuridica, scoraggia i consumatori, le imprese e le autorità pubbliche dall’effettuare transazioni per via elettronica e dall’adottare nuovi servizi”.

Pertanto, lo stesso “regolamento mira a rafforzare la **fiducia nelle transazioni elettroniche nel mercato interno** fornendo una base comune per interazioni elettroniche sicure fra cittadini, imprese e autorità pubbliche, in modo da migliorare l’efficacia dei servizi elettronici pubblici e privati, nonché dell’eBusiness e del commercio elettronico, nell’Unione europea.” (considerando n. 2).

La **necessità di affidarsi**, nel mercato globale, è ribadita dal **Regolamento n. 1183/2024 del Parlamento Europeo e del Consiglio dell'11 aprile 2024** che, recando modifiche al regolamento n. 910/2014 ed istituendo il c.d. **portafoglio europeo di identità digitale**, al considerando n. 18 reca: "Proteggere i cittadini e i residenti dell'Unione dall'uso non autorizzato o fraudolento dei portafogli europei di identità digitale è di grande importanza al fine di garantire la fiducia negli stessi e la loro ampia diffusione" ed aggiunge che "Agli utenti dovrebbe essere garantita una protezione efficace contro tale uso improprio".

Sul piano della "sicurezza delle reti e dei sistemi informativi dell'Unione" la **Direttiva n. 2016/1148 del Parlamento europeo e del Consiglio, del 6 luglio 2016 – NIS** -, esordiva, testualmente: "Le reti e i sistemi e servizi informativi svolgono un ruolo vitale nella società", cosicché "È essenziale che essi siano affidabili e sicuri per le attività economiche e sociali e in particolare ai fini del funzionamento del mercato interno".

Tale Direttiva è stata abrogata da quella recante **il numero 2022/2555 – NIS2** – che, rimarcando la necessità di garantire "misure per un livello comune elevato di cibersicurezza nell'Unione" nel considerando n. 3 sottolinea che "il numero, la portata, il livello di sofisticazione, la frequenza e l'impatto degli incidenti stanno aumentando e rappresentano una grave minaccia per il funzionamento dei sistemi informatici e di rete" e che "tali incidenti possono quindi impedire l'esercizio delle attività economiche nel mercato interno, provocare perdite finanziarie, minare la fiducia degli utenti e causare gravi danni all'economia e alla società dell'Unione".

Pertanto, tale Direttiva ribadisce come "**la preparazione e l'efficacia della cibersicurezza sono oggi più che mai essenziali per il corretto funzionamento del mercato interno**", aggiungendo che "la cibersicurezza è un fattore abilitante fondamentale per molti settori critici, affinché questi possano attuare con successo la trasformazione digitale e cogliere appieno i vantaggi economici, sociali e sostenibili della digitalizzazione".

Segue: la fiducia nel trattamento e nella valorizzazione dei dati personali

> Seguendo **uno schema che potremmo definire "per cerchi concentrici"** e che prenda le mosse, come visto, dalla rilevanza della fiducia nell'ambito delle transazioni e dei servizi commerciali, passando ad una visione più ampia del rapporto tra la fiducia stessa e l'ecosistema digitale è possibile rilevare come il legislatore europeo avesse già sottolineato l'importanza di tale valore nell'ambito della "protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati".

Difatti, il considerando n. 6 del **Regolamento n. 2016/679 del Parlamento europeo e del Consiglio del 27 aprile 2016 –** GDPR – sul presupposto che "la rapidità dell'evoluzione tecnologica e la globalizzazione comportano nuove sfide per la protezione dei dati personali", che "la portata della condivisione e della raccolta di dati personali è aumentata in modo significativo" e che "la tecnologia attuale consente tanto alle imprese private quanto alle autorità pubbliche di utilizzare dati personali, come mai in precedenza, nello svolgimento delle loro attività" rileva come "sempre più spesso, le persone fisiche rendono disponibili al pubblico su scala mondiale informazioni personali che le riguardano" cosicché – considerando n. 7 – "tale evoluzione richiede un quadro più solido e coerente in materia di protezione dei dati nell'Unione, affiancato da efficaci misure di attuazione, data l'importanza di creare il clima di fiducia che consentirà lo sviluppo dell'economia digitale in tutto il mercato interno".

In conclusione, il Regolamento ribadisce l'opportunità "che le persone fisiche abbiano il controllo dei dati personali che le riguardano e che la certezza giuridica e operativa sia rafforzata tanto per le persone fisiche quanto per gli operatori economici e le autorità pubbliche".

