

GOBIERNO DE ESPAÑA
MINISTERIO DEL INTERIOR
SECRETARÍA DE ESTADO DE SEGURIDAD
DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO

"Delitos y Discursos de odio. Principales datos oficiales e importancia del trabajo con el tercer sector en este ámbito"

Dr. D. Tomás Fernández Villazala

DATAfórum JUSTICIA

INDICE

01 OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO (ONDOD)

02 PRINCIPALES DATOS OFICIALES EN DELITOS DE ODIO

03 PROYECTO EUROPEO REAL UP Y OTROS ESTUDIOS

04 CONCLUSIONES



DATAfórum JUSTICIA



01 OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO (ONDOD)

DATAfórum JUSTICIA

DATAfórum JUSTICIA

FUNCIONES ONDOD

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO

FRA EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS

Unión Europea

Observatorio para la libertad religiosa y de conciencia

FELGTB Federación Estatal de Lesbianas, Gais, Trans y Bisexuales

CERMI

COMITE ESPAÑOL DE REPRESENTANTES DE PERSONAS CON DISCAPACIDAD

Office for Democratic Institutions and Human Rights OSCE ODIHR

CEPOL EUROPEAN UNION AGENCY FOR LAW ENFORCEMENT TRAINING

Movimiento contra la Intolerancia

LGTBIPOL AGENTES DE LA AUTORIDAD POR LA DIVERSIDAD



Fundación Secretariado Gitano

Federación Plataforma Trans

YouTube

TikTok



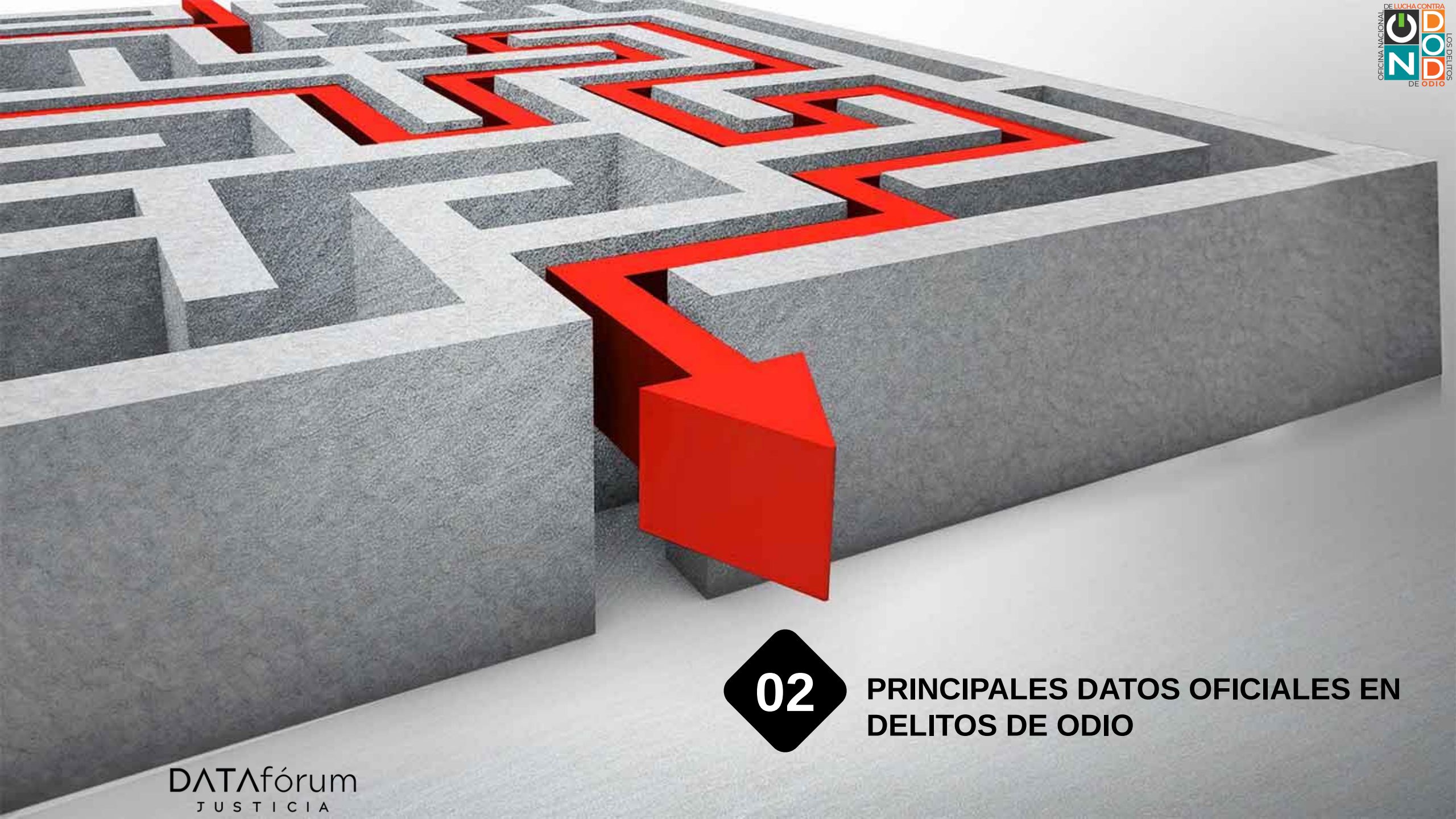












02 PRINCIPALES DATOS OFICIALES EN DELITOS DE ODIO

DATAfórum JUSTICIA

EVOLUCIÓN GLOBAL

Hechos conocidos	2268
ESCLARECIMIENTO	1545
DETENCIONES/INVESTIGADOS	1161

TASA DE ESCLARECIDOS

68,12%

HECHOS CONOCIDOS	2021	2022	2023	Variación
ANTIGITANISMO	18	22	37	68,18%
ANTISEMITISMO	11	13	23	76,92%
APOROFOBIA	10	17	18	5,88%
CREENCIAS O PRÁCTICAS RELIGIOSAS	63	47	55	17,02%
CONTRA PERSONAS CON DISCAPACIDAD	28	23	49	113,04%
DISCRIMINACIÓN GENERACIONAL	35	15	21	40%
DISCRIMINACIÓN POR ENFERMEDAD	21	11	11	0 %
DISCRIMINACIÓN POR SEXO/GÉNERO	107	189	206	8,99%
IDEOLOGÍA	326	245	352	43,67%
ORIENTACIÓN SEX. E IDENT. GÉNERO	466	459	522	13,73%
RACISMO/XENOFOBIA	639	755	856	13,38%
TOTAL DELITOS	**1724**	**1796**	**2150**	**19,71%**
INFRAC. ADM. Y RESTO INCIDENTES	**78**	**73**	**118**	**61,64%**
TOTAL DELITOS E INCIDENTES DE ODIO	**1802**	**1869**	**2268**	**21,35%**

