



Supporting professional fact-checking: how can NLP/AI help?

Carolina Scarton, University of Sheffield, UK

c.scarton@sheffield.ac.uk

 @carolscarton

SICSS-Edinburgh 2022, Edinburgh, UK, June 13 2022



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



What is a rumour?

Online rumours: “circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety so as to motivate finding out the actual truth” (Zubiaga et al., 2015)

- A rumour can become:
 - True
 - False
- Or remain unverified

Fine-grained classifications are also possible (partly false, misleading, etc.)



Thoughts Camera Action #ProvokingThought
@TCA_Media

5G and Corona on the £20 note 😕

#Covid_19 #cornoravirusuk #coronavirusuk
#5GCoronavirus #5GTOWERS



8:33 AM · Apr 6, 2020 · Twitter for Android



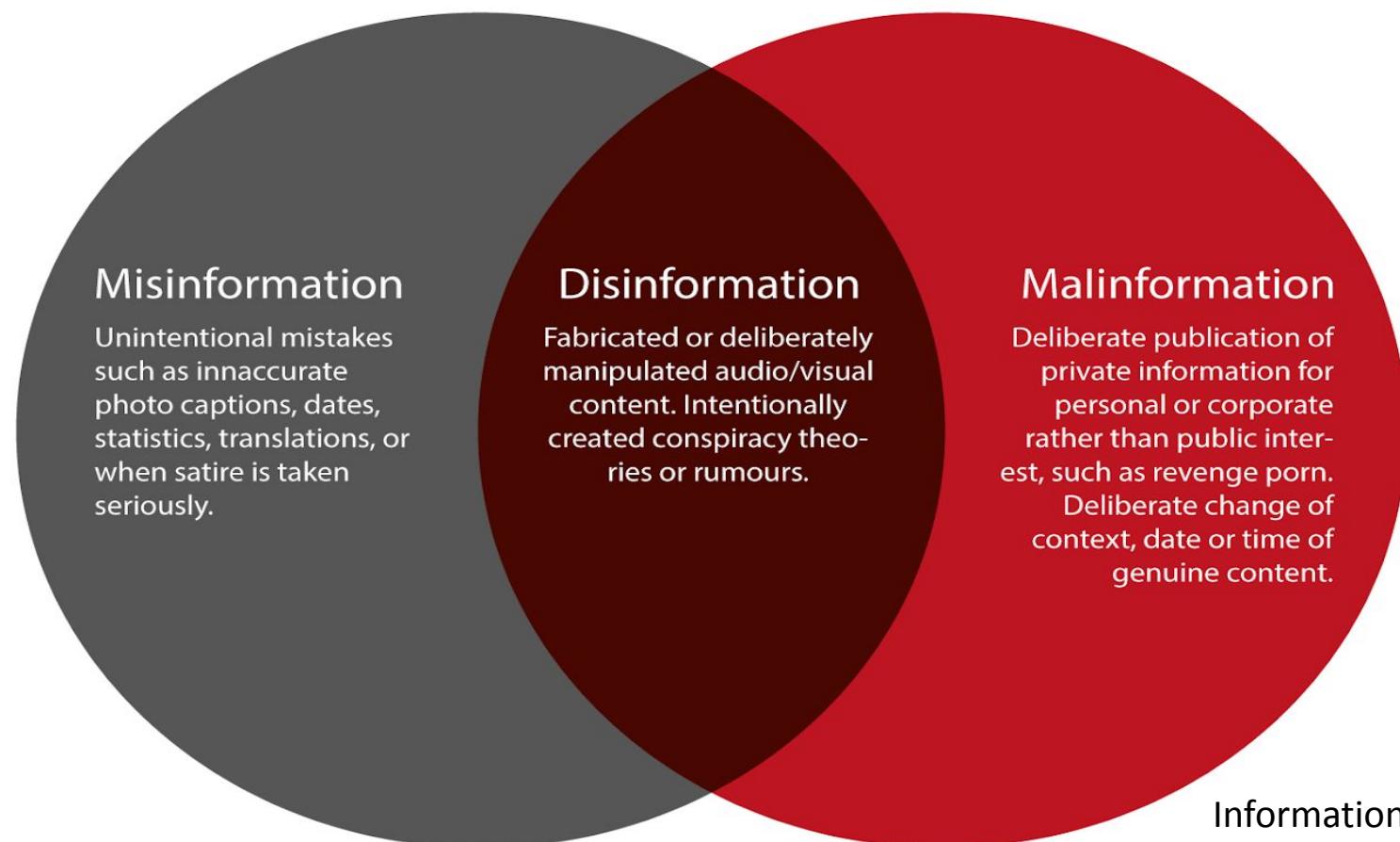
The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



TYPES OF INFORMATION DISORDER

FALSENESS

INTENT TO HARM



Information disorder theoretical framework
(Wardle, 2017; Wardle & Derakshan, 2017)





The 6 questions of disinformation

• What is being spread?

