PuntoSicuro

ARTICOLO DI PUNTOSICURO

Anno 25 - numero 5468 di Lunedì 25 settembre 2023

Una valutazione critica degli applicativi di intelligenza artificiale

Gli applicativi di IA vengono spesso utilizzati per manipolare il profilo di un soggetto, come appare sui media. Un documento di ETSI offre una linea guida per la messa in sicurezza di questo aspetto critico degli applicativi di intelligenza artificiale.

Questo documento esplora le tecniche, basate su applicativi di intelligenza artificiale, che permettono di manipolare i profili di alcuni soggetti o di creare profili fasulli, presentati in differenti formati sui media, come ad esempio audio, video o testo. Queste alterazioni o creazioni dal nulla sono spesso battezzate con l'espressione inglese "deepfakes".

Il documento descrive i diversi approcci tecnici possibili e analizza le minacce, che sono presentate da questi profili contraffatti, in corrispondenza dei diversi scenari di attacco. Il documento successivamente offre misure tecniche e organizzative che possono mitigare queste minacce, analizzandone sia l'efficacia, sia le limitazioni.

Il documento, come di consueto, si apre con la indicazione di riferimenti normativi o informativi, nonché con un glossario, che aiuta a meglio comprendere l'intero testo.

Il documento passa quindi ad analizzare i metodi di contraffazione o creazione dal nulla di profili personali, analizzando innanzitutto gli aspetti video, che possono comportare la sostituzione del volto, la modifica del volto o la creazione sintetica di un volto.

Successivamente si analizzano gli aspetti audio e gli aspetti testuali. Infine, vengono analizzate le combinazioni di queste tre tecniche di contraffazione.

Vengono successivamente analizzati gli scenari di attacco, che possono portare ad influenzare la pubblica opinione, oppure a diffamare una persona o modificarne l'immagine, da un punto di vista sociale. Altri attacchi possono essere diretti alla autentica di una persona, in particolare ai metodi di autentica biometrica.

Si passano poi ad esaminare i dati necessari per effettuare manipolazioni video, audio e del testo, nonché gli strumenti che possono essere usati per queste stesse tre categorie di manipolazioni.

Di particolare importanza è il paragrafo dedicato modalità di individuazione delle tecniche di manipolazione.

Si passa quindi all'illustrazione delle tecniche di contromisure, classificate in base ai tre particolari tipi di attacchi, sopra illustrati.

Si tratta di un documento oltremodo prezioso frutto di un lungo studio di specialisti, che raccomandiamo a tutti i lettori di leggere con estrema attenzione.

Le indicazioni offerte in questo documento permettono di analizzare i profili riferiti a soggetti veri o a soggetti artificiosamente creati, per verificare la credibilità dei profili stessi.

ETSI GR SAI 011 V1.1.1 - Securing Artificial Intelligence (SAI); Automated Manipulation of Multimedia Identity Representations (pdf)

Adalberto Biasiotti