LA NUOVA BUSSOLA QUOTIDIANA
DOMENICA
Direttore Riccardo Cascioli

è il cyber califfato 2.0

Così il Jihad usa l'IA per arruolare e far propaganda

ATTUALITà 23_02_2026

Martina Margaglio

Il panorama del terrorismo globale sta attraversando una metamorfosi senza precedenti. Superata l'era dei video sgranati caricati su forum attivi nel dark web, siamo entrati in una nuova fase del Cyber-Califfato 2.0. In un’epoca in cui l'innovazione

tecnologica non è più prerogativa esclusiva degli Stati o dei colossi della Silicon Valley, i gruppi jihadisti hanno dimostrato una capacità allarmante non solo di adattarsi, ma di colonizzare attivamente le nuove frontiere digitali. L'Intelligenza Artificiale è stata integrata nelle loro strategie e trasformata in un moltiplicatore di forza, sfruttata in ogni sua potenzialità per automatizzare il reclutamento, iperpersonalizzare l'indottrinamento e rendere la propaganda una macchina bellica invisibile e onnipresente.

Come rilevato dai principali think tank internazionali, tra cui la *Observer Research Foundation* (ORF), i gruppi estremisti oggi imitano le strategie di marketing di aziende e multinazionali non agendo più su una massa indistinta, ma puntando al singolo individuo attraverso la microsegmentazione. Questa tecnica permette di frammentare il pubblico in segmenti piccolissimi basati su vulnerabilità psicologiche e interessi locali, colpendo il destinatario con messaggi su misura.

Generano migliaia di contenuti diversi, ognuno ottimizzato per un profilo specifico, arrivando a manipolare la percezione del destinatario con una precisione degna delle più sofisticate campagne elettorali sempre più fondate su strategie di microtargeting e analisi dei dati. Questo processo di indottrinamento non avviene in un unico spazio, ma segue una struttura a imbuto studiata per eludere la sorveglianza e massimizzare l'efficacia psicologica.

Tutto inizia sui social media più popolari, che fungono da esca. Qui le reti di propaganda jihadista individuano gli utenti vulnerabili attraverso contenuti “gateway”, spesso inseriti nei feed di tendenza tramite il dirottamento degli hashtag. Come già segnalato dall'Europol nel 2023, queste reti usano l'IA per inondare il web di falsi filmati e immagini tradotti in moltissime lingue. Spesso creano dal nulla scene strazianti, come immagini di bambini feriti in guerre che non esistono, per suscitare rabbia e indignazione e aggirare al contempo i sistemi di moderazione automatica dei social, incapaci di distinguere sempre tra un contenuto reale e uno generato artificialmente. Poiché l’IA produce varianti ogni volta uniche, essa rende inefficaci i classici filtri basati sull'impronta digitale dei file (hash-matching), garantendo alla propaganda una persistenza senza precedenti.

Una volta individuato un profilo sensibile, l'interazione si sposta rapidamente verso spazi più protetti. L'utente viene, dunque, invitato tramite messaggi privati o link temporanei a unirsi a comunità su app di messaggistica criptata tra cui Telegram o RocketChat. In questi ecosistemi chiusi l'IA entra a pieno regime tramite chatbot sofisticati, capaci di simulare empatia e instaurare dialoghi personalizzati che siadattano allo stato emotivo dell'interlocutore.

Questo meccanismo crea un pericoloso senso di fiducia in individui vulnerabili, spesso giovanissimi nativi digitali isolati socialmente, che si ritrovano intrappolati in un confronto costante con entità artificiali progettate per radicalizzarli. In questa fase la propaganda utilizza le reti generative avversarie (GAN) per manipolare la percezione della realtà attraverso l'uso di avatar altamente verosimili e video deepfake. Questi strumenti permettono addirittura di riportare in vita leader carismatici del passato che sembrano incitare alla violenza nel presente, conferendo una sorta di autorità eterna al messaggio radicale.

L'ultima fase del percorso conduce a spazi online segreti, nascosti nel dark web o su reti protette diffuse, dove sono ospitati i materiali più radicali e i manuali operativi. In questo modo, l'utente viene isolato in una realtà parallela, protetta da una crittografia che rende il legame con la rete di propaganda difficilmente rintracciabile per le autorità.

Sul fronte del contrasto, le analisi più recenti evidenziano come le tecniche tradizionali di monitoraggio basate su parole chiave non siano più sufficienti. L'IA permette infatti di creare varianti infinite dello stesso messaggio, riuscendo a scivolare tra le maglie dei filtri automatici. Per rispondere a questa sfida, l'*intelligence* deve oggi puntare sulla *Data Fusion*: un sistema che incrocia enormi volumi di dati per agire su tre fronti. Primo, la traduzione istantanea di decine di dialetti locali; secondo, il riconoscimento visivo per localizzare brevi clip video in stile TikTok; terzo, l'analisi predittiva per intercettare quei contenuti di propaganda che sopravvivono ai filtri delle piattaforme e che sono sufficienti a mantenere attive reti di migliaia di persone.

Tuttavia, questa rincorsa tecnologica solleva seri dilemmi etici. L'uso di algoritmi predittivi rischia infatti di generare discriminazioni basate su pregiudizi etnici o religiosi, colpendo individui estranei ai fatti e minando la fiducia nelle istituzioni. In questa "guerra cognitiva", l'obiettivo dei terroristi non è solo colpire fisicamente, ma hackerare la percezione pubblica per alimentare le divisioni sociali.

Ci troviamo di fronte a una minaccia fluida e tecnologicamente avanzata che impone alle strategie antiterrorismo una complicata capacità di anticipazione. La sfida

resta complessa proprio a causa del continuo mutamento dei modelli comunicativi, che trasforma ogni difesa in una rincorsa costante contro una minaccia che non smette mai di evolversi.