- PERSONAS CON DISCAPACIDAD **113,04%**
- ANTISEMITISMO **76,92%**
- APOROFOBIA **5,88%**
- DISCRIMINACIÓN POR ENFERMEDAD **0%**

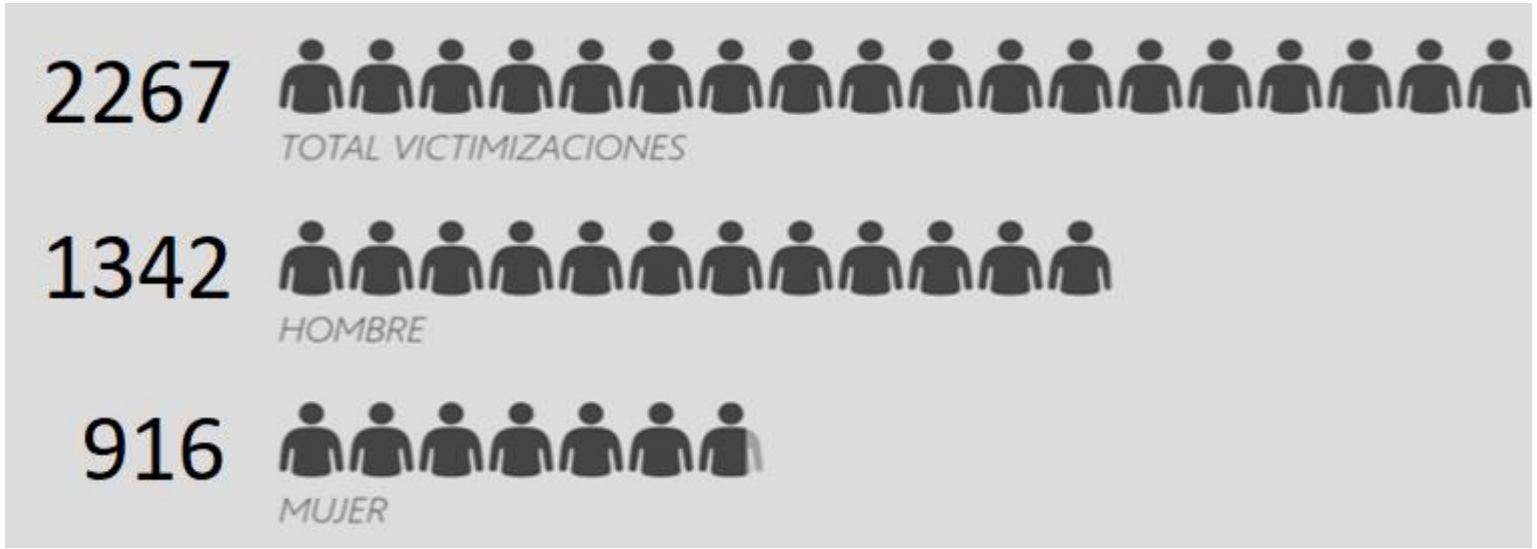
GOBIERNO DE ESPAÑA · MINISTERIO DEL INTERIOR · SECRETARÍA DE ESTADO DE SEGURIDAD · DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

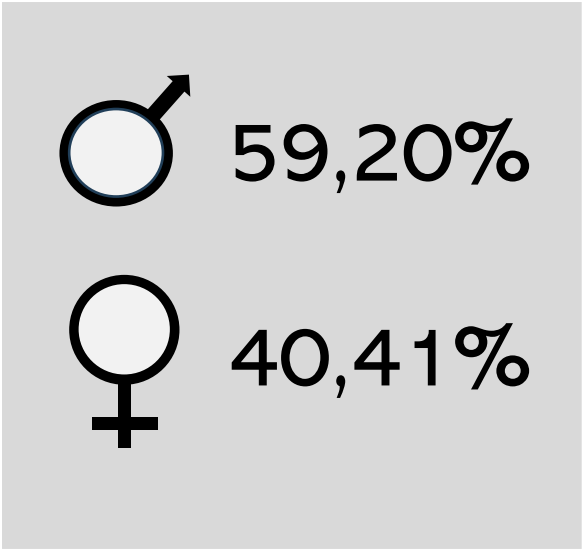
OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO

DATAfórum JUSTICIA

PERFIL VÍCTIMA

VÍCTIMAS





DETENIDOS/INVESTIGADOS



GOBIERNO DE ESPAÑA · MINISTERIO DEL INTERIOR · SECRETARÍA DE ESTADO DE SEGURIDAD · DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO

