

Proyecto PID2021-123983OB-I00 financiado por:



# COVID-19 Y SINOFOBIA EN TWITTER: UN ANÁLISIS DE LOS DISCURSOS ANTI-CHINA DURANTE LA PANDEMIA

Carolina Rebollo, [carolina.rebollo@dstso.uhu.es](mailto:carolina.rebollo@dstso.uhu.es)  
Alba Taboada Villamarín, [alba.taboada@uam.es](mailto:alba.taboada@uam.es)  
Universidad de Huelva. Grupo ESEIS / COIDESO  
Universidad Autónoma de Madrid

**CONSPIRACY THEORIES AND HATE SPEECH ONLINE:  
COMPARISON OF PATTERNS IN NARRATIVES AND SOCIAL MEDIA ABOUT  
COVID-19, IMMIGRANTS, REFUGEES AND LGBTIQ+ PEOPLE**

*II International Workshop on Methodological Advances and Applications in Science*

12-14 JULY, HUELVA

Proyecto PID2020-115095RB-I00 financiado por:



Ayuda Margarita Salas para la formación de jóvenes doctores financiada por:



Funded by the  
European Union  
NextGenerationEU



Proyecto PID2021-123983OB-I00 financiado por:



# ACKNOWLEDGEMENT

This paper is part of the I+D+i Project titled “**Conspiracy Theories and hate speech online: Comparison of patterns in narratives and social networks about COVID-19, immigrants, refugees and LGBTI people [NON-CONSPIRA-HATE!]**”, PID2021-123983OB-I00, funded by MCIN/AEI/10.13039/501100011033/ and by “ERDF A way of making Europe”.

It also is part of the I+D+i Project 'CONCERN' “**Confianza, sistemas científicos y negacionismo. Factores sociales de la vacunación en contextos epidémicos**”, PID2020-115095RB-I00, funded by MCIN/AEI/10.13039/501100011033.

It also benefited from the European Union, nextGeneration-UE and the Ministerio de Universidades, and the support of ESEIS (Social Studies and Social Intervention Centre) and COIDESO (Centro de Investigación en Pensamiento Contemporáneo e Innovación para el Desarrollo Social) at the Universidad de Huelva, Spain.

Proyecto PID2020-115095RB-I00 financiado por:



Ayuda Margarita Salas para la formación de jóvenes doctores financiada por:



Funded by the European Union  
NextGenerationEU



# INTRODUCTION

During the COVID-19 pandemic, there has been an increase in **cyber-racism and hate speech towards the Asian community**, particularly the Chinese as well as **conspiracy theories** blaming China for the **creation and spread of the virus**.

**Sinophobia** → fear or dislike of China, Chinese people, Chinese culture or Chinese diaspora abroad.

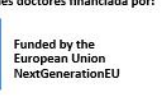


La Vanguardia (2020). <https://www.lavanguardia.com/cribeo/fast-news/20200203/473280657621/comunidad-china-critica-crecientes-actitudes-racistas-bajo-lema-no-soy-un-virus.html>

- Donald Trump's statement calling COVID-19 the "**Chinese virus**" or even the "**Kung flu**" led to the virus becoming increasingly associated with Asians.
- The WHO made a statement warning of the importance of the name of the disease for some **religious and ethnic communities**.

# OBJETIVES

- In the context of Twitter conversations about Covid-19 and China, we will **examine the predominant anti-Chinese discourses** on Twitter and **how they have evolved over the course of the pandemic**.
  - **Over three years**, from 2020 to 2022, to understand the temporal shifts in anti-China attitudes over the longer term.
- Twitter is useful to understand **public sentiments** around racial issues triggered by COVID-19.



# METHODS.

## DATA COLLECTION AND PRE-PROCESSING

Data collection tool	Tweepy library (Python)
Access API	API v2 Twitter
Search syntax	“COVID” AND (“China” OR “Chinese” OR “Wuhan”)
Language of tweets	English
Time window	01/01/2020 - 31/12/2022
Total number of tweets in the sample	3,000,973
Retweets	No

- DATA COLLECTION
- TOKENISATION
- LEMMATISATION

# METHODS.

## CLASSIFICATION OF TWEETS (I)

- **Spark NLP library** → an **open-source natural language processing library** developed by John Snow Labs, designed to be scalable and fast and can be used on large datasets.
- “Classifierdl\_use\_cyberbullying” model, a **pre-trained text classification model** used multiple times to **identify cyberbullying-related** content in a given text corpus.
- The algorithm was trained on **English-language Twitter data** and specifically labelled to identify **racist, neutral, or sexist narratives** by John Snow's team.
- **92% accuracy.**

# METHODS.

## CLASSIFICATION OF TWEETS (II)

- **Manually evaluation** of the machine-generated classification.
- We extracted a **simple random sample of 1,500 tweets for each year** (a total of 4,500 tweets) and, without consulting the results provided by the machine, **two experts coded the tweets as anti-Chinese/non anti-Chinese.**
- The results provided by the machine were compared with this manual coding, obtaining an **87% coincidence rate** in the tweets labelled as racist by the algorithm and by the two experts.