> L’analisi del GDPR consente di affermare che la “fiducia” nell’ecosistema digitale è qualcosa di più rispetto al solo contesto economico e commerciale ed è al significato fondante di questa parola che, con il presente lavoro, si vuole prestare attenzione, onde il punto di vista economico e del mercato sia una delle declinazioni possibili dell’affidabilità della persona fisica nei rapporti telematici “tout court”, caratterizzati tendenzialmente dall’assenza di relazioni fisiche e, per questo, idonei a generare maggiore attenzione e cautela da parte dei protagonisti del “villaggio globale”.

La fiducia nel rapporto tra il singolo e l’ecosistema digitale: il regolamento europeo AI Act

Il Regolamento n. 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, dedicato all’armonizzazione delle regole sull’intelligenza artificiale, al considerando n. 1 specifica lo scopo del Regolamento stesso, individuato non solo nel miglioramento del “funzionamento del mercato interno istituendo un quadro giuridico uniforme in particolare per quanto riguarda lo **sviluppo, l’immissione sul mercato, la messa in servizio e l’uso di sistemi di intelligenza artificiale**”, ma anche nella promozione della “**diffusione di un’intelligenza artificiale (IA) antropocentrica e affidabile**, garantendo nel contempo un livello elevato di protezione della salute, della sicurezza e dei diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell’Unione europea («Carta»), compresi la democrazia, lo Stato di diritto e la protezione dell’ambiente”.

> In linea con le richiamate finalità, il Regolamento si propone lo sviluppo dell’intelligenza artificiale – definita “famiglia di tecnologie in rapida evoluzione che contribuisce al conseguimento di un’ampia gamma di benefici a livello economico, ambientale e sociale nell’intero spettro delle attività industriali e sociali” (considerando n. 4) – e del suo quadro normativo “conformemente ai valori dell’Unione sanciti dall’articolo 2 del trattato sull’Unione europea (TUE), ai diritti e alle libertà fondamentali sanciti dai trattati e, conformemente all’articolo 6 TUE, alla Carta”, promuovendo una tecnologia antropocentrica della stessa intelligenza artificiale, da intendersi quale “strumento per le persone, con il fine ultimo di migliorare il benessere degli esseri umani”.

Il tutto “in considerazione dell’impatto significativo che l’IA può avere sulla società e della necessità di creare maggiore fiducia” (considerando n. 6).

Dunque, **la parola “fiducia” viene declinata non solo in riferimento al “mercato”** – quale fiducia nelle transazioni di natura economica – **ma quale aspetto fondante della “persona”** e del benessere della stessa.

L’intelligenza artificiale generativa nel regolamento europeo IA Act

Il significato attribuito al termine è confermato dall’analisi di altri “considerando” nei quali si articola la parte introduttiva del Regolamento: in particolare, ai fini del tema oggetto del presente lavoro, assume rilievo il **considerando n. 133 che affronta la questione dei sistemi di IA in grado di generare “grandi quantità di contenuti sintetici, che per gli esseri umani è divenuto sempre più difficile distinguere dai contenuti autentici e generati da esseri umani”**.

> Orbene, in relazione a tali sistemi il Regolamento europeo osserva come “l’ampia disponibilità e l’aumento delle capacità” degli stessi abbiano “un impatto significativo sull’integrità e sulla fiducia nell’ecosistema dell’informazione, aumentando i nuovi rischi di cattiva informazione e manipolazione su vasta scala, frode, impersonificazione e inganno dei consumatori” cosicché, prosegue il “considerando”,

> “Alla luce di tali impatti, della rapida evoluzione tecnologica e della necessità di nuovi metodi e tecniche per risalire all’origine delle informazioni, è opportuno imporre ai fornitori di tali sistemi di integrare soluzioni tecniche che consentano agli output di essere marcati in un formato leggibile meccanicamente e di essere rilevabili come generati o manipolati da un sistema di IA e non da esseri umani”.

Com’è di tutta evidenza il **Regolamento fa riferimento a quella forma di intelligenza artificiale definita “generativa”**, per tale intendendosi “qualsiasi tipo di intelligenza artificiale in grado di creare, in risposta a specifiche richieste, diversi tipi di contenuti come testi, audio, immagini, video”[1].