Thoughts Camera Action #ProvokingThought
@TCA_Media

5G and Corona on the £20 note 🤔

#Covid_19 #coronavirusuk #coronavirusuk
#5GCoronavirus #5GTOWERS



0:07 2.7K views

8:33 AM · Apr 6, 2020 · Twitter for Android

Botschaft der Russischen Föderation ✅
@RusBotschaft
Regierungsorganisation aus Russland

🇷🇺 Vladimir #Putin: In the difficult weeks of February-March 2014, residents of #Crimea and #Sevastopol displayed courage and patriotism and formed a wall against the neo-Nazis and radicals, those who staged a coup d'état in #Kiev.
🔗 Full text: [russische-botschaft.ru/de/2022/03/18/...](https://russische-botschaft.ru/de/2022/03/18/)



7:00 AM · Mar 22, 2022 · Twitter Web App

NOTORIOUS MASS MURDERER AND WHITE SUPREMATIC, SAM HYDE CONFIRMED AS SHOOTER AT YOUTUBE HQ #YoutubeHQShooting #YouTube #Shooting #MarchForOurLives #NeverAgain



3:13 PM - 3 Apr 2018





The 6 questions of disinformation

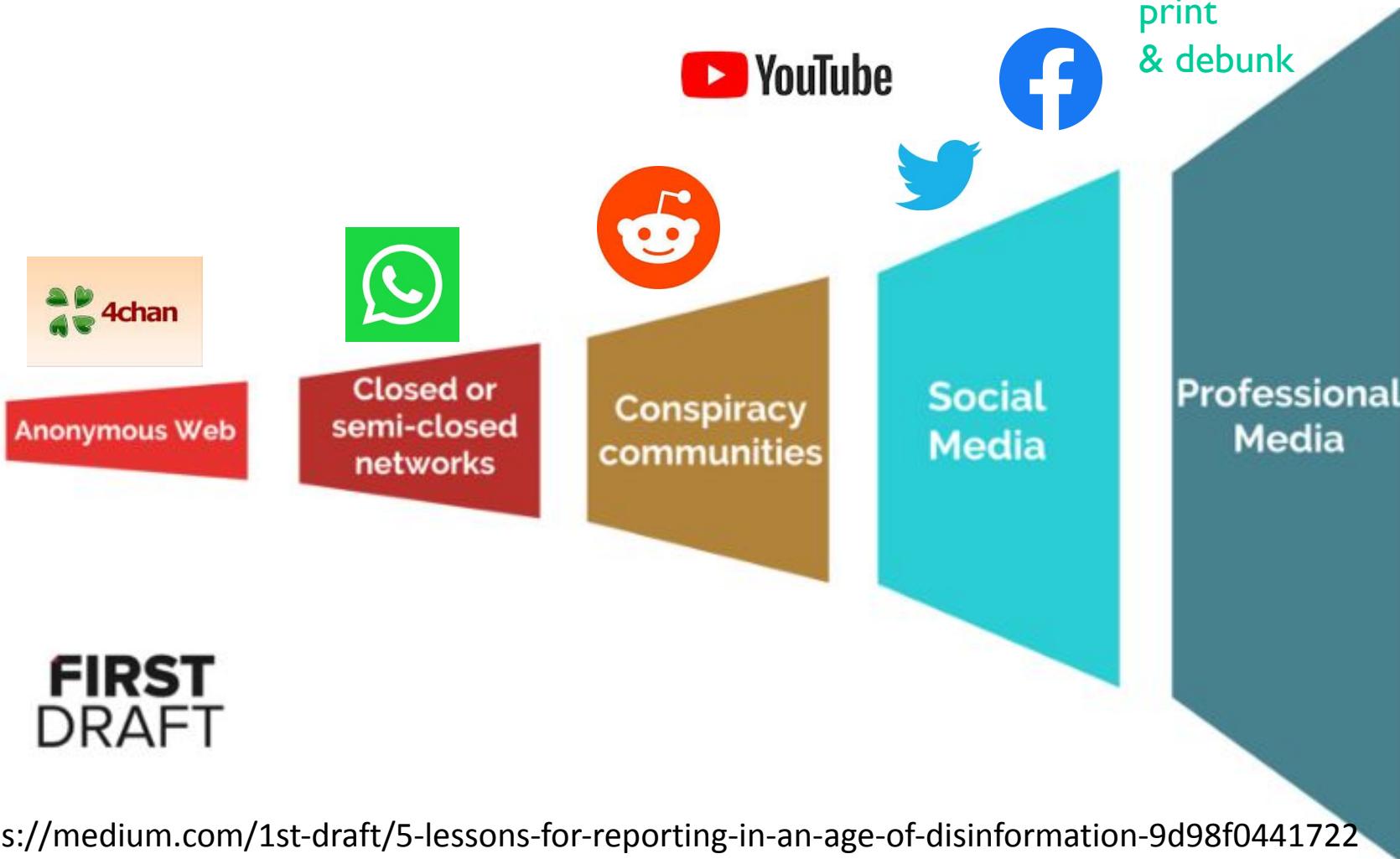
- **What** is being spread?





