

IA, GLOSSARIO MINIMO**ALLUCINAZIONE ARTIFICIALE**

L'allucinazione artificiale è un fenomeno tipico dei modelli di intelligenza artificiale generativa, come ChatGPT o DALLE, che consiste nella produzione di informazioni errate, inventate o fuorvianti. Questi sistemi, pur essendo addestrati su enormi quantità di dati, possono generare contenuti che sembrano plausibili ma che non hanno un riscontro nella realtà. Ad esempio, un modello di linguaggio potrebbe inven-



tare riferimenti bibliografici inesistenti o descrivere eventi mai accaduti. Le cause delle allucinazioni artificiali sono molteplici: dalla mancanza di dati sufficienti all'incapacità di verificare autonomamente le informazioni. Que-

sto problema è particolarmente critico quando l'IA viene utilizzata in settori sensibili come la medicina o il diritto. Le aziende tecnologiche stanno sviluppando soluzioni per ridurre il fenomeno, come l'integrazione di meccanismi di verifica e il miglioramento della qualità dei dataset di addestramento. Tuttavia, il problema delle allucinazioni artificiali sottolinea l'importanza del controllo umano sull'IA: è fondamentale che gli utenti siano consapevoli dei limiti di questi sistemi e utilizzino sempre un approccio critico.

Realtà visive distorte al tempo dell'IA: è sempre più difficile distinguere il vero dal falso

Bella foto, ma sarà fake?

di SEBASTIANO BATTIATO*

Seeing is Believing" addio! Con l'avvento delle recenti tecnologie basate appunto sulla cosiddetta Intelligenza Artificiale Generativa, sarà sempre più complicato, per gli esseri umani per loro natura intrinsecamente "analogici", riuscire a distinguere un contenuto multimediale "fake" da uno reale. Molteplici sono gli esempi che si susseguono, oramai ad un ritmo frenetico, in cui vuoi

tazionale utilizzata in ambito AI, per addestrare Reti Neurali Artificiali, cosiddette "profonde" cioè con un numero di parametri intrinseci particolarmente elevato e soprattutto ideata con la capacità di analizzare ed "imparare" strutture, pattern e correlazioni nascoste negli altrettanto numerosi dati di ingresso, forniti come esempio. La capacità di queste reti di apprendere da enormi quantità di dati non strutturati ha aperto la porta a nuove e impressionanti applica-

zioni, spingendo ulteriormente i confini di ciò che le macchine possono fare. Potremmo citare a tal scopo, proprio per rimanere su temi di cronaca recente "Now and Then" dei Beatles oppure la foto di Papa Francesco con indosso il piumino bianco, ecc. ma gli esempi sono destinati a crescere a tal punto che qualcuno profetizza che le future camere digitali dei nostri smartphone potrebbero essere sostituite "grossolanamente" da appositi AI-tools che generano immagi-

ni/video a comando. Nel nostro laboratorio abbiamo coinvolto il termine "Impostor Bias", proprio per indicare un bias cognitivo che ci porterà a dubitare sempre di più dei contenuti multimediali proposti dai media, perché potenzialmente frutto della cosiddetta AI-generativa.

Questo cambiamento potrebbe rivoluzionare non solo il campo della fotografia ma anche quello del cinema e della televisione, dove le immagini generate artificialmente potrebbero diventare la norma. I recenti rilasci di piattaforme sempre più stupefacenti non lascia adito a dubbi in tal senso. Senza dilungarsi sui dettagli tecnici che stanno alla base di tali "meraviglie", possiamo altresì sintetizzare il fatto che le moderne macchine basate sul cosiddetto Deep Learning generativo e più di recente grazie al supporto dei cosiddetti Foundational Models (stile Chat GPT per intendersi), sono in grado di analizzare quantità inimmaginabili di esempi e mediante tecniche "adversariali" o pseudo tali, generare contenuti "credibili", con una qualità percettiva mai vista prima.

La tecnologia AI ha quindi il potere di alterare radicalmente il nostro rapporto con la realtà mediata, sfidando i nostri processi di percezione e valutazione critica. Se a questo aggiungiamo la cosiddetta "democratizzazione della produzione audiovisiva", intendendo con ciò che i tools di sviluppo AI-based sono alla portata di tutti perché resi disponibili a basso costo, possiamo affermare che la certificazione della cosiddetta autenticità di un segnale multimediale diverrà sempre più attuale e complicata. E d'altra parte, anche quando palesemente artificiali, i nostri sistemi cognitivi, assorbono qualsivoglia contenuti visuali, che bypassando la parte razionale, fanno sì che si possano in qualche modo fissare nella nostra memoria inconscia. Si pensi per esempio al recente video pubblicato dall'entourage di Donald Trump, e alla sua visione distorta del territorio di Gaza, trasformato in resort di bassa lega che per quanto palesemente kitch e blasfema, comunque genera in noi sentimenti diversi ma comunque "reali" con un impatto emotivo più che tangibile.

