Intelligenza Artificiale, regole e rischi: un nuovo capitolo per il Diritto contro gli illeciti digitali

Giuseppe Accardo, Ingegnere e Chief AI Transformation Officer, Pollicino AIdvisory
Riccardo Perlusz, Senior Advisor, Legal-Technological Policy Analyst, Pollicino AIdvisory

1. Introduzione

Siamo stati per lungo tempo felicemente convinti che il mondo digitale viaggiasse su binari leciti e con percorsi chiari. I primi decenni dell'informatica[1] sono stati infatti una sorta di *giardino dell'eden*, chiuso in una bolla tecnologica sconosciuta ai più, dove l'illecito era francamente impossibile e dove dati e programmi erano sigillati all'interno di imponenti e inaccessibili *centri di calcolo*.
Questa convinzione, con l'arrivo della rete Internet e dell'informatica distribuita, con la sovrapposizione fra sfera privata, servizio pubblico e aziende, permessa da una tecnologia pervasiva e totalizzante, si è infranta contro le prime truffe online. La situazione si è resa ulteriormente complessa a seguito delle prime manipolazioni dei profili social, del controllo dei dati personali, dei comportamenti digitali aggressivi verso l'uno o l'altro attore del sistema globale, in cui tutto e tutti quotidianamente operavamo, spesso privi di tutele e di attenzioni.
Con lentezza, ma in continua progressione, le norme e le regole si sono adattate ai nuovi strumenti digitali, formando il *corpus iuris* dell'informatica giuridica che ha disciplinato i nuovi negozi giuridici, la validità delle prove digitali e ha identificato comportamenti illeciti con nuove fattispecie penali. La disciplina del digitale è quindi entrata a pieno titolo nel mondo giuridico.
Qui citiamo come antesignano di questa nuova fase, la normazione dei contratti perfezionati online che è stata regolamentata a partire dalla direttiva europea 2000/31/CE sulla comunicazione elettronica[2] poi recepita tramite la Lg. n.246 del 14 novembre 2005 e, per quanto riguarda i reati penali legati alla rete Internet, l'introduzione delle prime normative specifiche per contrastare crimini informatici, le truffe online e altri reati digitali.
Queste leggi si sono poi evolute nel tempo, seguendo la crescente diffusione della rete, in una nuova stagione in cui il diritto ha iniziato a disciplinare molti aspetti delle tecnologie digitali, arrivando a ipotizzare fra i diritti dei cittadini europei, anche quelli digitali[3].

[1] Gli autori si riferiscono al periodo antecedente alla diffusione delle tecnologie informatiche delle reti Internet.
[2] Nota come *Direttiva sul Commercio Elettronico* – Dir . 2000/31/CE 8 giugno 2000
[3] Si veda Dichiarazione europea sui diritti e i principi digitali per il decennio digitale (2023/C 23/01) in https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32023C0123(01)

2. L'avvento dell'IA

L'elemento di rottura della progressione fin qui descritta risiede nell'Intelligenza Artificiale, quale fattore di accelerazione tecnologica, che introduce oggi un ulteriore mutamento strutturale in questo modello di rischio, legato all'autonomia *"agentica"* dei processi IA, alla possibilità di effettuare parallelamente innumerevoli processi, consentendo di processare enormi quantità di dati disponibili a tempo zero tramite le infrastrutture della rete. Elementi questi, che portano ad un livello molto più alto ed ampio il rischio di illeciti, di comportamenti aggressivi o di azioni di controllo che queste tecnologie consentono di effettuare in modo più diretto e mirato.

Assunto che ognuno di noi è esposto costantemente nell'universo della rete Internet, i cui contenuti sono spesso (a torto) da noi stessi considerati una realtà parallela attendibile e sicura per la nostra esistenza sociale, il pericolo di trovarsi coinvolti, o peggio, vittime di questi illeciti è oggettivamente sempre più elevato.