The 6 questions of disinformation

- Where/How it spreads?



**FIRST
DRAFT**

<https://medium.com/1st-draft/5-lessons-for-reporting-in-an-age-of-disinformation-9d98f0441722>



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



The 6 questions of disinformation

- Where/How it spreads?



- A network of **seemingly UK-based far-right** news sites (operated from Eastern Europe)
- Shared and amplified through **thirteen related Facebook pages**
- **2.4 million likes** (more than any UK Facebook political page)

Source: (Reynolds, 2018)





The 6 questions of disinformation

- **Who** is spreading it? / **Why** it spreads?

Fake Profiles

- IRA operated **cyborg accounts** spread misinformation during the 2016 US elections
 - **Jenna Abrams** - widely considered a real American
- Fake accounts try to **gain credibility by following genuine accounts**
 - As many as 60% of Donald Trump's followers being suspected fake accounts (Campoy, 2018) → over 40% for Hillary and Obama

Advertising and Clickbait

- **Online advertising** used extensively to **make money** from junk news sites
- A **clickbait** post is designed to provoke **emotional response** in its readers





The 6 questions of disinformation

- **Who** is spreading it? / **Why** it spreads?

Micro-targetting or dark ads

- Online adverts that are **visible only to the users that are being targeted**
 - e.g. voters in a marginal UK constituency (Cadwalladr, 2017a)
- **Do not appear** on the advertiser's timeline or in the feeds of the advertiser's followers
- **Micro-targeting:** fine-grained ad targeting, based on job titles or demographic data





The 6 questions of disinformation

- **Who** is spreading it? / **Why** it spreads?

Political advertising online

- **Issues around digital campaigning by political organisations** during the 2016 UK EU membership referendum (DCMS report, 2018)
- **£350m false claim: 10.2 times more tweets** than the 3,200 tweets by the Russia-linked accounts suspended by Twitter
 - More than **1,500 tweets from different voter**





The 6 questions of disinformation

- Who is spreading it? / Why it spreads?

Amplifiers: Bots

- Bots disseminated anti-vaccine messages
- Russian trolls promoted discord
- Accounts masquerading as legitimate users create false equivalency



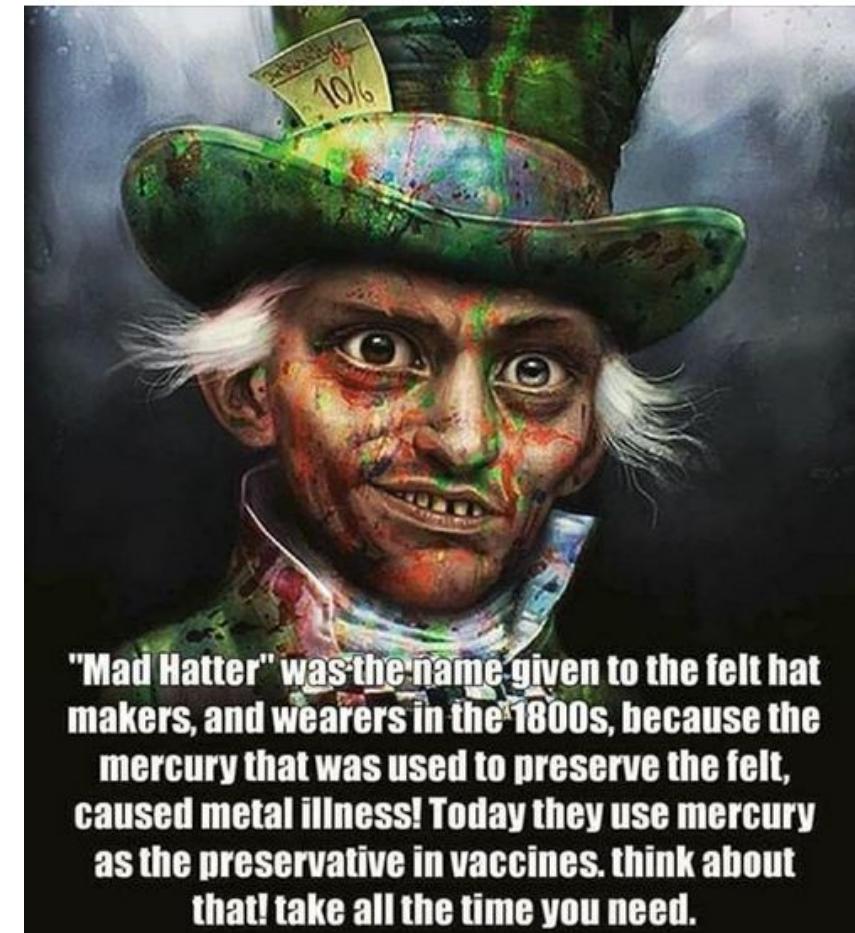
Am J Public Health. 2018 October; 108(10): 1378–1384.

