IA generativa e disinformazione: progressi, sfide e opportunità

5 Marzo 2024

Quattro progetti europei uniscono le forze contro la disinformazione con modelli avanzati di Intelligenza Artificiale. Disponibile online il nuovo libro bianco dell'osservatorio europeo sui media digitali

In tutta Europa e nel mondo gli sforzi nel campo della ricerca stanno portando allo sviluppo di modelli di IA all'avanguardia per il rilevamento e l'analisi della disinformazione online. In questo contesto, FBK gioca un ruolo di leadership. Col coordinamento di Riccardo Gallotti, principal investigator, il progetto AI4TRUST, finanziato dall'Unione Europea nell'ambito del programma Horizon Europe (HORIZON-CL4-2021-HUMAN-01-27 AI to fight disinformation) è stato avviato un anno fa e nei prossimi due anni mira alla costruzione di una piattaforma che combinerà l'apporto dell'intelligenza artificiale e quello di esperti fact-checker.

Lo scopo del sistema sarà monitorare diversi social media e fonti di informazione in tempo pressoché reale, utilizzando algoritmi IA all'avanguardia, per analizzare testi, audio e video in otto lingue diverse. Questo strumento permetterà di selezionare i contenuti con un alto rischio di disinformazione in modo da segnalarli per la revisione a fact-checker professionisti, il cui contributo fornirà ulteriori riscontri per il miglioramento continuo degli algoritmi impiegati. Saranno inoltre sviluppati report adattati alle esigenze di chi lavora nei media, col fine di fornire informazioni affidabili per prevenire la diffusione incontrollata della disinformazione.

Dentro questo percorso si inserisce la collaborazione con altri progetti che, con differenti focus, competenze e tecnologie, stanno affrontando lo stesso problema. Dalla condivisione fra **AI4Trust** e altri tre progetti, **AI4Media**, **vera.ai** e **Titan**, anch'essi finanziati dal programma Horizon Europe, è nata una pubblicazione sull'IA generativa e la disinformazione. Si tratta di un documento agile che in meno di quaranta pagine analizza il ruolo dell'IA nella creazione della disinformazione, le tecnologie di rilevamento impiegabili e le sfide etiche e legali che presenta.

Il libro bianco dal titolo "Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities" e a cui ha contributo anche Riccardo Gallotti è stato pubblicato nel mese di febbraio 2024 ed è disponibile online sul sito web dell'osservatorio europeo sui media digitali (European Digital Media Observatory).

Il testo parte da una classificazione sistematica della disinformazione generata sinteticamente, descrivendone le tipologie (testi, audio, immagini e video generati dall'IA), la prevalenza e l'impatto sulle elezioni.
Prosegue discutendo i recenti progressi nel rilevamento della disinformazione generata sinteticamente.
Presenta alcuni strumenti basati sull'IA selezionati per assistere i professionisti della verifica delle informazioni (fact checkers) e i cittadini.
Nel seguito si concentra sulle questioni etiche e legali emergenti, all'intersezione tra disinformazione e IA generativa, in particolare prendendo in esame il concetto di inquinamento dei dati e dell'informazione, il cui sovraccarico rappresenta un rischio di portata globale, e in questo senso quest'anno rappresenta un banco di prova importante visto che circa metà della popolazione mondiale, oltre 4 miliardi di persone, sarà chiamata alle urne.
Infine, oltre a fornire una ricchissima bibliografia in tema per approfondire, descrive le sfide da affrontare:

- IA generativa, allucinazioni e qualità dei dati formati/provenienti da modelli linguistici di grandi dimensioni (LLM)
- Superare la fiducia infondata dei cittadini nell'IA
- Sviluppo di nuovi strumenti per individuare i contenuti generati dall'IA
- Oltre l'inglese: sono necessari nuovi strumenti di rilevamento multilingue
- Accesso ai dati per i ricercatori
- Scarsità di fondi per la ricerca

Lo studio mette in evidenza la centralità della ricerca relativa a immagini e video generati dall'IA, comprese campagne coordinate e disinformazione generata da ChatGPT o altri strumenti analoghi. Per sfruttare al meglio i dati limitati e i fondi disponibili, i ricercatori anziché essere ostacolati dalle attuali pratiche di ricerca (in cui diversi gruppi tendono a competere tra loro per produrre i modelli migliori e le pubblicazioni più citate) potrebbero iniziare a collaborare per perseguire l'obiettivo comune di conseguire progressi rapidi. Affinché ciò diventi possibile a beneficio della società, i finanziatori e i responsabili politici dovrebbero fornire un adeguato quadro di finanziamento e collaborazione che consenta collaborazioni transfrontaliere e inter-progettuali di lungo periodo. L'impatto sociale e geopolitico di un simile approccio congiunto e coordinato per contrastare l'influenza e la disinformazione online sarebbe molto significativo ed è altamente necessario, poiché la posta in gioco non è mai stata così alta in termini di manutenzione dell'integrità elettorale, sostegno della fiducia verso i media e la democrazia e impatto sulla salute dei cittadini, per citare solo alcuni esempi.

Per quanto riguarda i prossimi passi più concreti, gli autori del libro bianco chiedono in particolare all'UE:

- una migliore mediazione e un più ampio e migliore accesso ai dati allo scopo di addestrare nuovi modelli di rilevamento basati sull'IA, così come superare i limiti di approcci all'accesso ai dati proposti dalle piattaforme social espressamente allo scopo di consentire ai ricercatori di applicare nuovi modelli di IA per contrastare la disinformazione online.
- la fornitura di finanziamenti per la creazione di set di dati completi sulla formazione multilingue per i ricercatori di tutti i paesi europei. La creazione di nuovi modelli richiede dati etichettati come "esseri umani" per migliorare gli algoritmi di IA e valutare le loro prestazioni su diversi tipi di disinformazione, che abbracciano molti paesi e lingue europei. Tale iniziativa congiunta e ben finanziata per la creazione di dati consentirà quindi ai ricercatori di unire le forze nella creazione di questi set di dati estremamente necessari, ma costosi. Dati che le piattaforme già possiedono, poiché vengono creati (ma non condivisi!) come effetto

collaterale degli sforzi di moderazione dei contenuti.

LINK

https://magazine.fbk.eu/it/news/ia-generativa-e-disinformazione-progressi-sfide-e-opportunita/

TAG

- #ai4trust
- #aigenerativa
- #audio
- #chatgpt
- #disinformazione
- #IA
- #intelligenzaartificiale
- #media
- #social media
- #societàdigitale
- #testo
- #verifica dei fatti
- #video

AUTORI

- Giancarlo Sciascia