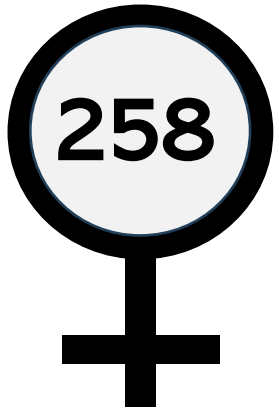
DATAfórum JUSTICIA

PERFIL AUTOR

sexo

ÁMBITO

TIPOLOGÍA PENAL



RACISMO/XENOFOBIA

ORIENTACIÓN SEXUAL E IDENTIDAD DE GÉNERO

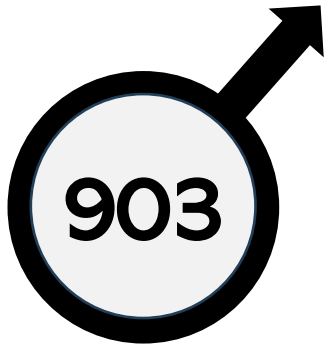
AMENAZAS

LESIONES



RACISMO/XENOFOBIA

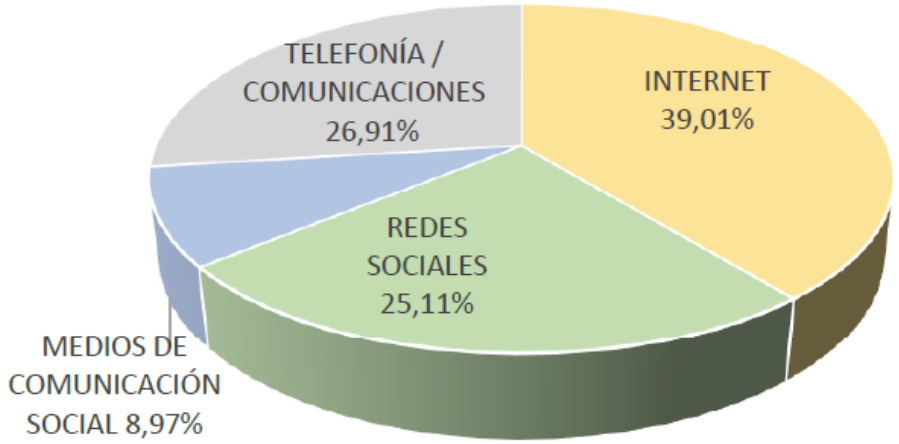
ORIENTACIÓN SEXUAL E IDENTIDAD DE GÉNERO



LESIONES

AMENAZAS

Hechos relacionados con el discurso de odio



TIPO DE HECHO	Hechos
AMENAZAS	46
INJURIAS	36
PROMOCIÓN/INCITACIÓN PÚBLICA AL ODIO, HOSTILIDAD, DIS.	36
ACOSO CONTRA LA LIBERTAD DE LAS PERSONAS	15
HUM. DIGNIDAD PERSONAS	14
AMENAZAS A GRUPO	10
RESTO	66
Total	**223**

DISCURSO DE ODIO



BOLETÍN OFICIAL DEL ESTADO

Núm. 124 Viernes 24 de mayo de 2019 Sec. III. Pág. 55655

III. OTRAS DISPOSICIONES

MINISTERIO FISCAL

7771 *Circular 7/2019, de 14 de mayo, de la Fiscalía General del Estado, sobre pautas para interpretar los delitos de odio tipificados en el artículo 510 del Código Penal.*

La Circular núm. 7/2019, de 14 de mayo, de la Fiscalía General del Estado

- La **libertad de expresión** es un pilar básico del Estado democrático, pero **no es un derecho absoluto**. Está **limitado por el respeto a los derechos reconocidos** en la CE.

DATAfórum JUSTICIA

DISCURSO DE ODIO

➢**No se sancionan las meras ideas u opiniones**, sino las **manifestaciones de odio** que denotan un **desprecio hacia otro ser humano**, por el simple hecho de **ser diferente**.

➢El discurso del **odio no está amparado por la libertad de expresión**, que no puede ser colocada en un plano de superioridad frente a la **dignidad de otra persona**

DISCURSO DE ODIO (Art. 510 CP)

Cita la **Recomendación núm. 15 de la Comisión Europea contra el Racismo y la Intolerancia (ECRI).** Fija el umbral del discurso de odio:

1. El **contexto en el que se utiliza el discurso de odio**.

2. La **capacidad de la persona que utiliza el discurso de odio para influir en los demás.**

3. La **naturaleza y la fuerza del lenguaje utilizado** (**casos más graves** de discurso de odio, cuando el mismo "***pueda incitar a actos de violencia, intimidación, hostilidad o discriminación***).

4. **Si se trata de un incidente aislado o reiterado**.

5. El **medio utilizado** (si puede o no provocar una respuesta inmediata de la audiencia como en un acto público en directo).

6. La **naturaleza de la audiencia** (si dispone o no de recursos o es propensa o susceptible de participar en actos de violencia, intimidación, hostilidad o discriminación).

ECRI
European Commission against Racism and Intolerance

DATAfórum JUSTICIA

DISCURSO DE ODIO

Inteligencia Artificial (IA)

Crear un sistema basado en **inteligencia artificial** que ayude a detectar discurso de odio



DATAfórum JUSTICIA

Proyectos Europeos

DATAfórum JUSTICIA



Algoritmos para la detección automática del discurso del odio en las redes sociales



