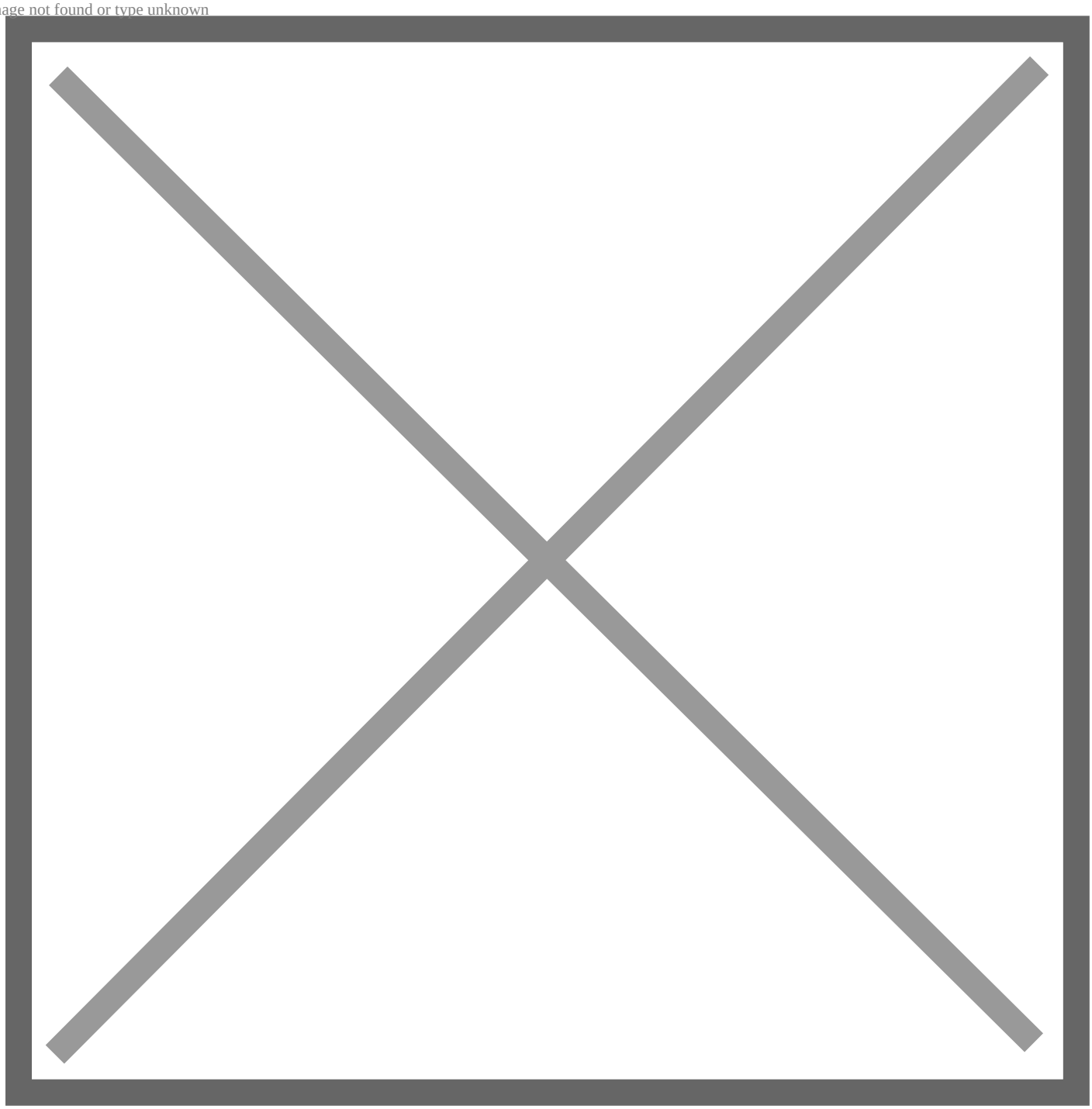
Image not found or type unknown



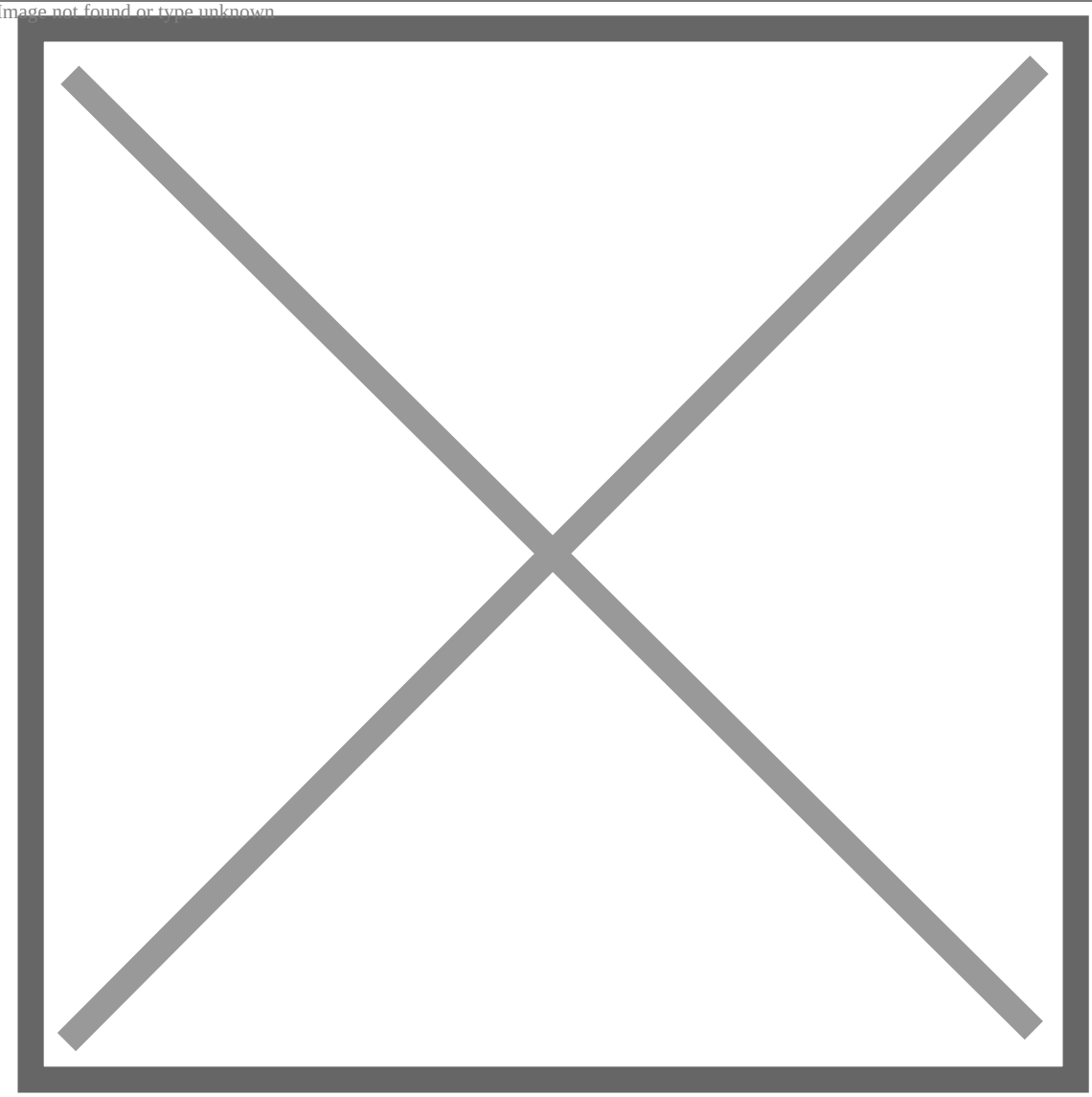
ERRORE DI ZUCKERBERG

Meta punta alla libertà di espressione, ma libera la violenza in rete

ECONOMIA 12_03_2025

Image not found or type unknown

Daniele Ciacci



Negli ultimi giorni Meta è stata sul filo del rasoio. Solo il 27 febbraio infatti ha risolto un errore tecnico che ha fatto apparire contenuti violenti nei feed personali dei reel di Instagram di numerosi utenti in tutto il mondo. L'azienda si è poi scusata per l'incidente, avvenuto dopo una serie di segnalazioni sui social media riguardanti contenuti inappropriati visualizzati nonostante alcuni utenti avessero attivato l'impostazione "Controllo Contenuti Sensibili" per filtrare il suddetto materiale.

In un comunicato stampa l'azienda di Zuckerberg ha dichiarato infatti di essere riuscita a limitare il dilagare di questi contenuti violenti, scusandosi per il disagio arrecato da un simile problema. Tuttavia, il portavoce di Meta non ha voluto dichiarare l'informazione più importante: quale è stata la causa effettiva del disguido. L'azienda è rimasta silente in merito.

Secondo le politiche di Meta, i contenuti particolarmente violenti dovrebbero essere