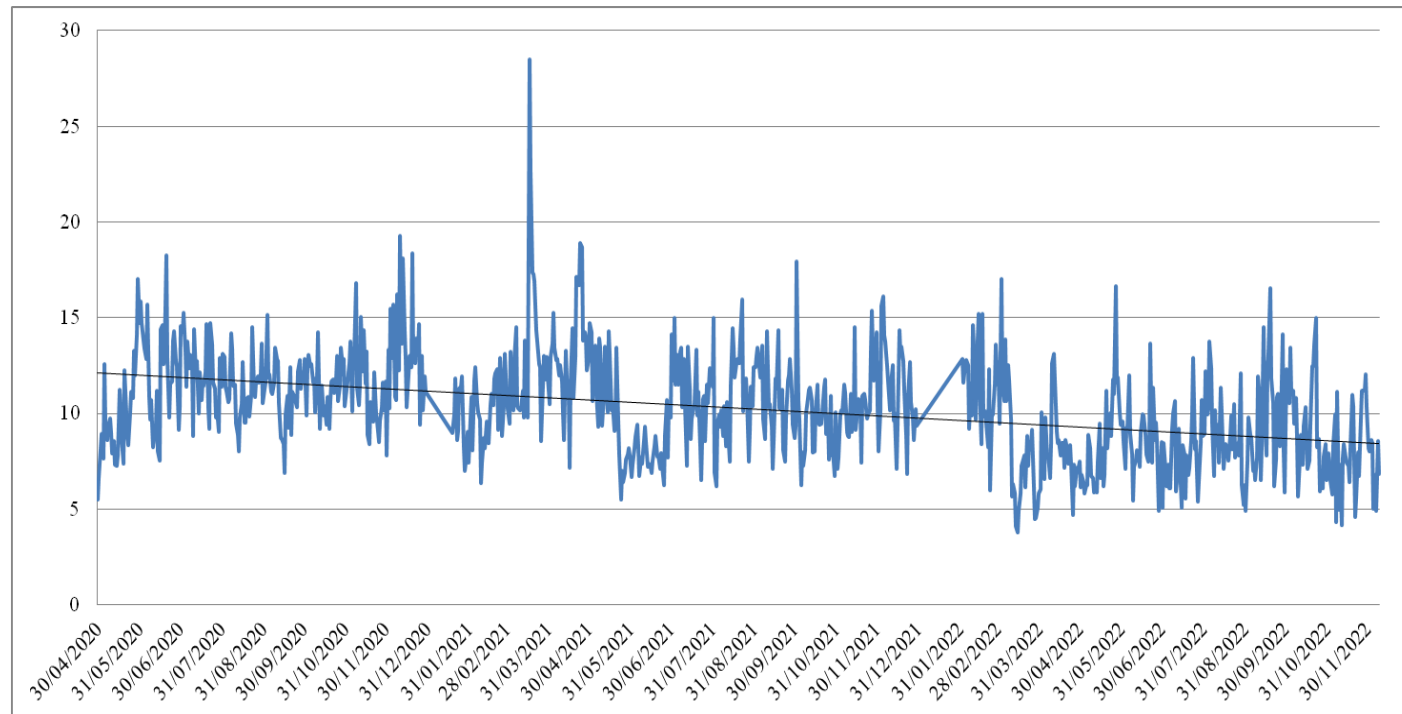
# METHODS.

## TOPIC MODELING

- **LDA algorithm** → 10 clusters.
- 10 clusters into 4 themes.
- After reviewing the main keywords of each topic, a **random sample of each topic was analysed in depth** by reading and interpreting the tweets and a title was given to summarize the content of each topic.

# RESULTS

Figure 1. Percentage distribution of tweets classified as anti-Chinese/racist over the total number of tweets.



Source: Authors.

- **10% of the sample** contained a significant degree of anti-Chinese or racist speech [**89% accuracy**].
  - 37,5% in 2020
  - 34,3% in 2021
  - 28,1% in 2022.
  