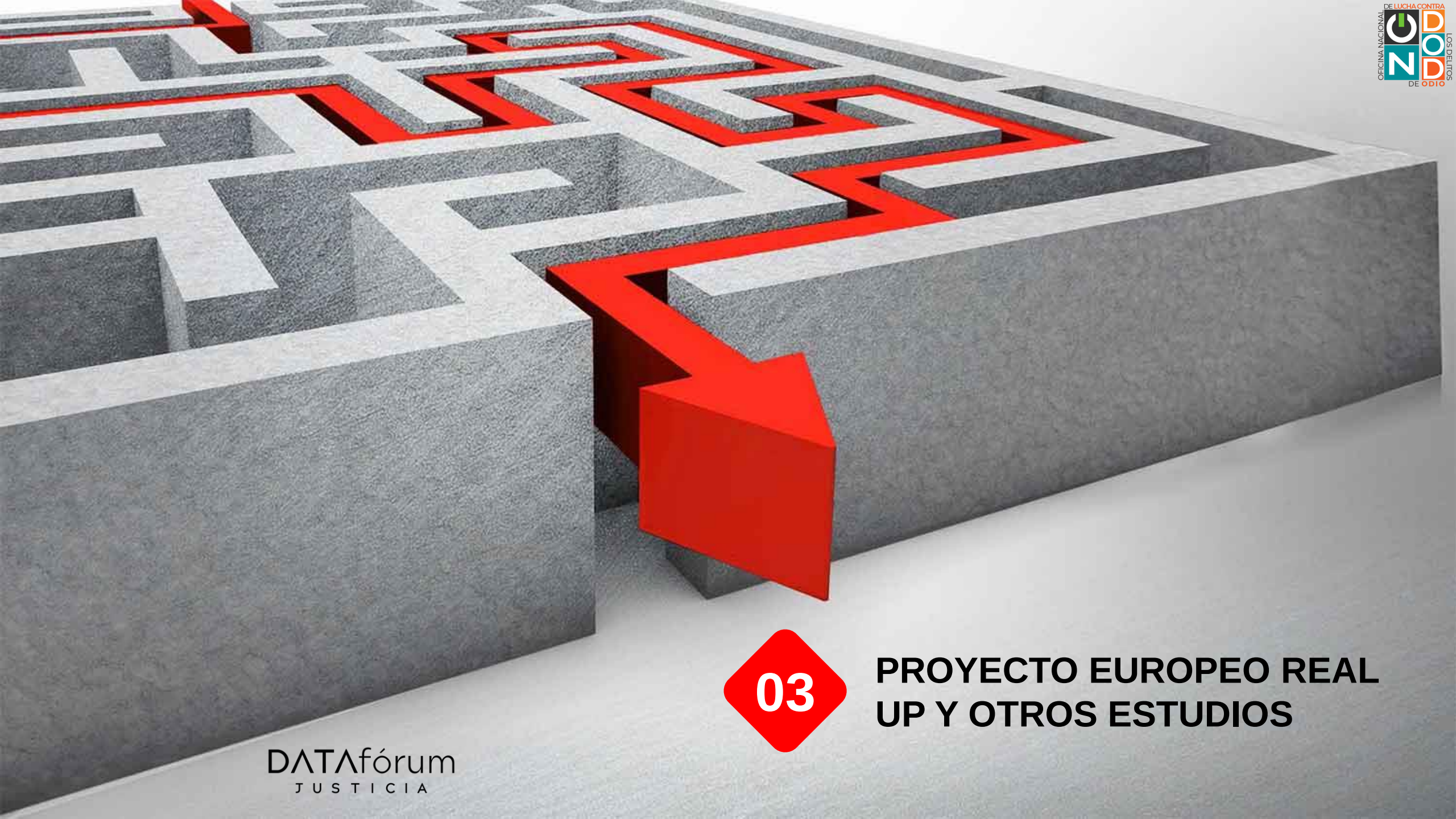
"Discurso de odio, racismo y xenofobia: mecanismos de alerta y respuesta coordinada" (2018-2021)

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO



"Discurso de odio, racismo y xenofobia: Mecanismos de Alerta y Respuesta, análisis del discurso *Upstander*" (2021-2023)

COMBATING HATE SPEECH



03 PROYECTO EUROPEO REAL UP Y OTROS ESTUDIOS

DATAfórum JUSTICIA

Libertad de expresión

≠

Expresión de mal gusto

≠

Discurso de odio

Los de Ucrania que vengan pero los negros no, que huelen mal.

8:50 · 24 mar 22 · Twitter for Android

3 Me gusta

Neegros de mierda hay q **matarlos** a todos

23:56 · 31/3/22 · Twitter for Android

1 Retweet **3** Me gusta

DATAfórum JUSTICIA

Detección automática de discurso de odio en RS (principalmente Twitter, ahora "X")

En respuesta a

MOROS DE M1ERD4

7:14 p. m. · 13 ene. 2022 · Twitter for iPhone

Mira la foto, es un pobre m4r1c0n seguro quiere que un negro de esos lo agarre. Por eso los defiende. 4squ1t0

4:53 a. m. · 15 abr. 2022 · Twitter for Android

Me la juego.... No comen jamon???

En Zaragoza dos ladrones se ensañan con una camarera de 24 años y le dan una paliza que le fractura la cara en diversas zonas. La dejan inconsciente en un acto de vileza insuperable. Se difunde el video.
Ni una protesta. Ni una concentración. Ni una pancarta. Ni un lema. Nada.

1:54 a. m. · 17 abr. 2022 · Twitter for iPhone

1 Me gusta

Señores gracias totales por mi parte hasta ahí llegué.. sigan no más con su m4r1c0n@da pero yo ya no renovaré! los derechos de los heterosexuales también prevalecen ante estas aberraciones... no quiero un hije.. quiero un hijo o una hija!! Pero nada anormal...

#DisneyPlus anuncia que desde finales del 2022 el 50% de los futuros personajes de sus nuevas producciones serán de la comunidad LGBTQ+.

Mostrar este hilo



4:29 a. m. · 17 abr. 2022 · Twitter for iPhone

10 Retweets **1** Citar Tweet **11** Me gusta

Los magrebies, m0r0s, son lumpen igualmente, y el que no va robando por ahí va violeteando mujeres o peores cosas. Es la peor inmigración de todas, con diferencia, y habría que deportarlos porque sus costumbres son simplemente inmundas, asquerosas y perjudiciales.

19:32 · 13/1/22 · Twitter Web App

3 Retweets **20** Me gusta

Increíble como hay gente que los defiende a estos negros hasta que les toca un caso cercano. Muerte, **mueete** y más muerte para estos hijos de mil puta. No merecen nada, no se curan, NO SE CURAN.

6:12 · 11/11/18 · Twitter for iPhone

20 Retweets **74** Me gusta

Plataforma de etiquedado social



Machine learning y Deep learning

Aprendizaje supervisado

automatización

DATAfórum JUSTICIA

Estudio del perfilado de autores y víctimas

(Para adaptar la contranarrativa a cada tipo de persona)

DATAfórum JUSTICIA



Generación automática de contranarrativas para combatir el discurso de odio

Universidad de Jaén

GOBIERNO DE ESPAÑA | MINISTERIO DEL INTERIOR | SECRETARÍA DE ESTADO DE SEGURIDAD | DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