Oggi acquisiamo informazioni, entriamo in contatto con soggetti terzi, cerchiamo e valutiamo prodotti, acquistiamo ed effettuiamo transazioni finanziarie quasi prevalentemente attraverso la rete internet, utilizzando le infrastrutture dei data centers e agendo con applicazioni informatiche che sempre più spesso alla base utilizzano le architetture applicative complesse dell'Intelligenza Artificiale e della gestione dei dati tramite applicativi Big Data del cyberspazio, ovvero in giurisdizioni incerte e nel labirinti del Multi-Cloud[4].

Ma proprio per le sue potenzialità tecniche, l'IA può essere uno strumento capace di manipolare l'informazione e quindi l'opinione pubblica, diventando un efficace mezzo di disinformazione e propaganda tramite la generazione massiva e automatizzata di *fake news*, articoli o notizie apparentemente giornalistiche, post sui social, generati tramite gli automatismi dei *modelli di LLM*[5] con linguaggio naturale, credibile e usando una qualsiasi lingua o idioma. Queste operazioni consentono di influenzare appartenenze o elezioni politiche, scelte dei mercati finanziari, valutazioni sulla reputazione di personalità o di aziende ed in tal senso vengono impropriamente utilizzate. La capacità di processo dei motori IA consente oltremodo di creare migliaia di account "credibili", simulando discussioni spontanee, amplificando artificialmente i trend di prodotti o modificando le curve di crescita o penetrazione di prodotti, falsificando quello che dallo scorso secolo veniva indicato come "opinione pubblica". L'impatto sugli utenti della rete è, come detto, profondo, poiché queste tecnologie vengono anche utilizzate per inquinare il dibattito democratico e polarizzare le società.

[4] Si riferisce a un'architettura cloud avanzata che integra cloud pubblici, privati e ibridi, utilizzando tecnologie emergenti di più fornitori come automazione, intelligenza artificiale e iperconvergenza, più facilmente gestibili attraverso l'interoperabilità tra diverse piattaforme geografiche e distribuite.

[5] In sintesi, un LLM è un modello linguistico di grandi dimensioni che elabora e produce testo in modo autonomo, basandosi su pattern appresi dai dati.

3. Casi di studio

L'IA agisce su tre direttrici: amplifica la potenza e la capacità operativa, semplifica il livello di conoscenza necessario per accedere a questi strumenti dell'illecito e permette di creare infrastrutture applicative specializzate.

Alcuni esempi. Recenti casi dimostrano come siti definiti *"Pink Slime"* ovvero siti progettati per attirare traffico senza offrire un reale valore dando false notizie, titoli sensazionalistici o contenuti copiati illecitamente da altre fonti e che generano ricavi pubblicitari attirando traffico di visitatori, utilizzino l'IA per fabbricare velocemente deepfake o manipolare informazioni attraverso la diffusione di notizie artefatte. La sfida principale risiede appunto nella velocità di propagazione di questi falsi contenuti generati dall'IA che possono inondare le piattaforme della rete, prima ancora che i meccanismi di *fact-check*ing riescano a intervenire, rendendo l'informazione un terreno di scontro algoritmico, dove la verità diventa sempre più difficile da discernere. In questo scenario, l'IA risulta purtroppo vincente: la rapidità degli algoritmi nel reagire a notizie sintetiche può innescare distorsioni dei prezzi immediate, provocando gravi perdite finanziarie e minando la stabilità e l'integrità dei mercati globali.

Le reazioni sono spesso tardive e lente e l'incapacità dell'utente medio di discernere verità da menzogne rende lo strumento particolarmente insidioso.

Analoghi strumenti hanno invece la potenzialità di manipolare i mercati digitali e finanziari[6], tramite la produzione coordinata di falsi ma credibili contenuti che promuovono un titolo in borsa, una valuta *crypto* o un asset quotato, generando distorsione dei prezzi e danni agli investitori retail. In questo caso le tecnologie AI consentono di analizzare contestualmente molteplici flussi di dati, generando flussi paralleli di recensioni artefatte, su scala anche globale, realizzando meccanismi automatici di concorrenza sleale che riescono a incidere in modo artefatto sui meccanismi logaritmi che regolano la domanda dei mercati.