L’obiettivo del Regolamento non è quello di vietare l’utilizzo di siffatti sistemi ma, nell’ottica di **preservare la fiducia dell’utente nei confronti dell’ecosistema digitale e dell’informazione**, di imporre ai fornitori di quelli “che generano contenuti audio, immagine, video o testuali sintetici” la garanzia “che gli output del sistema di IA siano marcati in un formato leggibile meccanicamente e rilevabili come generati o manipolati artificialmente” (cfr. articolo 50 comma 2 del regolamento).

Tale obbligo riguarda sia i sistemi in grado di generare testi “allo scopo di informare il pubblico su questioni di interesse pubblico” sia quelli in grado di generare o manipolare una immagine o un contenuto audio o video che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona”: il regolamento, in questa parte, si occupa di disciplinare quei fenomeni che possono determinare la nascita delle c.d. “fake news” nonché del “deep fake”.

“Fake news” e “Deep fake”

Sul punto è bene precisare **che sia le “fake news” che il “deep fake” hanno, in comune, la “falsità” della informazione**, intesa in senso ampio, caratteristica che si ritrova proprio nel termine “fake” che rinvia alla falsificazione ed alla contraffazione.

Difatti per fake news si intendono “notizie false o tendenziose, diffuse attraverso i media tradizionali o i social media on line” e la cui rilevanza “è aumentata nel mondo attuale della «post-verità»”[2]mentre il “deep fake” è definito dall’articolo 1 comma 1 n. 60 dello stesso regolamento come “un’immagine o un contenuto audio o video generato o manipolato dall’IA che assomiglia a persone, oggetti, luoghi, entità o eventi esistenti e che apparirebbe falsamente autentico o veritiero a una persona”.

> Fatta questa premessa è da rilevare che, mentre con riferimento ai sistemi di intelligenza artificiale idonei a generare o manipolare testi “allo scopo di informare il pubblico su questioni di interesse pubblico” il regolamento non fa alcun riferimento esplicito alla “falsità” della informazione (le fake news), con riferimento a quei sistemi di intelligenza artificiale idonei a generare o manipolare immagini o contenuti audio o video il regolamento riconosce espressamente che tali contenuti apparirebbero falsamente autentici o veritieri ad una persona (il deep fake).

Orbene, a norma dell’articolo 50 comma 4 del Regolamento sia nel caso di **intelligenza artificiale che genera o manipola testi,** sia nel caso di **intelligenza artificiale che genera o manipola immagini o contenuti audio o video,** i deployer di tali sistemi rendono noto che il testo o il contenuto “è stato generato o manipolato artificialmente”.

“Fake news” e “deep fake” oltre le prescrizioni formali: una questione culturale

Nella prospettiva della incidenza che i sistemi generativi possono avere sull’affidabilità dell’ecosistema digitale, da intendersi non solo nella sua dimensione economica ma anche come contenitore di informazioni – il già ricordato ecosistema dell’informazione – assumono particolare rilievo le considerazioni contenute nella **“Relazione sull’attività svolta” della “Commissione per**

l’intelligenza artificiale per l’informazione” istituita presso la Presidenza del Consiglio dei Ministri, Dipartimento per l’informazione e l’editoria.

In tale relazione si legge che la Commissione, in attuazione dei propri compiti istituzionali – premesso che “nel mondo dell’informazione l’intenzione, lo sviluppo e l’utilizzo dell’innovazione condizionano la natura e la ricezione del messaggio, la percezione del cittadino e il processo di formazione della coscienza collettiva” – **ha individuato alcuni ambiti “di impatto strategico”** fra i quali figura non solo la “trasparenza, riconoscibilità e tracciabilità dei contenuti prodotti dall’IA generativa tramite certificazione standardizzata”, ma anche il “contrasto alla disinformazione”.

> In tale prospettiva la Commissione ha avanzato il timore “che **l’IA generativa ne determini la proliferazione**” posto che questi sistemi “possono essere utilizzati per dare visibilità o addirittura creare notizie false, difficili da distinguere da quelle vere, alimentando disinformazione e propaganda” ovvero “per creare immagini o video falsi, i c.d. «**deep fake**», che possono divenire strumento di truffe, ricatti, campagne diffamatorie e altre manipolazioni”.

La Commissione sottolinea **il fenomeno dell’utilizzo massivo dei sistemi di intelligenza artificiale** “per personalizzare i contenuti in base alle preferenze individuali degli utenti” che “può, inoltre, causare una maggiore polarizzazione dell’opinione pubblica, alimentando divisioni e conflitti”[3]: l’utilizzo di sistemi di intelligenza artificiale generativa nel campo della informazione amplifica il **fenomeno delle “echo chambers”** che raffigurano metaforicamente delle “casse di risonanza” e rappresentano degli spazi limitati in cui si aggregano degli individui aventi idee e/o credenze simili dando luogo a delle comunità polarizzate.