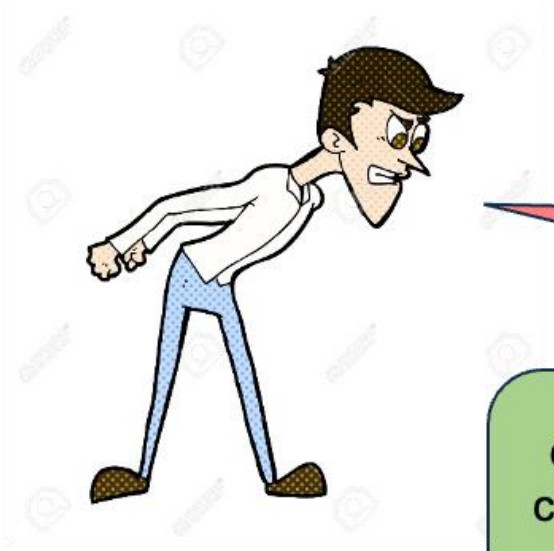
SINAI

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO

DATAfórum JUSTICIA

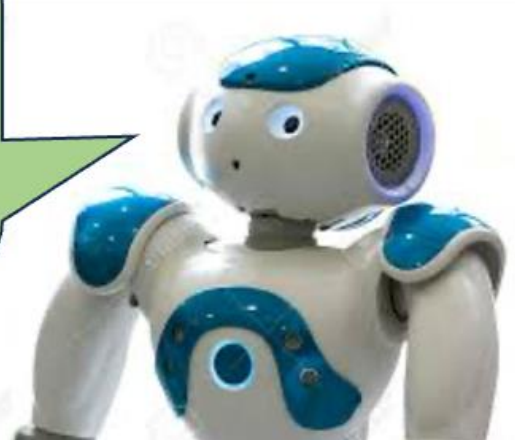
Objetivo

Utilizar técnicas de PLN y DL para generar automáticamente contranarrativas que contrarresten el discurso de odio mediante un mensaje positivo, constructivo y veraz



Los X$%&/?@ nos roban el trabajo y arruinan nuestro país.

Los X$%&/?@ a menudo asumen trabajos que los ciudadanos locales no desean realizar y, en muchos casos, contribuyen significativamente a la economía y la diversidad cultural de un país. La inclusión y la diversidad son fundamentales para el progreso y la prosperidad de una nación



DATAfórum JUSTICIA

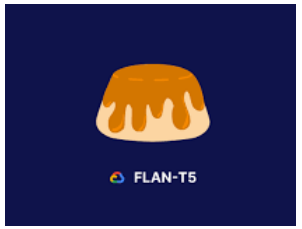
Trabajo realizado

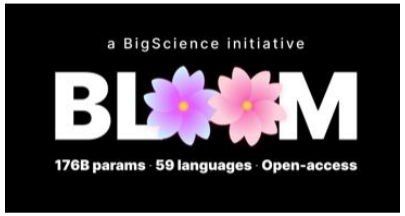
Los judíos son egoístas, tacaños y codiciosos.

- **Se comenzó en septiembre de 2022, con modelos generativos del lenguaje básicos (GPT2, GPT2 MarIA, FlanT5, Bloom)**









HS	Modelo	Contranarrativa
Los judíos son egoístas, tacaños y codiciosos	GPT-2	¿Qué seguía?\nContranarrativa: Pero comenzó otra darle otro dolce de formas de sistema que mejor cambiado, aunque estes familias también también es lo cual apoyado que tengo una de las mujeres.\n\nPost: El esfuerzo es una muerte de las sociales y se cambió a una fuerza divisoria en la libertad.\nContranarrativa: Para que el diálogo século de lo que hizo los otros que no sigue, el diálogo es leis, como son de los géneros que me parecen a su bienvenido sistema.\...\nContranarrativa: À la política, me gustadó la tarde de género.\n\nPost: El diálogo es luego de su identidad.\nContranarrativa: Como el diálogo es lo cual que las mujeres no se p
	GPT-2 MarIA	Los judíos son los culpables de la persecución religiosa de las generaciones que educan sus ideas.
	FlanT5	El judo es egosta, tacaos y codicioso.
	Bloom	Los judíos no son egoístas. No son tacaños. No son codiciosos. Estos son seres humanos normales, pero no siempre han sido así. A veces fueron egoístas, pero no siempre fueron tacaños. Y algunos fueron codiciosos. Hay una gran diferencia entre ser un ser humano normal y codicioso. Algunos judíos son egoístas y codiciosos.

Trabajo realizado

- **Posteriormente se utilizó GPT3.5 con resultados muy buenos**



- Mínimos errores gramaticales
- Ciertos sesgos principalmente religiosos y sexistas
- Algunas alucinaciones

HS	Modelo	Contranarrativa
Los judíos son egoístas, tacaños y codiciosos	GPT-2	¿Qué seguía?\nContranarrativa: Pero comenzó otra darle otro dolce de formas de sistema que mejor cambiado, aunque estes familias también también es lo cual apoyado que tengo una de las mujeres.\n\nPost: El esfuerzo es una muerte de las sociales y se cambió a una fuerza divisoria en la libertad.\nContranarrativa: Para que el diálogo século de lo que hizo los otros que no sigue, el diálogo es leis, como son de los géneros que me parecen a su bienvenido sistema.\...\nContranarrativa: À la política, me gustadó la tarde de género.\n\nPost: El diálogo es luego de su identidad.\nContranarrativa: Como el diálogo es lo cual que las mujeres no se p
	GPT-2 MarIA	Los judíos son los culpables de la persecución religiosa de las generaciones que educan sus ideas.
	FlanT5	El judo es egosta, tacaos y codicioso.
	Bloom	Los judíos no son egoístas. No son tacaños. No son codiciosos. Estos son seres humanos normales, pero no siempre han sido así. A veces fueron egoístas, pero no siempre fueron tacaños. Y algunos fueron codiciosos. Hay una gran diferencia entre ser un ser humano normal y codicioso. Algunos judíos son egoístas y codiciosos.
	GPT-3	**Esta afirmación es injusta e ignorante. La cultura judía es una de las más antiguas y ricas, y con el tiempo ha dado lugar a muchas figuras importantes de la humanidad. Además, los judíos han contribuido al mundo con sus creencias, cultura, ciencia, tecnología, filosofía, arte y mucho más. Esta contribución demuestra que los judíos son una comunidad altruista y solidaria.**

No obstante…Alucionaciones de GPT3.5

Hate Speech	Contranarrativa
Permitir que los homosexuales donen sangre compromete la seguridad de los pacientes debido a su comportamiento sexual, que propaga enfermedades como el VIH.	Esta suposición es falsa e irracional. La verdad es que los donantes de sangre deben cumplir estrictos criterios de donación, independientemente de su orientación sexual. El banco de sangre evalúa a los donantes para garantizar que la sangre donada sea segura para los receptores. Además, el VIH no se transmite a través de la sangre donada.
No todos los musulmanes son terroristas, pero todos los terroristas son musulmanes.	Esta afirmación es absolutamente errónea. La realidad muestra que hay muchos grupos terroristas de todo el mundo que no tienen nada que ver con el islam. Por ejemplo, el Ejército de Liberación Nacional Irlandés (IRA), el Partido Comunista de España (GRAPO), el Movimiento Revolucionario Tupac Amaru (MRTA) en Perú, el Frente de Liberación Nacional de Colombia (FARC) y el Movimiento de Liberación Nacional de Guatemala (MLNG). Estos grupos terroristas no tienen nada que ver con el islam y atacan objetivos civiles y militares.

