Il WEF propone nuovi metodi di censura basati sul potenziamento dell'IA

Il World Economic Forum (WEF) ha recentemente proposto nuovi metodi per monitorare e censurare dalle piattaforme social quelle che vengono definite «opinioni estreme», disinformazione e materiale pedopornografico. In un articolo intitolato «La soluzione agli abusi online? Intelligenza artificiale (IA) più intelligenza umana», si sostiene che i metodi di controllo tradizionali sui social network non siano più efficaci e occorra, dunque, potenziare l'IA tramite **nuovi set di apprendimento** che le consentano di raccogliere informazioni da «milioni di fonti», rendendola così in grado di "**decifrare" l'intelligenza umana** e di bloccare i contenuti "nocivi" prima ancora che arrivino sulle reti social. Sebbene il WEF tenga a precisare che l'articolo rappresenta l'opinione dell'autrice – Inbal Goldberger – esso, essendo ospitato sul suo sito ufficiale, non può che rappresentare almeno in parte anche i programmi della celebre organizzazione internazionale. Considerato peraltro che il suo fondatore – Klaus Schwab – è un fervente sostenitore delle tecnologie più avanzate e dell'IA, appartenendo alla schiera degli adepti della nuova religione tecno-scientista.

Nel mondo occidentale sedicente democratico non è la prima volta che si tenta di disciplinare ed eventualmente censurare i contenuti della rete dietro l'espediente subdolo della disinformazione e della sicurezza degli utenti, coniando nuove espressioni come "incitamento all'odio" e "opinioni estreme" assolutamente vaghe e sotto il cui ombrello potrebbe rientrare qualunque manifestazione legittima di dissenso verso l'ideologia dominante. Beninteso, che la rete e i social network vadano in qualche modo disciplinati è evidente: meno chiaro, invece, è il confine superato il quale – con il pretesto della disinformazione – si finisce per adottare una vera e propria **forma di censura mascherata**.

Già qualche mese fa, la Commissione europea aveva adottato il nuovo codice di condotta contro la disinformazione, basato soprattutto sulla cooperazione con i "moderatori" delle piattaforme online. A differenza di quest'ultima iniziativa europea, la proposta dell'autrice del WEF si affida incondizionatamente alle risorse delle **tecnologie avveniristiche**, limitando sempre di più il contributo umano in favore del **potenziamento dell'IA** e proiettandoci così in uno scenario dalle tinte distopiche. È evidente, infatti, come ci si muova nella direzione di un sempre maggiore controllo dei pensieri e delle opinioni delle persone da parte degli **algoritmi**, tanto che Yuval Noah Harari – anche lui membro del WEF – ha dichiarato che «entro 10, 20 o 30 anni tali algoritmi potrebbero anche dirti cosa studiare al college e dove lavorare, chi sposare e persino per chi votare».

Nello specifico, nell'articolo si sostiene che gli attuali metodi di controllo dei contenuti siano inefficaci per diverse ragioni: innanzitutto la rapidità con cui i responsabili degli abusi adottano tattiche sempre più sofisticate per eludere i rilevamenti e, in secondo luogo, i **limiti dell'IA** stessa. Quest'ultima, infatti, non è in grado di **distinguere i contesti** (ad esempio, non è in grado di capire se l'immagine di un nudo appartenga ad un contesto

pornografico piuttosto che a un'opera d'arte figurativa), né di rilevare minacce in lingue nelle quali non è stata addestrata. A differenza dell'IA, i moderatori umani possono capire più lingue e interpretare diverse culture: «questa precisione, tuttavia, è limitata dalla specifica area di competenza dell'analista», si legge nell'articolo. In generale, la tesi di fondo è che gli sforzi combinati di intelligenza umana e IA «non sono ancora sufficienti per **rilevare in modo proattivo i danni prima che raggiungano le piattaforme**», che sarebbe l'obiettivo ultimo per garantire un controllo veramente efficace dei contenuti della rete.

Il modo che l'autrice presenta per raggiungere questo obiettivo è «un **approccio basato sull'intelligenza**»: si tratta di introdurre negli insiemi di apprendimento dell'IA l'intelligenza umana, integrandola al suo interno, oltreché un sistema di acquisizione multilingue. In questo modo, «l'IA sarà in grado di rilevare nuovi abusi online su larga scala, prima che raggiungano le piattaforme tradizionali». Tutto ciò, si legge, «ci consentirà di creare un'IA con l'intelligenza umana integrata. Questa IA più intelligente diventa più sofisticata con ogni decisione di moderazione, consentendo infine un rilevamento quasi perfetto, su larga scala». Si tratta, dunque, di **raccogliere informazioni al di fuori dei canali social** da milioni di utenti, **monitorando costantemente le persone e le idee**, eliminando quindi quelle ritenute non in linea con gli standard delle piattaforme prima ancora che approdino su queste ultime.

È evidente che dietro questa frenetica corsa all'individuazione di sistemi sempre più sofisticati di monitoraggio e rimozione dei contenuti digitali vi sia non tanto la volontà di proteggere gli utenti, quanto l'incapacità da parte delle istituzioni di arginare un **dissenso sempre più debordante**, inasprito dagli ultimi avvenimenti politici e sociali che hanno alimentato il malcontento generale e la sfiducia negli organi rappresentativi istituzionali. E poiché – come sostiene anche l'eminente studioso Noam Chomsky – la società "liberal-democratica" si fonda sul **consenso**, nella fattispecie un consenso artificiale costruito attraverso tecniche **ingegneria sociale**, ciò non può essere tollerato. Di qui, il rischio che si vada nella direzione di un **controllo tecnologico sulla mente**, sui comportamenti e sulle opinioni sempre più avanzato e opprimente, ma allo stesso tempo anche abilmente dissimulato.

[di Giorgia Audiello]