È imperativo, quindi, sviluppare nuovi strumenti fo-

rensi e metodologie per riconoscere e autenticare i contenuti digitali, assicurando che la verità e la realtà non vengano compromesse. D'altra parte gli addetti ai lavori stanno già rilasciando opportuni adeguamenti alle best practice di settore, in cui in analogia a quanto fatto finora, si va alla ricerca con metodi tradizionali o con ulteriori tools basati sempre su AI, di tracce della "innaturalità" dell'informazione rappresentata nel segnale da analizzare. Per quanto le macchine siano in grado di

Se anche il kitch e il blasfemo generano sensazioni "reali"

La ricerca di rivelatori di tracce di innaturalità

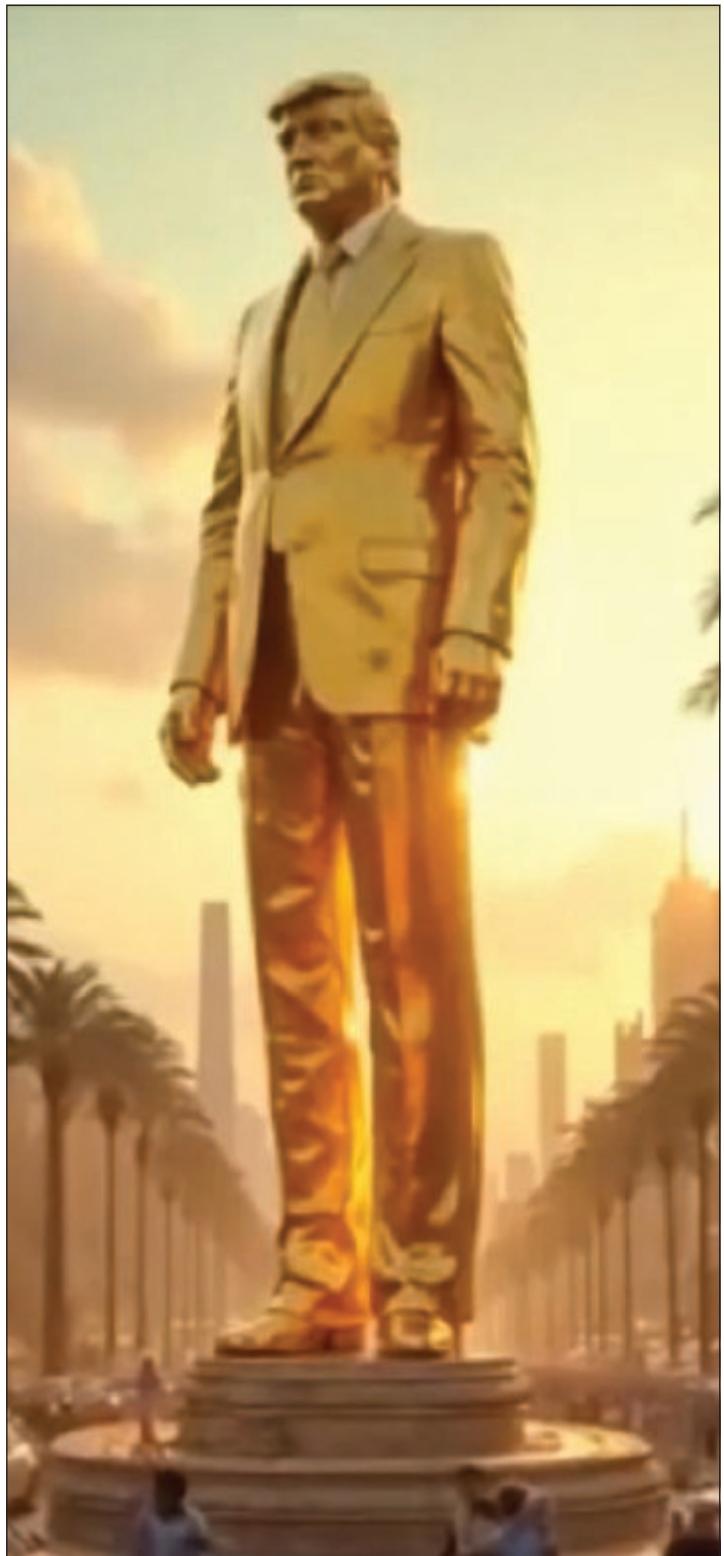
La necessità di trasparenza e utilizzo consapevole

ingannare il nostro vettore oculare biologico, non tutte le innamorate statistiche naturali sono riprodotte fedelmente per cui il buon caro vecchio investigatore/technico forense ante-litteram è ancora necessario, utile e determinante nella ri-

soluzione di diatribe sulla presunta autenticità di clip multimediali.

La continua evoluzione della AI, tuttavia, richiede un costante aggiornamento delle tecniche forensi, così come un impegno accademico e industriale per tenere il passo con le innovazioni tecnologiche. Il futuro ovviamente è da scrivere. Anche in questo caso però dobbiamo assolutamente ribadire come le scoperte scientifiche e la comunità accademica siano assolutamente da considerarsi come stabile approdo cui affidarsi per garantire l'utilizzo consapevole, trasparente e ove possibile sicuro e scevro da potenziali bias di sviluppo e/o dei dati utilizzati delle attuali e future macchine calcolatrici AI-based.

*Professore Ordinario di Informatica Università di Catania



sintetiche, volti animati perfettamente integrati e indistinguibili dal presunto originale. Recent studi hanno inoltre dimostrato come il famoso "Test di Turing", orientato a verificare se solo tramite chat fosse possibile decidere della presunta "umanità" dell'interlocutore, sia destinato a fallire nel breve termine. Questo test, a cui spesso ci si riferisce come il Gioco dell'imitazione (da cui il titolo del biopic film proprio su Alan Turing - The Imitation Game), non aveva previsto l'avvento di una tecnologia capace di simulare con tale precisione le interazioni umane. L'Intelligenza Artificiale ora non solo imita, ma può anche creare da nulla nuove realtà, rendendo ancora più sfumate le linee tra ciò che è reale e l'immaginario.

Il termine "Deep" sta a sottolineare la strategia compu-