

Más allá del cabello: raíces

Detección de discurso de odio en redes sociales

Beyond hair: Roots. Hate speech detection on social media

Por Verónica Neri Mendoza, Yulia Ledeneva y René Arnulfo García Hernández

Resumen: El discurso de odio en redes sociales representa una amenaza para quien lo sufre. Este documento explora cómo la frase “Pelo a la cintura, india segura” es discriminante hacia las mujeres indígenas en México. Por medio de un conjunto de datos y la implementación de algoritmos de inteligencia artificial (IA), este texto busca contribuir a visibilizar y sensibilizar sobre el uso estas expresiones.

Palabras clave: discurso de odio, indígenas, redes sociales, discriminación, IA.

Abstract: Hate speech on social media represents a threat to those who suffer from it. This document explores how the phrase “Long hair, ‘india’ for sure” has become a discriminating towards indigenous women in Mexico. Through a set of data and the implementation of artificial intelligence (IA) algorithms, we seek to contribute to the visibility and awareness of the use of these expressions.

Keywords: hate speech, indigenous people, social media, discrimination, IA.

El discurso de odio es una forma de comunicación que expresa desprecio hacia un individuo o grupo, basándose en su origen étnico, creencias o género (ONU, 2023), y se ha acrecentado con el uso de las redes sociales, fomentando un ambiente de hostilidad que puede llevar a actos violentos y normalización de la discriminación (Acevedo y López, 2023; Campos *et al.*, 2015). En ese sentido, es urgente desarrollar tecnologías capaces de reconocer los patrones lingüísticos asociados a este tipo de discurso para evitar su proliferación.

Actualmente, en el Centro Universitario UAEMéx Tlaxiaco, se está formando un conjunto de datos que

ayude a implementar algoritmos de inteligencia artificial (IA) para la detección de discurso de odio con enfoque indígena en México. Extrayendo comentarios públicos en la red social X, una de las frases más repetidas esta recopilación es “Pelo a la cintura, india segura” que ha sido convertida en un arma para ofender y denigrar a las mujeres.

Históricamente, el término *indio (a)* se ha utilizado de manera despectiva para referirse a los pueblos originarios (Campos *et al.*, 2015). El origen de la oración es difícil de rastrear: no existe registro de cuándo y dónde surgió. Probablemente, se creó en la cultura popular, transmi-



Tecnología

Imagen: freepik.es



tiéndose de generación en generación a través de chistes, dichos y conversaciones cotidianas, pero ¿qué tiene de malo el cabello largo?

Según Fray Bernardino de Sahagún (1499-1590), una característica distintiva de la vestimenta indígena era el cabello largo, dividido en trenzas; lo describe como un elemento de identidad cultural.

Figura 1. Comentarios con discurso de odio



Fuente: Elaboración propia

En estos ejemplos (Figuras 1 y 2) se observa cómo se asocia el largo del cabello con una imagen preconcebida de las mujeres indígenas. Su

representación estereotipada en personajes como la India María ha normalizado la burla y la discriminación hacia este grupo.

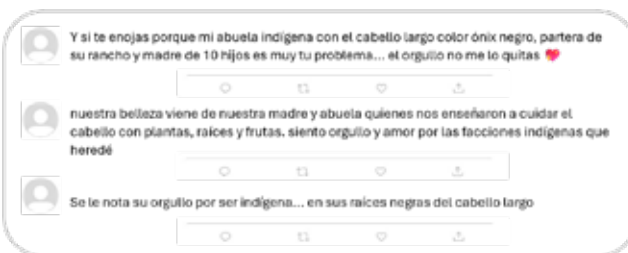
Figura 2. Chistes ofensivos



Fuente: Elaboración propia

Además de las ofensas explícitas, es importante mencionar que parte de la normalización de estos discursos tiene su origen en el humor, el cual disfraza las ofensas con bromas, no obstante, es evidente el menosprecio. Asimismo, hay textos que no utilizan palabras malsonantes, pero su sentido sí es denigrante (Figura 3).

Figura (3). Odio disfrazado de ironía



Fuente: Elaboración propia

¿CÓMO FUNCIONA LA IA?

Las personas podemos reconocer determinados mensajes verbales por las experiencias de vida que han moldeado nuestra percepción; sin embargo, para que la IA identifique discriminación en los textos es necesario implementar métodos de aprendizaje automático, esto consiste en enseñar a una computadora usar datos como entrenamiento, como redes neuronales que emulan al cerebro humano para tomar decisiones (Gómez-Adorno et al., 2024).

El conjunto de datos que se está desarrollando está a cargo de un equipo interdisciplinario para realizar un eti-

quetado manual de los textos (con o sin odio). Luego, el conjunto de datos se divide en dos partes:

Conjunto de entrenamiento: Alimenta al algoritmo y le enseña a reconocer patrones lingüísticos.

Conjunto de prueba: Verifica si la máquina aprendió y puede clasificar nuevos textos de forma precisa.

A continuación (Tabla 1), se presenta un ejemplo de la clasificación realizada por el algoritmo en relación con los comentarios mostrados en las figuras 1, 2 y 3.