rimossi a monte, inclusi video di smembramenti, organi interni visibili e corpi carbonizzati. Questo filtro dovrebbe essere garantito da un sistema di Intelligenza Artificiale in grado di limitarli, seppur con qualche errore. Infatti, alcuni contenuti sono permessi se utili a sensibilizzare l'utenza su temi come l'abuso dei diritti umani o sui conflitti armati: in questi casi, all'inizio del reel è presente un'etichetta di avvertimento che chiede conferma di voler prendere visione di un contenuto che potrebbe potenzialmente turbare
l'utente.

È però significativo che questo incidente coincida con l'annuncio di Meta di aggiornare le proprie politiche di moderazione per promuovere maggiormente la libertà di espressione. Dal 7 gennaio, l'azienda ha modificato i sistemi automatizzati concentrandosi su "violazioni illegali e di alta gravità" come terrorismo, sfruttamento sessuale infantile e frodi, piuttosto che su "tutte le violazioni delle politiche". Per violazioni meno gravi, Meta si affiderà maggiormente alle segnalazioni degli utenti.

La conseguenza di questo nuovo corso sulle politiche di moderazione di Meta è principalmente l'eliminazione del programma di fact-checking negli Stati Uniti su Facebook, Instagram e Threads, tre delle più grandi piattaforme social con oltre 3 miliardi di utenti globali. A seguito di questa