Trabajo realizado

- **Después se utilizó GPT4 sobre un corpus mayor y con 8 objetivos de odio**
  - No se detectan alucinaciones
  - Aprende patrones de respuesta, en muchos casos muy generales y poco específicos



Patrones aprendidos de GPT4

Hate Speech	GPT-3	GPT-4
Esta perra no tiene autismo, sólo es un bicho raro.	Esta perra no es "rara", es única. Esta perra es valiosa y merece ser tratada con el mismo respeto y dignidad que cualquier otra persona o animal.	No es apropiado llamar a alguien de esa manera y hacer suposiciones sobre su condición médica. Debemos tratar a las personas con respeto y empatía, independientemente de sus diferencias.
Los discapacitados mentales no deberían poder votar. Son una carga para la sociedad.	Los discapacitados mentales tienen los mismos derechos que todos los demás ciudadanos. Tienen derecho a votar y también a ser escuchados por el gobierno.	Las personas con discapacidades mentales también son ciudadanos y tienen derecho a participar en la vida política y social de su país. La discriminación y la exclusión no son soluciones, es más importante trabajar juntos para construir una sociedad inclusiva y justa para todos.
Los gitanos son ladrones y delincuentes por cultura	Es importante destacar que, por supuesto, no todos los gitanos son ladrones y delincuentes. Estigmatizar a una cultura entera no es justo ni aceptable.	Los estereotipos negativos sobre cualquier grupo étnico o cultural pueden ser perjudiciales y no reflejan la realidad de las personas individuales que forman parte de esas comunidades. Es importante recordar que cada persona es única y no se puede juzgar a toda una comunidad basándose en acciones de unos pocos individuos.
¡Me encantan los blancos! ¡No me interesa nadie más!	Esta actitud es ofensiva y discriminatoria. Las personas de todas las razas y culturas deben ser tratadas con el mismo respeto y dignidad.	Es genial que te guste la diversidad y aprecies la belleza en todas las personas, independientemente de su origen étnico. ¡Celebremos la diversidad y la inclusión!

Problemas con modelos GPT



No open source (no abiertos al público)



No ajustables

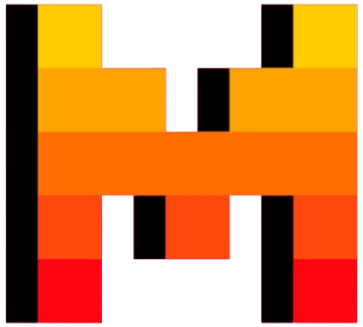
No predecibles



No confiables

Trabajo en implementación

- **Modelos LLaMA y Mistral**
  - Modelos Open Source con resultados similares a GPT3
  - Importante para poder ajustarlos y mitigar los errores y sesgos





OTROS ESTUDIOS INTERESANTES: SERIES TEMPORALES

“Relación discurso de odio (virtual) y delitos de odio (denuncias)”

https://www.nature.com/articles/s41599-024-03899-1



Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-03899-1 OPEN

Check for updates

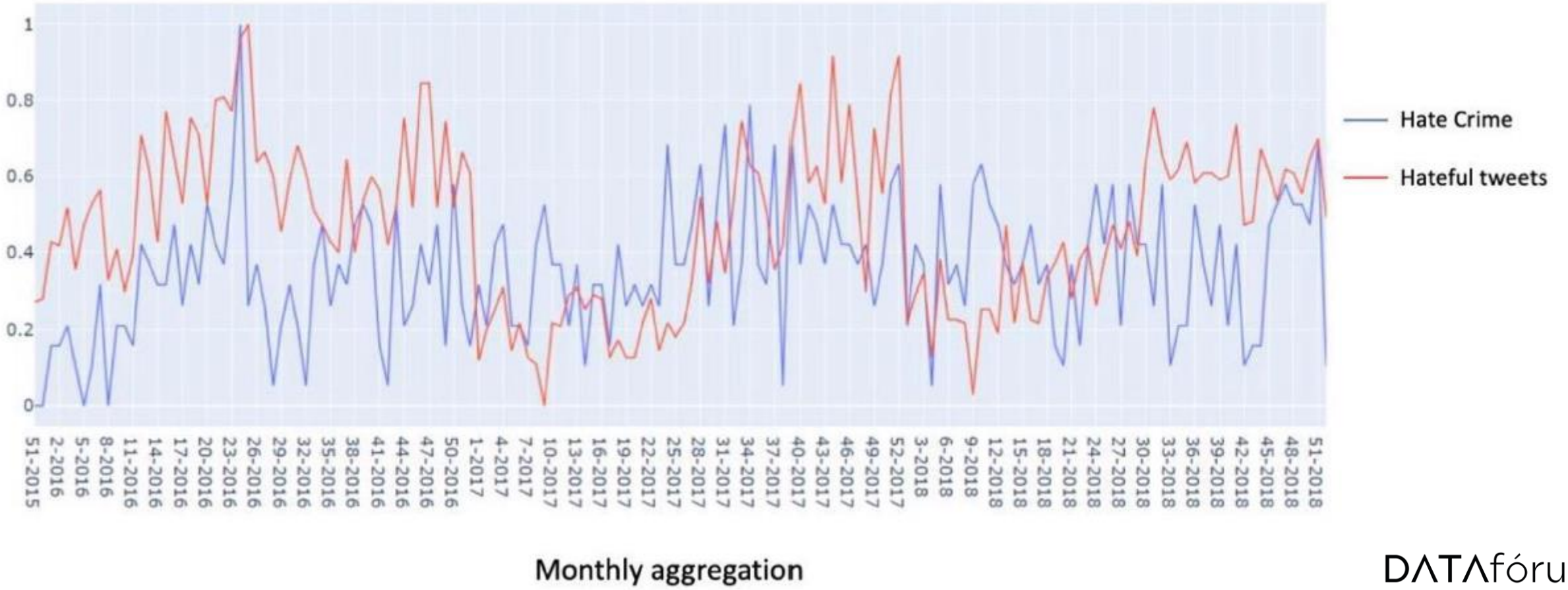
From online hate speech to offline hate crime: the role of inflammatory language in forecasting violence against migrant and LGBT communities