Published online 2018 October. doi: [10.2105/AJPH.2018.304567](https://doi.org/10.2105/AJPH.2018.304567)

PMCID: PMC6137759

PMID: [30138075](https://pubmed.ncbi.nlm.nih.gov/30138075/)

Not safe to eat in fish Or to wear on your head...but in vaccines, it's safe in the body? 😊





The 6 questions of disinformation

- **Who** is spreading it? / **Why** it spreads?

Amplifiers: Fake groups

- **Astroturfing:** creating artificial appearance of “grass-roots” support
- Initially **seeded with fake accounts**, before drawing in genuine users
- Other fake groups are created to “spread **sensationalistic** or **heavily biased** news or headlines, often distorting facts to fit a narrative” (Weedon, Nulan & Stamos, 2017)





The 6 questions of disinformation

- **Who** is spreading it? / **Why** it spreads?

Amplifiers: genuine human users

- The **main amplifiers** behind viral misinformation and propaganda are (Vosoughi et al, 2018)
 - **Confirmation bias** → reading news that **conforms to the individual's political views**
 - **Homophily** → Individual's information sharing and commenting behaviour is influenced by the **behaviour of their online social connections**
- Confirmation bias and homophily lead to the creation of **online echo chambers** (Quattrociocchi et al. 2016)





The 6 questions of disinformation

- Who is spreading it? / Why it spreads?

Amplifiers: mainstream media

globo.com | g1 | globoesporte | gshow | vídeos

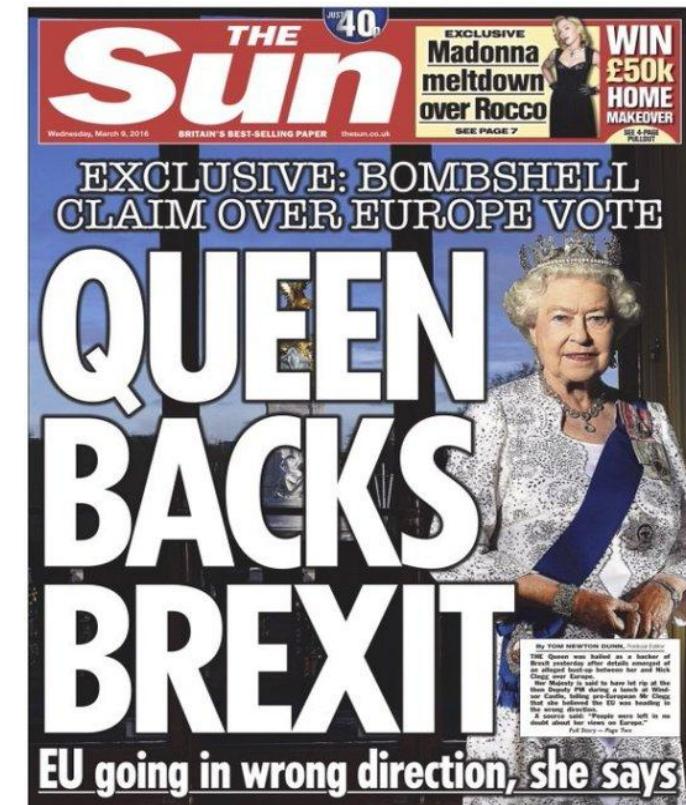
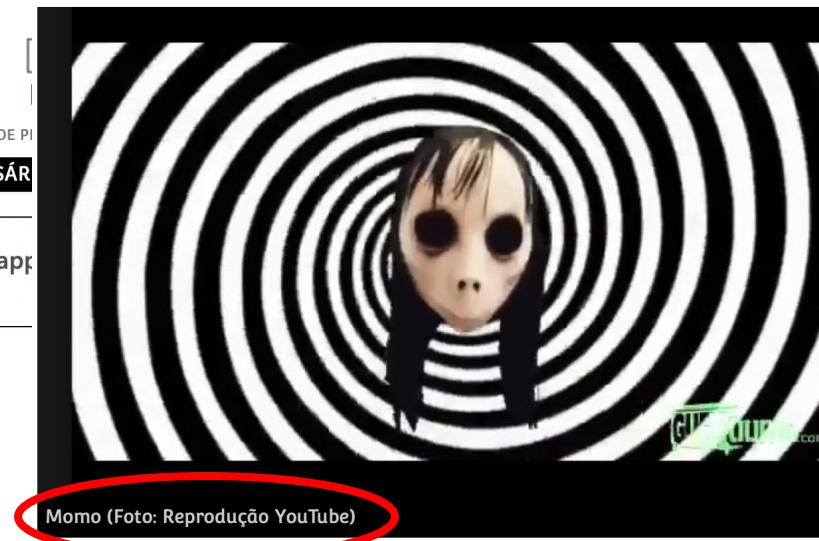