Altri strumenti IA entrano nell'armamentario aggressivo di sodalizi criminali. Pratiche come l'*"AI Washing*", che consistono nel creare meccanismi di sopravalutazione di prodotti, attribuendone falsamente contenuti o componenti in tecnologie AI avanzate o azioni di *"Pump and Dump"* cioè particolari illeciti realizzati facendo salire artificialmente nei mercati il prezzo di assets tramite una diffusione di informazioni false per poi rivenderlo in tempo utile a lasciando ad altri investitori le perdite legate alla svendita del titolo, sono oggi azioni coordinabili

[6] Si veda" *BIS -Financial stability implications of artificial intelligence - Executive Summary", 26 giugno 25 scaricabile in* https://www.bis.org/fsi/fsisummaries/exsum_23904.htm?utm_source=chatgpt.com

tramite applicativi evoluti di Intelligenza Artificiale in grado di operare da una rete di server *"botnet",* infrastrutture utilizzate illecitamente dopo averne preso il controllo tramite accessi illegali da parte degli stessi *cybercriminali*. Queste azioni saturano i contenuti dei social e le piattaforme di trading con informazioni artefatte con contenuti e false informazioni, generando un'illusione che spinge gli investitori retail verso acquisti rischiosi o vittime prescelte, verso operazioni finanziarie o acquisti incauti.

Secondo il Rapporto 2023 della Federal Trade Commission (FTC) degli Stati Uniti[7], le frodi online hanno raggiunto cifre da capogiro, con oltre 8 miliardi di dollari di perdite segnalate solo negli Stati Uniti. Questo numero è in continuo aumento ogni anno, con una crescita esponenziale delle frodi legate a phishing, scam telefonici, truffe finanziarie e furto d'identità. A livello mondiale, secondo il Centro di Analisi e Condivisione delle Informazioni sulla Sicurezza Informatica (CISA), le perdite globali dovute alle frodi online si aggirano intorno ai 30 miliardi di dollari all'anno. Le truffe online colpiscono principalmente le persone che non sono consapevoli delle tecniche più comuni, come le e-mail phishing, i falsi negozi online e i ransomware. Nel 2022 un fornitore primario di servizi social avrebbe stimato che oltre il 5% degli account attivi su Facebook risultavano falsi o duplicati, che equivarrebbe a oltre 100 milioni di account. Inoltre, si stima che probabilmente una notizia su sette circolante in rete non avrebbe in verità alcun fondamento. Infine, il MIT di Boston ha invece rilevato che la velocità di propagazione in rete di *fake news* è ben sei volte maggiore della circolazione di notizie vere.

È ormai evidente che l'ambito delle frodi, truffe ed in generale del cybercrime è oggi nell'era dell'IA più ampio, aggressivo e penetrante. Con l'informatica distribuita abbiamo di fatto aperto le nostre case a nuove categorie di criminali, consentendo a soggetti spesso irrintracciabili di operare a nostro danno da paesi remoti ed impermeabili alle azioni di contrasto delle autorità.

Qui le tecniche sono quelle del *phishing avanzato* e *social engineering*, tramite e-mail o messaggi personalizzati, che tuttavia l'IA ha reso più insidiosi e credibili perché privi di errori linguistici, moltiplicabili in migliaia di azioni parallele e contemporanee, spesso adattate automaticamente al profilo psicologico della vittima designata. Le infrastrutture aggressive utilizzate, definite *chatbot malevoli,* interagiscono spesso autonomamente in tempo reale con le vittime, utilizzano dati reperiti in rete direttamente riconducibili alla vittima. Il volume di fuoco è talmente ampio che in poche ore e prima che sia possibile attivare controreazioni, gli illeciti in numero rilevantissimo risultano perfezionati.