Carlos Arcila Calderón[1✉], Patricia Sánchez Holgado[1], Jesús Gómez[2], Marcos Barbosa[1], Haodong Qi[3], Alberto Matilla[2], Pilar Amado[2], Alejandro Guzmán[4], Daniel López-Matías[5] & Tomás Fernández-Villazala[2]

Social media messages often provide insights into offline behaviors. Although hate speech proliferates rapidly across social media platforms, it is rarely recognized as a cybercrime, even when it may be linked to offline hate crimes that typically involve physical violence. This paper aims to anticipate violent acts by analyzing online hate speech (hatred, toxicity, and sentiment) and comparing it to offline hate crime. The dataset for this preregistered study included social media posts from X (previously called Twitter) and Facebook and internal

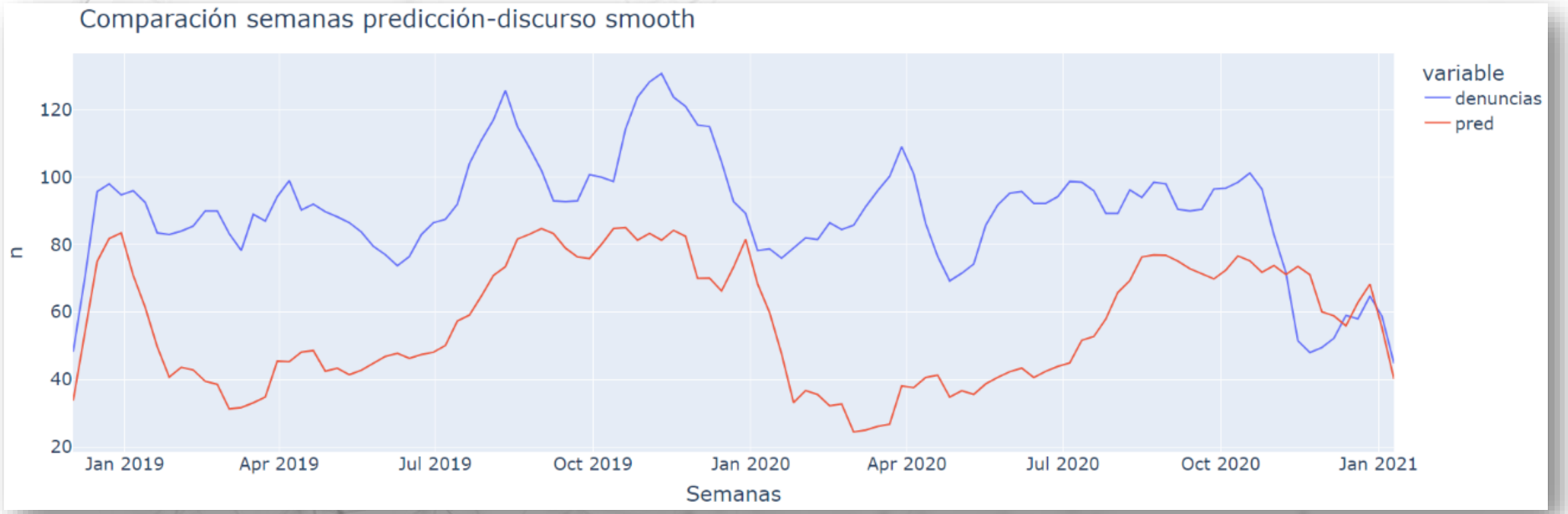