decisione, Zuckerberg ha anche dichiarato di aver subito pressione censoria da parte dell'amministrazione Biden per eliminare contenuti che mostravano gli effetti collaterali dei vaccini contro il Covid19. Pur sembrando una svolta che ha l'obiettivo principale di ingraziarsi Donald Trump, Meta negli ultimi anni ha fatto sempre più affidamento sui suoi strumenti di moderazione automatizzata, appunto per abbandonare un fact-checking che spesso nascondeva l'inquisizione dei poteri forti.

Allora, forse, l'errore dei giorni scorsi altro non è che un tentativo mal riuscito di Meta di equilibrare efficacemente i contenuti consigliati e la sicurezza degli utenti, non avendo più come vincolo l'argine delle segnalazioni dei fact checker. Ecco, quindi, la comparsa dei primi errori di bilanciamento, come la diffusione su Instagram di contenuti sui disturbi alimentari verso gli adolescenti.

Va anche notato che tra il 2022 e il 2023 Meta ha ridotto la sua forza lavoro di circa 21mila dipendenti, tagliando significativamente i gruppi responsabili dell'integrità delle informazioni e della sicurezza. Questo incidente infelice potrebbe quindi essere il primo inciampo verso una strada che, però, dovrebbe liberare l'azienda tech dalle grinfie di una politica sempre più interessata a strumentalizzarla.