[7] Si veda https://www.ftc.gov

Ulteriori esempi recenti includono *"truffe vishing"* dove l'IA clona la voce di parenti per richiedere denaro in emergenza, o frodi aziendali tramite deepfake video di dirigenti che autorizzano bonifici milionari. Nel dark web, strumenti come *"WormGPT"* automatizzano la creazione di malware e pagine di phishing, mentre campagne di social engineering sfruttano l'IA per analizzare profili social e inviare esche personalizzate. Altri casi emersi riguardano attori statali che usano agenti IA per infiltrare curriculum falsi in aziende IT o generare propaganda coordinata su larga scala, rendendo l'attacco cyber più rapido, economico e difficile da rilevare. L'aumento drastico del numero e dell'efficacia delle truffe online dimostra come l'automazione di alcune attività criminali si sia evoluta soprattutto con il supporto tecnologico dei prodotti IA di scrittura e di traduzione dei contenuti, ma anche con la generazione di documenti falsi, incredibilmente realistici.

4. Analisi, dati e sorveglianza

La nostra presenza in rete non ci rende solo noti, ma anche sorvegliabili, controllabili e profilabili.

Le tecnologie IA consentono oggi la raccolta e analisi automatica e massiva di ampie informazioni provenienti dalle molteplici fonti della rete, aprendo alla possibilità di essere analizzati da prodotti IA in grado di effettuare una profilazione psicologica o politica. Questi automatismi, impossibili sino a qualche anno fa, consentono oggi in paesi totalitari, grazie ai nuovi algoritmi IA, l'identificazione automatica del dissenso e la generazione automatica di contro-narrazioni a favore delle proprie politiche dominanti, raggiungendo target a livello planetario e non solo all'interno delle proprie giurisdizioni.

La tecnologia IA rende oggettivamente la rete e l'intero cyberspazio un ambito di maggior rischio, abbassando il costo di operazioni, il tempo di esecuzione e le competenze richieste per effettuarle. Ciò che fino a ieri richiedeva team specializzati, tempi lunghi per la raccolta e l'analisi dei dati e la pianificazione delle operazioni, ora tutto questo può essere fatto con impegni tecnologici e investimenti ridotti.

Dietro la nebbia dell'etica IA che abbiamo alzato per discutere di Intelligenza Artificiale, è evidente che si possano anche celare pericoli ben più attuali e prossimi alle nostre vite e realtà quotidiane. La rete globale internet ha messo le nostre esistenze digitali in vetrina e la tecnologia IA sta potenzialmente dando nuovi strumenti per poter realizzare attività sempre più sofisticate e penetranti. Questo nel bene e nel male.

Alcuni esempi ne dettagliano la portata. Il 23 aprile 2025 Anthropic pubblica un report su abusi rilevati (*Uso "illecito/non etico")*, includendo un'operazione mirante

ad influenzare scelte commerciali tramite il modello IA Claude non solo per generare contenuti, ma anche per orchestrare il comportamento di *bot social* utilizzati per decidere quando mettere like, commentare, condividere, tramite false identità digitali. Anthropic avrebbe poi identificato la fonte e inibito le utenze. Nello stesso report vengono citati anche casi di *recruitment fraud, ovvero* truffe nei processi di false procedure di assunzione, con modelli più convincenti per le vittime grazie a riscrittura linguistica dei LLM. Sono inoltre noti tentativi legati all'uso di credenziali generate da IA per l'accesso non autorizzato a dispositivi IoT spesso da soggetti con ridotta competenza ed esperienza di sviluppo ma che hanno utilizzato i modelli generativi pubblici di IA per generare codici o procedure applicative.

In tal senso il 27 agosto 2025 Anthropic ha pubblicato un *report di threat intelligence* (n.d.r. documento che raccoglie, analizza e presenta informazioni dettagliate sulle minacce informatiche a cui un'organizzazione, un settore o una nazione può essere esposta) che descrive una operazione di estorsione basata su dati sottratti in cui viene citato l'uso di codice applicativo realizzato tramite LLM Claude Code; il caso indica almeno diciassette organizzazioni colpite e richieste di riscatto che talvolta superavano 500.000 dollari (con minaccia di esposizione pubblica dei dati) dimostrando come lo strumento IA consenta oggi la estorsioni su larga scala e la possibilità di abbassare la barriera tecnica di accesso al cybercrime, rendendo anche imprudenti adolescenti, pericolosi cybercriminali.

