PuntoSicuro

ARTICOLO DI PUNTOSICURO

Anno 27 - numero 5787 di Mercoledì 12 febbraio 2025

Il rapporto internazionale sulla sicurezza dell'intelligenza artificiale

Occorre un documento di ben 300 pagine per analizzare tutti i rischi legati all'utilizzo di applicazioni di intelligenza artificiale.

Un numeroso gruppo di esperti, provenienti da tutte le parti del mondo, ha pubblicato, a gennaio 2025, un rapporto internazionale sulla sicurezza degli applicativi di intelligenza artificiale. È un documento che merita un'attenta lettura da parte di tutti gli esperti di informatica, perché evidenzia situazioni di rischio, sino ad oggi forse non sufficientemente inquadrate e valutate.

Il documento mette in evidenza alcuni aspetti già noti, ed altri meno noti.

Tanto per cominciare, è indubbio che le prestazioni di applicativi di intelligenza artificiale siano cresciute in modo esponenziale negli ultimi tempi, ma al contempo sono cresciuti in maniera esponenziale i rischi collegati a tali applicazioni.

Il documento mette in evidenza come molti soggetti, che hanno deciso di investire in applicativi di intelligenza artificiale, non hanno inquadrato correttamente i rischi connessi all'utilizzo di questi applicativi. In particolare, maggiore è lo sviluppo degli applicativi, maggiori sono i rischi, messi in evidenza sia dopo un'analisi a freddo, sia dopo l'esame delle informazioni recuperate sul campo.

In sintesi, il documento è articolato in tre parti fondamentali:

- Cosa possono fare gli applicativi di intelligenza artificiale?
- Quali sono i rischi associati a questi applicativi?
- Quali tecniche di messa sotto controllo sono disponibili?

Non potendo in poche righe sintetizzare questo documento, disponibile in allegato, riteniamo opportuno evidenziare almeno i principali rischi che sono stati presi in considerazione e analizzati in profondità.

Vengono innanzitutto esaminati i rischi legati ad un uso criminoso dell'applicativo, che può portare a danni nei confronti di specifiche persone, per contenuti falsi, oppure alla manipolazione della pubblica opinione, o a rischi informatici o addirittura ad attacchi biologici chimici.

Altri rischi discendono dal fatto che l'applicativo potrebbe non funzionare correttamente, creando danni, anche di origine involontaria. Inoltre vi sono problemi legati all'affidabilità dell'applicativo, al fatto che l'applicativo potrebbe avere degli orientamenti sociali e politici, nei confronti di particolari categorie di persone, ed infine al fatto che sia possibile giungere a una perdita completa di controllo dell'applicativo, con conseguenze difficilmente valutabili.

La terza categoria di rischio fa riferimento ai rischi di tipo sistemico, che possono andare da un impatto negativo sul mercato del lavoro, fino a violazioni della privacy e danni all'ambiente.

Anche il rischio di violazioni del copyright dei testi, che vengono analizzati dagli applicativi di intelligenza artificiale, rappresenta un'area meritevole di approfondimenti.

Dopo questa analisi dei rischi, il documento passa ad analizzare le possibili misure di messa sotto controllo, offrendo indicazioni importanti sia per chi sviluppa questi applicativi, sia per chi li deve utilizzare.

In sintesi, un documento preziosissimo, che siamo ben lieti di mettere a disposizione dei nostri lettori, ringraziando le diecine di esperti, da tutte le parti del mondo, che hanno contribuito a compilare il documento.

International AI Safety Report - The International Scientific Report on the Safety of Advanced AI - January 2025 (pdf)

Adalberto Biasiotti