Crescer *

Momo aparece em vídeos de slime do YouTube Kids e ensina as crianças a se suicidarem, diz mãe

Professora relata à CRESCER como descobriu que sua filha teve acesso ao boneco assustador enquanto assistia a alguns de seus vídeos favoritos no aplicativo

<https://revistacrescer.globo.com/Criancas/Seguranca/noticia/2019/03/momo-aparece-em-videos-de-slime-do-youtube-kids-e-ensina-criancas-se-suicidarem.html>



<https://www.ipso.co.uk/news-press-releases/press-releases/ipso-upholds-complaint-against-the-sun-s-queen-backs-brexit-headline/>

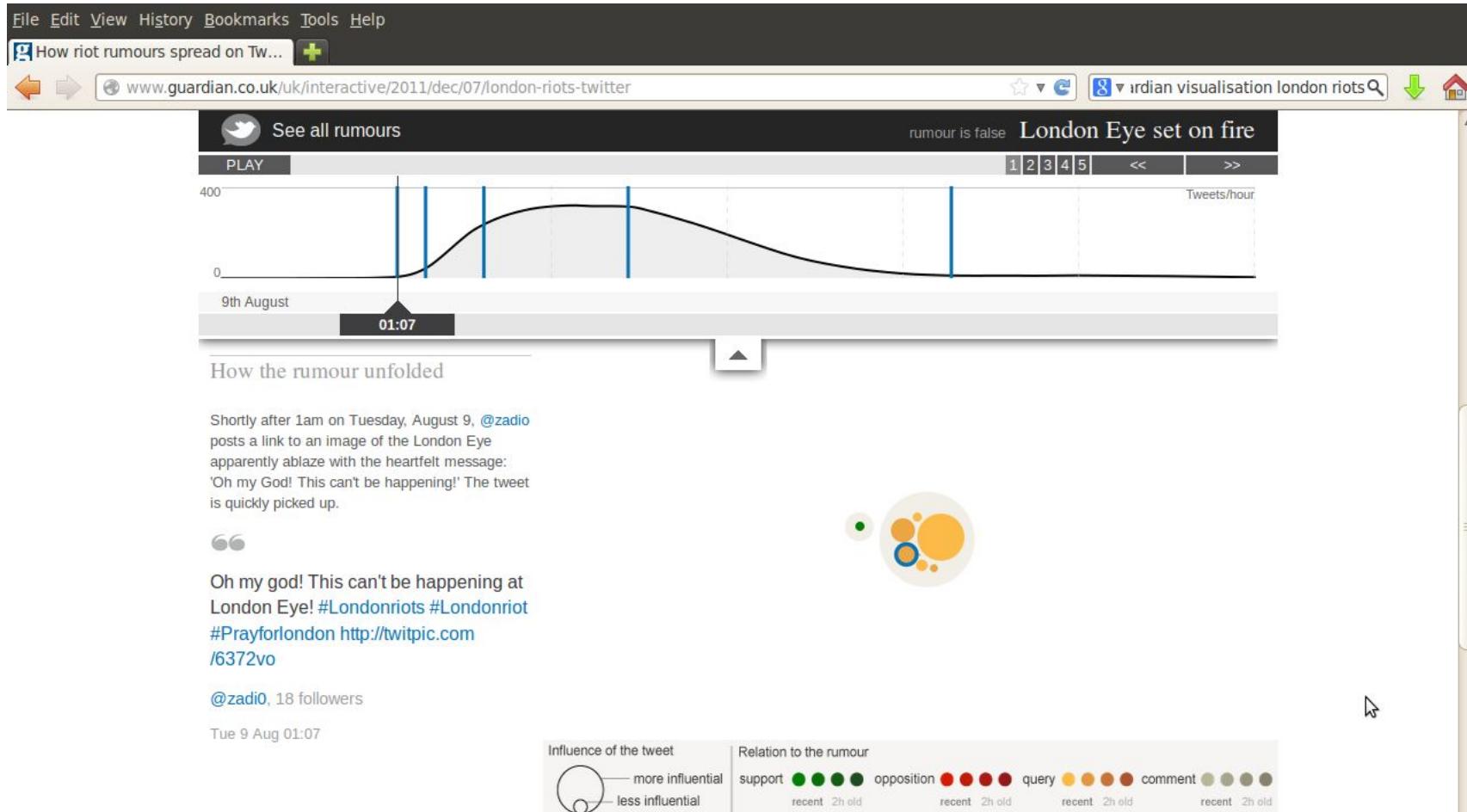




The 6 questions of disinformation



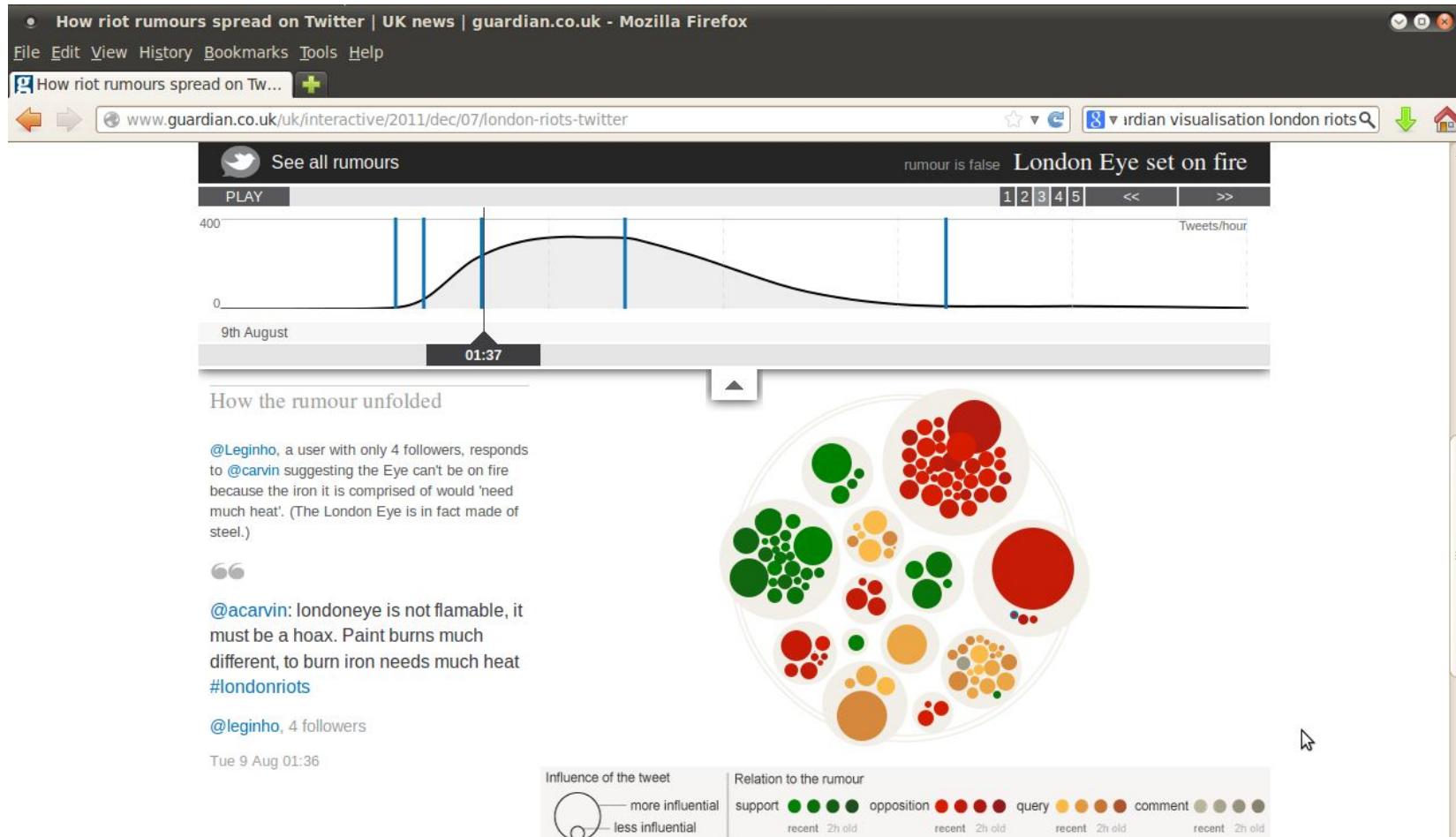
• When it spreads?