- **Slight decrease** over the months except for peaks of high activity.
  
- March and April 2021 → report published by the WHO about the **origin of the virus**.

# IMPERIALISM, GEOPOLITICAL TENSIONS AND DISCRIMINATION (39.6%)

## Economic dependence

- Messages accusing the sale of masks and vaccines to African and Latin American countries to limit their economic autonomy by acquiring debt that is impossible for these countries to pay off.
- "China is invading us", "they want to expand their empire", "China is buying the world".

## Islamophobia

- Anti-Islamic narratives due to the trade relations between China and some Arab countries about vaccination doses. Hatred towards the Muslim population, exposing the idea of a dictatorial political binomial between communism and Islamism that threatens Western democracies.

## Geopolitical tensions

- China vs India. Instrumentalisation of the pandemic to justify strategic territorial attacks that have been a conflict between the two nations for decades

# ORIGINS OF THE PANDEMIC AND RESPONSIBILITY (23.1%)

## Biological weapon

- Hypotheses, sometimes labelled as conspiracy theories, suggesting that the COVID-19 virus is a biological weapon intentionally created by the Chinese government.
- Also the vaccine as a weapon of control.

## Revenge

- Call for revenge and sanctions on China and its people for the creation of a weapon capable of murdering the population

## Mistrust

- Racist and hateful discourses about the origins and spread of the virus that point to a perverse alliance between the Chinese and Dr. Anthony Fauci.

# RESTRICTIONS ON MOBILITY, CENSORSHIP AND LACK OF FREEDOM (20.1%)

## Censorship

- Constant censorship implemented by China. It is feared that these practices will be established by other Western governments and endanger democracies.
- Policies against the spread of COVID are also strongly attacked. These regulations are a pretext for controlling the world's population and curtailing civil liberties.

## COVID Zero Policy

- Chinese government portrayed as an unscrupulous entity that is capable of murdering dogs and isolating children and minors.

# POLITICAL ALLIANCE AND CONSPIRACY AGAINST THE US GOVERNMENT (17.2%)

## Anti-communism

- Biden allies with Xi Jinping to organise a corrupt election using the pandemic as a tactical tool.
- Chinese population is described with derogatory terms such as "communist invaders", or they are assumed to have perverse intentions, sometimes guided by the CCP.

## Realdonaltrump

- Messages of explicit racism linked to messages of support for Trump and his Twitter account

# CONCLUSIONS

- Anti-Chinese discourses denote **typical characteristics of racism and xenophobia** but also have a strong historical, economical and political component:
  - **Confusion about the object of the hatred** or towards whom it is directed.
  - More explicit attacks on the Chinese government, well-known institutions and the media, BUT **danger** to the **perception of the Asian community**, especially those living abroad.
- Users who incite hatred tend to use issues such as China's controversial politics as an **outlet to spread anti-Asian sentiments** (Lin et al., 2022)
- Linkage of these narratives with **conspiracy theories** and hoax propagation.
  - COVID-19 was the first global pandemic to occur in a **global digital media context**.
  - Rapid circulation of information, but also **disinformation**.

# CONCLUSIONS

- The debates about the origin of the virus, the responsibility for the health crisis and the targeting of the Chinese population continue and are still present for **months after the start of the pandemic**.
- Although the data show a decrease in this trend, they also suggest that **some socio-political phenomena** that are a priori unrelated to the pandemic may **rekindle narratives that incite xenophobia or racism** (like the Ukraine-Russia war).

Proyecto PID2021-123983OB-I00 financiado por:



# COVID-19 Y SINOFOBIA EN TWITTER: UN ANÁLISIS DE LOS DISCURSOS ANTI-CHINA DURANTE LA PANDEMIA

Carolina Rebollo, [carolina.rebollo@dstso.uhu.es](mailto:carolina.rebollo@dstso.uhu.es)  
Alba Taboada Villamarín, [alba.taboada@uam.es](mailto:alba.taboada@uam.es)  
Universidad de Huelva. Grupo ESEIS / COIDESO  
Universidad Autónoma de Madrid

**CONSPIRACY THEORIES AND HATE SPEECH ONLINE:  
COMPARISON OF PATTERNS IN NARRATIVES AND SOCIAL MEDIA ABOUT  
COVID-19, IMMIGRANTS, REFUGEES AND LGBTIQ+ PEOPLE**

*II International Workshop on Methodological Advances and Applications in Science*

12-14 JULY, HUELVA

Proyecto PID2020-115095RB-I00 financiado por:



Funded by the  
European Union  
NextGenerationEU