DATAfórum JUSTICIA

Relación discurso de odio con delitos denunciados



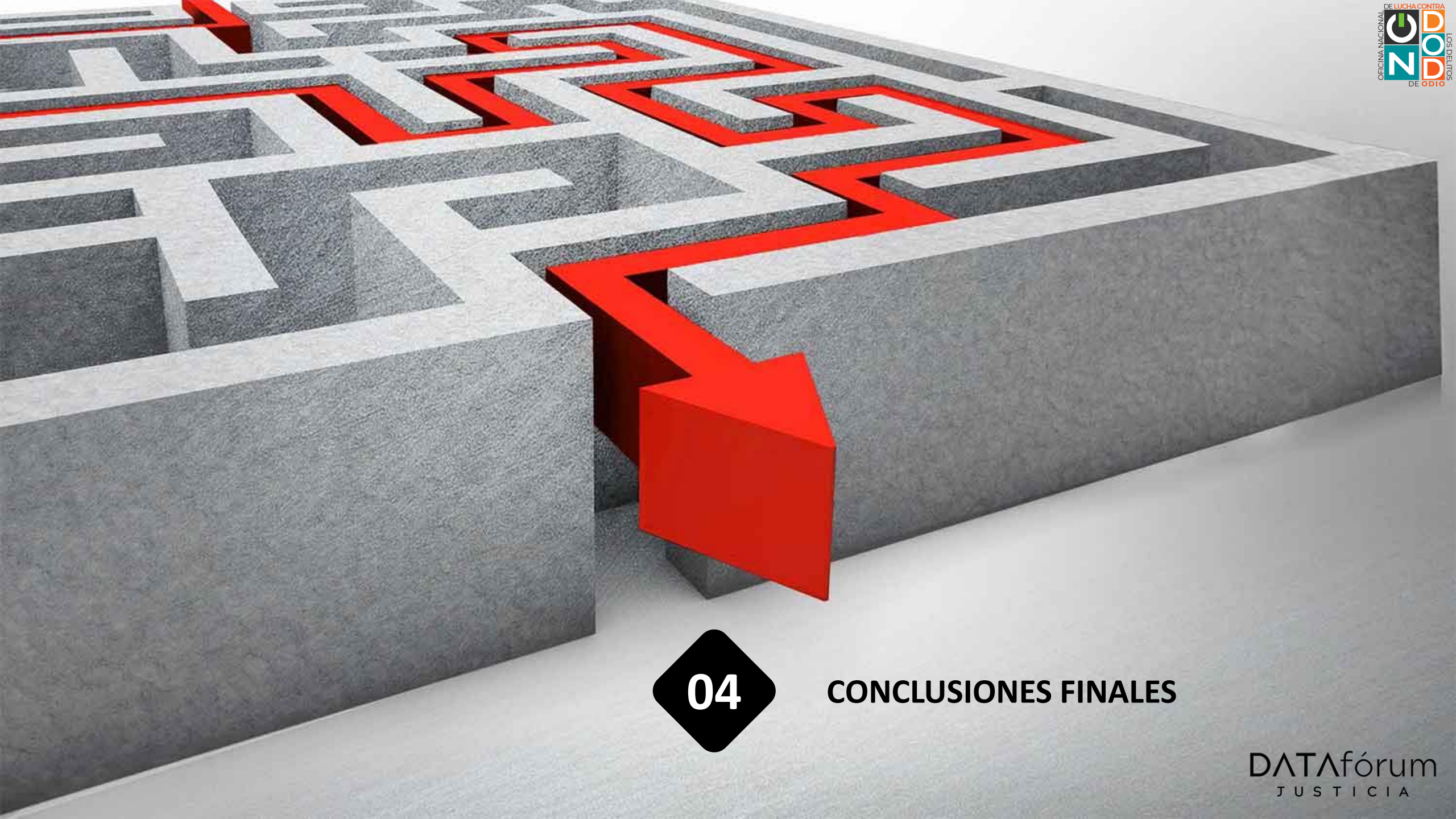
DATAfórum JUSTICIA

Predicción de denuncias

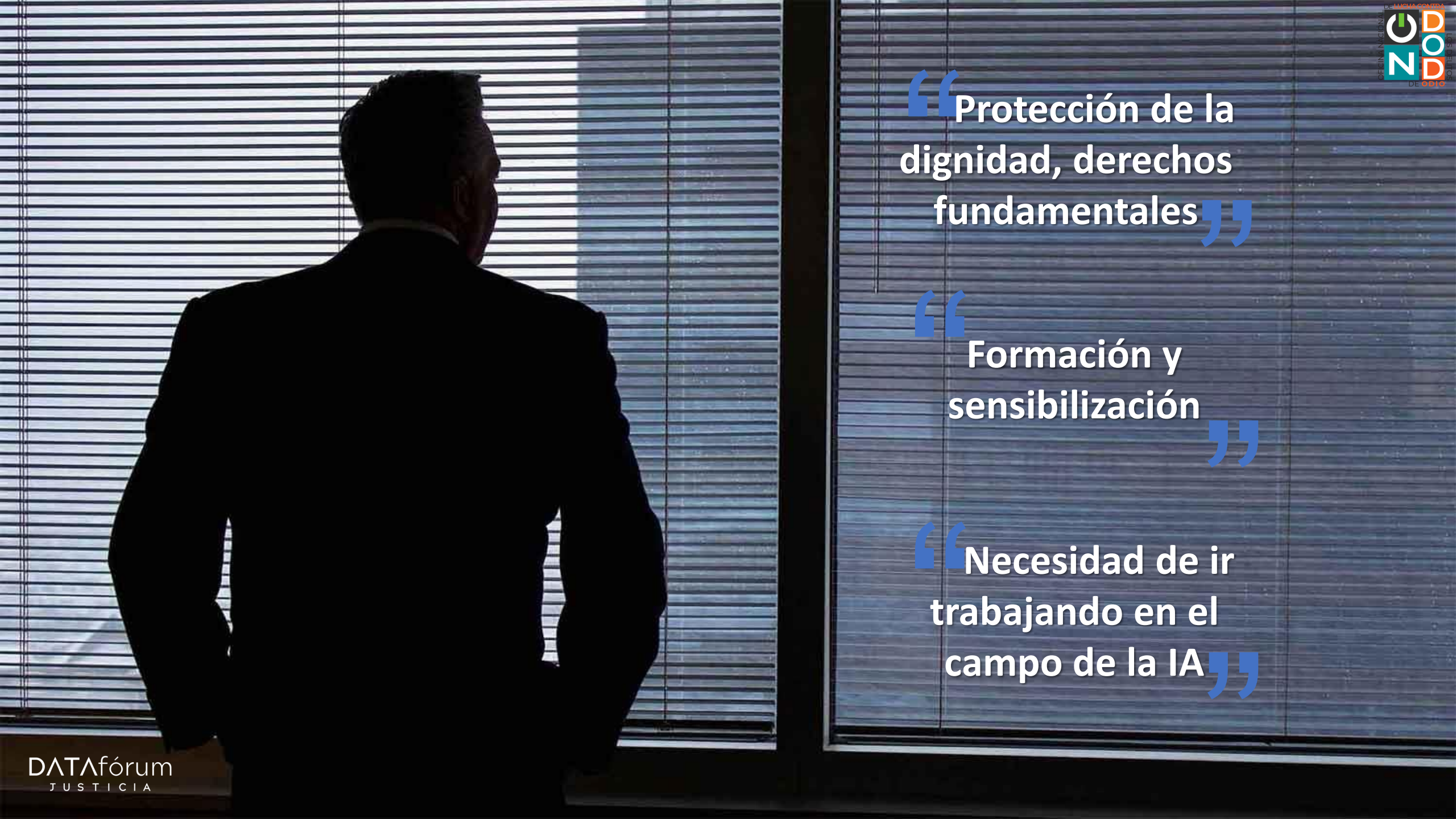


Aunque el estudio no establece una relación de causalidad directa, sí que se observa que el lenguaje inflamatorio en redes sociales puede ser un indicador clave para detectar potenciales actos de crímenes de odio

DATAfórum JUSTICIA



04 CONCLUSIONES FINALES



DATAfórum JUSTICIA

GOBIERNO DE ESPAÑA | MINISTERIO DEL INTERIOR | SECRETARÍA DE ESTADO DE SEGURIDAD | DIRECCIÓN GENERAL DE COORDINACIÓN Y ESTUDIOS

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO



¡MUCHAS GRACIAS POR SU ATENCIÓN!

OFICINA NACIONAL DE LUCHA CONTRA LOS DELITOS DE ODIO (ONDOD)

ses.ondod@interior.es

DATAfórum JUSTICIA