The 6 questions of disinformation

● When it spreads?





A diet rich in alkaline foods can eliminate the coronavirus

Date: 02/04/2020

Country: Indonesia

Fact-checker: TEMPO

Language: Indonesian

Date: 06/04/2020

Country: Venezuela

Fact-checker: Efecto Cocuyo

Language: Spanish

Date: 12/09/2020

Country: Turkey

Fact-checker: Teyit

Language: Turkish



Date: 01/03/2020

Country: Spain

Fact-checker: Maldita.es

Language: Spanish

Source: WhatsApp

Date: 02/04/2020

Country: US

Fact-checker: LeadStories

Language: English

Date: 16/06/2020

Country: Spain

Fact-checker: EFE Verifica

Language: Spanish

Date: 09/12/2020

Country: Brazil

Fact-checker: AosFatos

Language: Portuguese

Flag icons made by Freepik from www.flaticon.com

Singh et al. 2021: The False COVID-19 Narratives That Keep Being Debunked: A Spatiotemporal Analysis

<https://arxiv.org/pdf/2107.12303>



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.





Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.
- Involves (but not limited to):
 - **Identifying** potential online rumours
 - predict impact





Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.
- Involves (but not limited to):
 - **Identifying** potential online rumours
 - predict impact
 - **Deciding** what can be verified
 - e.g. identifying where evidence can be found
 - consider time to debunk vs. amplification





Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.
- Involves (but not limited to):
 - **Identifying** potential online rumours
 - predict impact
 - **Deciding** what can be verified
 - e.g. identifying where evidence can be found
 - consider time to debunk vs. amplification
 - Finding **previously debunked cases**





Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.
- Involves (but not limited to):
 - **Identifying** potential online rumours
 - predict impact
 - **Deciding** what can be verified
 - e.g. identifying where evidence can be found
 - consider time to debunk vs. amplification
 - **Finding previously debunked cases**
 - **Collecting evidence**
 -





Professional fact-checking

- Complex task involving **forensic analysis**
 - videos, images, text, etc.
- Involves (but not limited to):
 - **Identifying** potential online rumours
 - predict impact
 - **Deciding** what can be verified
 - e.g. identifying where evidence can be found
 - consider time to debunk vs. amplification
 - Finding **previously debunked cases**
 - Collecting **evidence**
 - Writing the **debunk**