Tabla 1. Análisis del texto a través de la máquina para identificar discurso de odio

Tarea	Discurso de Odio	Sin discurso de Odio
Identificar palabras o expresiones a través de frecuencias para identificar patrones.	* Palabras más frecuentes, el número de repeticiones se encuentra entre paréntesis. Pelo (4) India (6) María (3) Trenza/trenzuda (3)	Indígena/indígenas (3) Orgullo (3) Cabello (3) Madre (2)
Analizar las construcciones gramaticales.	* Hace uso de adjetivos como: India pl#%& gata pl#%& india	Indígenas ónix negro belleza
Analizar emociones del texto.	* En los textos de ejemplo se identifican expresiones de: odio, ira o desprecio. pl#%& gata, pl#%&, chl#%&! * El uso de emojis ayuda a contextualizar emociones 😏 (burla, broma)	respeto y aprecio. orgullo, belleza, amor y raíces ❤️ (amor, aprecio)

Con base en la frecuencia de palabras obtenidas, la máquina considera discurso de odio aquellos textos que contienen las palabras *indias*, *pelo*, *María* y *trenza/trenzuda*. Además de los adjetivos *india*, *pl#%&* y *gata*, así como las expresiones *pl#%& gata*, *pl#%&*, *chl#%&!* y los emojis 😏. Mientras que los textos sin discurso de odio estarían caracterizados por términos como *indígena/indígenas*, *orgullo*, *cabello* y *madre*. Por otro lado, adjetivos indígenas, *ónix negro* y *belleza* ayudarían a reforzar esta clasificación. De igual forma, el uso de emociones de aprecio y el emoji ❤️.

En conclusión, para que una máquina pueda identificar discurso de odio es necesario realizar diferentes análisis que permitan



Ilustración: Luis Ángel Velázquez



Tecnología

ADEMÁS DE LAS OFENSAS EXPLÍCITAS, ES IMPORTANTE MENCIONAR QUE PARTE DE LA NORMALIZACIÓN DE ESTOS DISCURSOS TIENE SU ORIGEN EN EL HUMOR, EL CUAL DISFRAZA LAS OFENSAS CON BROMAS, NO OBSTANTE, ES EVIDENTE EL MENOSPRECIO

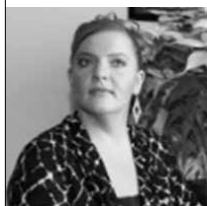
reconocer patrones, conceptos y relaciones semánticas que en conjunto tienen como finalidad ayudar a la IA a comprender el lenguaje y ofrecer decisiones cercanas a las humanas. Cabe mencionar que es un proceso complejo y no existe un método único para detectarlo. La elección del estilo de cabello, además de ser un aspecto cultural, es una decisión personal y no es motivo de discriminación. 🗨️

Referencias

- Acevedo, Eugenia Iturrana y Jaime López Reyes (coords. y eds.) (2023). *Pueblos indígenas frente al racismo mexicano. Caja de herramientas para identificar el racismo en México II*. UNAM.
- Campos, Mario et al. (2015). *Mensajes de odio y discriminación en las redes sociales*. Secretaría de Gobernación, Consejo Nacional para Prevenir la Discriminación.
- Gómez-Adorno, Helena et al. (2024). "Overview of HOMO-MEX at IberLEF 2024: Hate Speech Detection Towards the Mexican Spanish speaking LGBT+ Population", en *Procesamiento del Lenguaje Natural*, núm 73, 393–405.
- ONU (2023). *Objetivos de Desarrollo Sostenible - Programa de las Naciones Unidas para el Desarrollo*. <<https://www.undp.org/es/sustainable-development-goals>>.
- Sahagún, Fray Bernardino Ribeira de (1499-1590). *Historia General de las Cosas de la Nueva España*, tomo II. <<https://www.biblioteca-antologica.org/es/wp-content/uploads/2018/05/SAHAGÚN-Historia-General-de-la-Nueva-España-II.pdf>>.



Verónica Neri Mendoza es egresada del Doctorado en Ciencias de la Computación de la UAEMéx. Sus áreas de investigación son inteligencia artificial, procesamiento del lenguaje natural, minería de textos y reconocimiento de patrones.



Yulia Ledeneva es Doctora en Ciencias de la Computación por el CIC-IPN. Es profesora-investigadora de la UAEMéx, miembro del SNI, Nivel II, y autora de más de 85 publicaciones científicas en revistas de alto impacto y congresos internacionales. Sus áreas de investigación son inteligencia artificial, procesamiento del lenguaje natural, algoritmos genéticos y métodos basados en grafos.



René Arnulfo García-Hernández es Doctor en Ciencias de la Computación por el INAOE. Es profesor-investigador de la UAEMéx, miembro del SNI, Nivel II y autor de más de 85 publicaciones científicas en revistas de alto impacto y congresos internacionales. Sus áreas de investigación son reconocimiento de patrones, computación evolutiva, minería de textos y procesamiento del lenguaje natural.