Arduo difendersi dall' ingerenza dell' IA

Data 05 settembre 2023
Categoria professione

Abbiamo gia' espresso i nostri dubbi circa il possibile uso fraudolento dell' Intelligenza Artificiale nelle attivita' umane, invocando disposizioni che rendano obbligatoria l' attribuzione dei vari contenuti sulle piattaforme web, e nello specifico, se siano autentici o falsi prodotti dall' IA. Ora persone piu' esperte di noi hanno sottolineato con maggior vigore queste esigenze.

A fine gennaio, OpenAI (la Societa' creatrice di CHAT-Gpt) ha annunciato di aver prodotto un classificatore IA avente funzione di riconoscere il testo generato dalle intelligenze artificiali, distinguendolo da un testo scritto invece da un essere umano.

Questo strumento, secondo i produttori, sarebbe utile per contrastare il fenomeno della diffusione di false informazioni tramite IA.

Purtroppo i risultati ottenuti finora si siano fin qui rivelati deludenti, per cui il progetto e' stato abbandonato o quanto meno sospeso in attesa di individuare strumenti tecnici piu' efficaci.
Non e' stato comunicato, pero', ne' la tempistica prevista ne' i settori interessati (immagini, filmati, testi?)

Attualmente quindi il programma finora studiato non è più disponibile a causa del suo basso tasso di precisione.

La stessa Casa produttrice riconosce esplicitamente la necessita' (da noi gia' sottolineata) di far capire agli utenti se gli audio o i contenuti visivi siano generati o no dall'intelligenza artificiale.

Va specificato che, fin dal debutto dell' attuale strumento (quello ritirato), ne erano stati resi noti i limiti, come ad esempio:

-Inaffidabilità con i testi molto brevi (sotto le 1.000 battute) e, talvolta, anche con quelli più lunghi;
Alcuni testi scritti da un essere umano potrebbero essere erroneamente etichettati come generati dall'IA;
-Funzionamento poco efficace con lingue diverse dall'inglese;
-Contenuti difficili da valutare; I testi prodotti da IA e successivamente modificati manualmente potrebbero non essere correttamente riconosciuti.

In definitiva, e' possibile riconoscere il testo generato dall'IA?
Forse in futuro lo si potrà fare con un ragionevole livello di precisione, ma non sembra una prospettiva a breve termine, considerando che nemmeno i creatori di ChatGPT ci sono finora riusciti.

Gli informatici suggeriscono paradossalmente di dar valore, per identificare il prodotto umano, alle sgrammaticature e ai refusi come indicatori unici di autenticita', pero' non e' impossibile che alla fine un algoritmo non impari a simulare anche quelli.

DanieleZamperini

http://www.pillole.org/public/aspnuke/news.asp?id=8076
http://www.scienzaeprofessione.it/public/nuke/modules.php?name=News&file=article&sid=2484
https://www.punto-informatico.it/impossibile-riconoscere-testo-generato-ia/