In ultimo, il 13 novembre 2025 sempre Anthropic pubblica un case study su quella che definisce la prima campagna di "*cyber spionaggio orchestrato tramite IA*" dove il codice generato tramite il LLM di Claude è stato impiegato per svolgere gran parte del lavoro tattico di ricognizione, individuazione vulnerabilità. Di particolare interesse notare come Anthropic abbia stimato che l'AI abbia eseguito circa 80–90% del lavoro con supervisione umana limitata.

Di fronte a questo scenario tanto complesso quanto preoccupante, risulta ancora incompleto il quadro normativo a contrasto, dove i nuovi illeciti digitali con strumenti IA trovano oggi unicamente norme indirette e non contestualizzate, ovvero prive di idonee misure aggravanti, centrate sullo strumento delle nuove tecnologie usate nel reato.

Complessa e onerosa oltremodo la formazione e l'accertamento delle prove, con un aggravio aggiuntivo nell'iter investigativo e processuale. Algoritmi con strutture proprietarie, impenetrabili, applicativi eseguiti in reti di infrastrutture Hypercloud sono una complessità analitica e investigativa non di poco conto.

5. Prospettive future

Da pochi mesi l'Italia ha approvato una legge organica sull'IA (Lg. 132/2025 entrata in vigore il 10 ottobre 2025), che introduce la nuova fattispecie penale dell'art. 612-quater c.p. che punisce l'illecita diffusione di contenuti generati o alterati da IA (deepfake e materiali manipolati), con pene edittali da 1 a 5 anni di reclusione. Da notare che rispetto all'originaria versione della fattispecie, che ne prevedeva una pena massima di tre anni, la pena edittale dell'articolo novellato ne consente oltremodo la misura cautelare detentiva.

Questa legge è attuata in raccordo al Regolamento AI Act europeo e costituisce un riferimento normativo per l'uso lecito dell'IA. La medesima norma ha peraltro fornito al governo alcuni poteri legislativi per formare decreti attuativi e per la novazione delle fattispecie penali esistenti, con l'obiettivo di adeguarle allo scenario di un uso dell'IA a supporto di comportamenti illeciti. Il nostro paese è risultato fra i più veloci nel muoversi verso una disciplina penalistica specifica. Oggi più che mai è opportuno che il legislatore e gli uffici competenti sappiano agire con opportuno tempismo e con precisione, colmando prima possibile le lacune normative esistenti, a tutela dei *cittadini-utenti* della rete. Tuttavia, un'azione efficace nel mondo globalizzato del digitale non può prescindere da una tutela quanto meno geografica e transnazionale, il che rende tutto molto complesso. In un mondo dove l'intelligenza artificiale interagisce già in molte delle nostre attività e relazioni, è sicuramente indispensabile che il Diritto disciplini opportunamente i nuovi comportamenti per poter effettivamente bilanciare innovazione e tutela dei diritti fondamentali, garantendo una rete sicura, etica e responsabile per tutti. È assodato che solo attraverso una legislazione completa, chiara, aggiornata e pronta a fronteggiare le nuove dinamiche digitali, sia realmente possibile proteggere i cittadini da manipolazioni, frodi e abusi nell'era dell'IA, garantendo al contempo libertà, progresso e soprattutto fiducia nelle tecnologie digitali.

Il diritto deve accompagnare la tecnologia. Comprendere, governare e contrastare gli usi illeciti dell'IA non è più un'opzione, ma una necessità, per preservare una società digitale equa, trasparente e sicura per tutti.

Pollicino Aldvisory
Via Visconti di Modrone 8
20122, Milano, Italy
info@pollicinoaidvisory.eu