La formazione delle **echo chambers** è favorita dai meccanismi di funzionamento delle piattaforme on line (algoritmi di ricerca o news feed) che consentono all’utente di personalizzare le notizie da visualizzare, così da costruire i social media in modo tale da sfruttare la tendenza degli individui a interagire con soggetti ideologicamente vicini ed escludendo invece coloro che hanno un punto di vista diverso[4].

> La disinformazione, dunque, non è solo “informazione falsa” ma anche aggregata e selezionata per soddisfare le esigenze dei singoli utenti ai quali viene sottratta la possibilità di un confronto critico con punti di vista differenti: il tutto avviene con la “complicità” dei c.d. bias cognitivi che finiscono per replicare, nell’ecosistema digitale, percezioni errate o deformate, pregiudizi e ideologie, errori cognitivi che impattano nella vita di tutti i giorni, non solo su decisioni e comportamenti, ma anche sui processi di pensiero[5].

La soluzione proposta dalla Commissione per risolvere i problemi posti all’applicazione dei sistemi di IA al settore dell’informazione **va oltre l’apposizione dei marcatori che consentono di riconoscere un contenuto come generato da tali sistemi e guarda a colui che, in genere e nel settore delle nuove tecnologie, viene definito l’anello debole, ossia la persona**[6]: in questa prospettiva nella relazione è sottolineato **il ruolo strategico ricoperto dal “supporto alla formazione continua dei professionisti dell’informazione**, per lo sviluppo di specifiche competenze digitali, che consentano di fare fronte all’emersione delle nuove figure professionali, nonché di sfruttare l’intelligenza artificiale per amplificare le capacità del giornalista e migliorare quindi la qualità e l’offerta dei prodotti dell’informazione”.

La **promozione di attività di *awareness*** destinata ai professionisti dell’informazione – prosegue la Commissione – “consentirebbe loro una corretta rappresentazione delle capacità e dei limiti dei modelli di IA, ossia, in altri termini, una corretta rappresentazione delle conclusioni effettivamente deducibili dai processi di generazione ed analisi delle informazioni, determinando, dunque, anche una

possibile mitigazione degli effetti dei *bias* insiti nel modello o introdotti dall'utilizzatore"[7].

Conclusioni

Dalle considerazioni che precedono emerge che **la supervisione dell'uomo sui sistemi di intelligenza artificiale**, anche generativa, costituisce un aspetto fondamentale per evitare di affidare ad un algoritmo, il cui "ragionamento" risulta, almeno in parte, incomprensibile agli stessi cultori della materia[8], non solo la creazione ma la stessa selezione del patrimonio informativo ricco e variegato che un corretto ecosistema dell'informazione è in grado di fornire.

In tal modo sarà possibile **garantire meno diffidenza dell'utente nell'ecosistema della informazione a vantaggio di una maggiore fiducia** idonea a generare il circolo virtuoso della crescita sociale dell'individuo.

NOTE

[1] Treccani, definizione di "Intelligenza artificiale generativa" (consultato il 24 luglio 2024)

[2] G. Suffia, voce "Fake News" in AAVV, Dizionario Legale Tech, Milano, pagg. 415 ss.

[3] Su tale aspetto cfr. i risultati della ricerca sull'impatto dei sistemi di IA nel campo della informazione contenuti nel "Quarto rapporto ital communication IISFA sull'intelligenza artificiale in Italia del 24 luglio 2024", pagg. 17 ss

[4] Alba Calia, Voce "Echo chambers" in AAVV, Dizionario Legale Tech, Milano, pagg. 374 s.

[5] Sui "bias cognitivi" cfr. B. Fiammella, Intelligenza Artificiale, euristica e bias cognitivi applicati agli algoritmi, in Altalex si veda qui (consultato il 26 luglio 2024)

[6] Sul concetto di "anello debole" nel settore della sicurezza informatica si sono sviluppate le considerazioni inerenti il c.d. "fattore umano"

[7] Sull'aspetto inerente l'importanza della formazione nella diffusione dei sitemi di IA cfr quanto scritto nella "Strategia italiana per l'intelligenza artificiale 2024-2026" del Dipartimento per la trasformazione digitale e dell'Agenzia per l'Italia digitale AGID, pagg. 29 ss.

[8] Si pensi, sul punto, al fenomeno delle c.d. black box