

NEUROCIÈNCIA

Els missatges d'odi a xarxes s'assemblen a trastorns de personalitat

Una investigació de la Universitat de Texas situa aquest tipus de missatge molt a prop dels vinculats a trastorns com el narcisisme, el límit de personalitat i l'antisocial

David Bueno

Entrar en segons quins fòrums d'internet o en xarxes socials com X pot ser, en moltes ocasions, com caminar per un terreny minat. Un comentari desafortunat o simplement que no agradi a algun dels participants i pot esclatar una tempesta d'insults i amenaces. L'odi digital és tan habitual que sovint el considerem part inevitable de les xarxes socials, i tendim a associar-lo a la mala educació d'alguns internautes. Però què passaria si el llenguatge de l'odi a internet no fos només un reflex de mala educació? Què passaria si amagués patrons que recorden altres formes de comunicació humana relacionades a determinades alteracions de la personalitat?

Un treball publicat recentment per dos investigadors de la Universitat de Texas, Andrew William Alexander i Hongbin Wang, a *Plos Digital Health*, mostra que els discursos d'odi que es troben a les xarxes socials comparteixen trets lingüístics característics amb els textos que fan persones amb trastorns de personalitat. Han cartografiat aquest fenomen amb tècniques matemàtiques i, quan l'han situat en un mapa conceptual del llenguatge, han vist que es troben molt a prop dels discursos típics vinculats a trastorns com el narcisisme, el límit de personalitat i l'antisocial.

De manera resumida i simplificada, el narcisisme es manifesta en una necessitat constant d'admiració i una empatia limitada. El trastorn límit de personalitat, al seu torn, s'associa a una muntanya russa emocional, amb relacions intenses i molta por de ser abandonat. I l'antisocial es caracteritza per la manca de respecte per les normes i pels drets dels altres, amb tendència a manipular les persones i les situacions i amb molt poc remordiment pels actes comesos. Això no vol dir, diuen explícitament els investigadors i és molt important deixar-ho ben clar, que les persones amb aquests diagnòstics psiquiàtrics siguin més agressives, sinó que el llenguatge de l'odi a les xarxes socials té una estructura que recorda la desregulació emocional pròpia d'aquestes condicions.

Amb ajuda de la IA

Per arribar a aquestes conclusions, els autors van comparar milers de missat-



L'estudi és molt curós i no parla de persones sinó d'un tipus de llenguatge. GETTY

ges de comunitats d'odi i de fòrums sobre salut mental. Van recollir-los de 54 comunitats de Reddit, que és una plataforma de fòrums en línia que funciona com una gran comunitat de comunitats. Hi havia grups d'odi, fòrums de desinformació, comunitats sobre trastorns psiquiàtrics i grups de control. Cada missatge es va convertir en un vector

matemàtic de 1.536 dimensions mitjançant intel·ligència artificial. Després, amb tècniques de topologia de dades, van construir un mapa que mostrava quines comunitats són lingüísticament més properes. La topologia de dades és una tècnica matemàtica que serveix per entendre l'estructura amagada de conjunts de dades molt grans i complexes.

En lloc de mirar només punts concrets, analitza la forma global com s'agrupen en funció de la seva semblança.

El resultat va ser contundent: els discursos d'odi es col·loquen just al costat de les comunitats de trastorns de personalitat, molt més a prop entre elles que la dels grups de control. Van detectar punts en comú entre els grups d'odi i les comunitats sobre trastorns psiquiàtrics, específicament un ús intens d'expressions emocionals, una tendència a percebre l'altre com una amenaça i una comunicació marcada pel conflicte. Curiosament, els fòrums de desinformació presentaven un patró diferent. El seu llenguatge s'assemblava més al dels grups de control, amb una lleugera connexió amb trastorns d'ansietat. En altres paraules, no és el mateix difondre notícies falses o *fake news* que odi.

Cal dir que això no implica que les persones que fan missatges d'odi hagin de tenir algun d'aquests trastorns, sinó que tenen estils comunicatius similars. En altres paraules, les expressions dels *haters*, com se'ls acostuma a anomenar, poden sonar com les d'algú que lluita amb una regulació emocional difícil. Aquest paral·lelisme planteja una idea molt in-

Segons l'estudi, els fòrums de notícies falses tenen un patró diferent dels que fan servir el llenguatge d'odi

teressant: si les teràpies destinades a millorar l'empatia i la gestió emocional funcionen en pacients amb trastorns de personalitat, ¿podrien inspirar estratègies per reduir la toxicitat a internet?

El llenguatge de l'odi, diuen Alexander i Wang, no és només una qüestió d'ideologia; també és una manera de comunicar-se marcada per la desregulació emocional. Aquest fet permet entendre millor aquest fenomen i obrir noves vies d'acció. En proposen tres, que altrament són molt lògiques: incidir en educació emocional per poder gestionar millor les emocions davant la impulsivitat que comporta l'àmbit digital; afavorir una moderació més humana que en lloc de limitar-se a censurar explori estratègies que promoguin l'empatia i la reflexió, i desenvolupar eines de detecció per identificar l'odi abans que esclati.

Sempre cal tenir ben present, però, que no es pot vincular trastorns de personalitat amb odi, ja que significaria una estigmatització que podria alimentar prejudicis contra persones vulnerables; que aquests estudis analitzen textos, no persones, i que, per tant, de cap manera poden servir per fer diagnòstics, i que el desenvolupament d'algoritmes hauria de ser molt curós, per evitar una censura excessiva que confongués el llenguatge emocional lícit i legítim amb odi. La millor estratègia no és només esborrar missatges, sinó aprendre a parlar i escoltar d'una altra manera.

David Bueno és fundador de la Càtedra de Neuroeducació UB-EDUIST