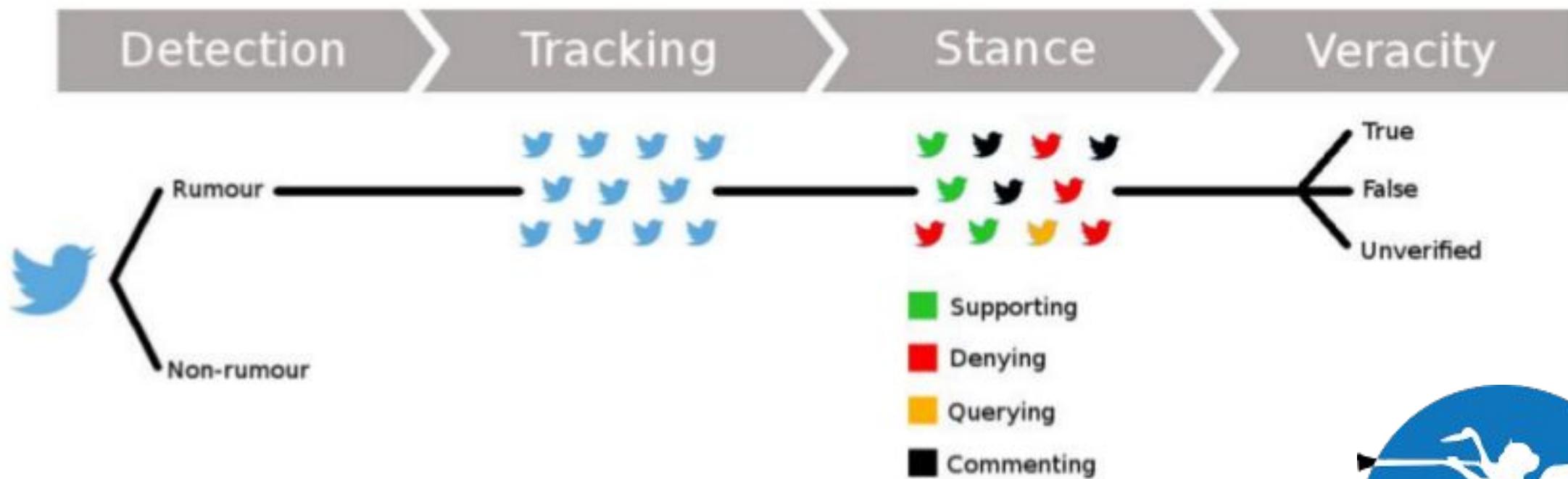
Tools for professional fact-checking

- Perhaps a "step back" from automatic fact-checking
- Telling a fact-checker whether a **rumour is true/false** is not useful
- **Automatically writing a debunk** is also not very useful
- Tools should **support their task** in each stage of the debunking process
 - e.g. detecting tampered images / deep fake; claim retrieval; stance classification...





Tools for professional fact-checking

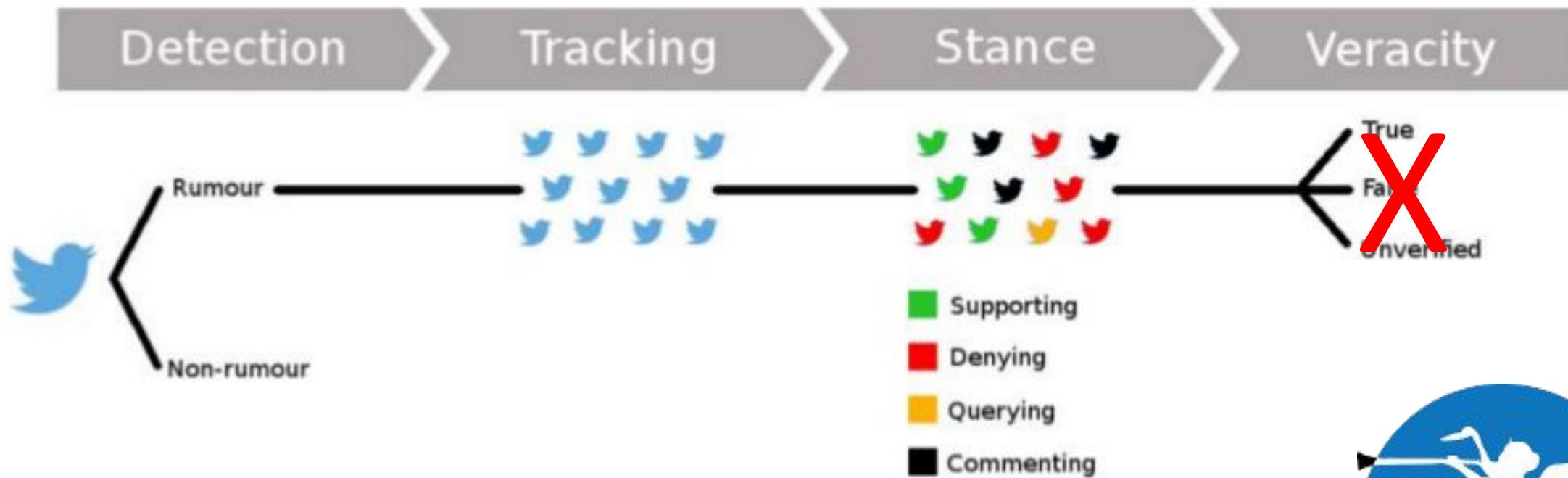


Source: (PHEME project)





Tools for professional fact-checking



Source: (PHEME project)





Truly Media: Collaborative Cross-Modal Verification Workbench



Feeds Explore Monitor

+ Add collection

George Floyd update 6 days ago

Howard University update 19 days ago

COVID-19 & 5G updated 4 hours ago

CrowdTangle This collection will contain... updated 4 hours ago

COVID casualties updated a month ago

Hanau Shootings updated 2 months ago

Coronavirus

George Floyd update 6 days ago

Howard University update 19 days ago

COVID-19 & 5G updated 4 hours ago

CrowdTangle This collection will contain... updated 4 hours ago

COVID casualties updated a month ago

Hanau Shootings updated 2 months ago

COVID-19 and 5G

European Commission EU #UnitedAgainstCoronavirus @EU_Commission 12:08PM - 10 Jun 2020

In the midst of the pandemic, it is important we follow the science and get our facts straight:

The #coronavirus does not spread through 5G telecommunication networks.

Here are 4 important facts 📈 #EUvsDisinfo #FactsMatter <https://medium.com/@EuropeanCommission/5g-and-the-coronavirus-the-one-myth-weve-all-heard-of-10d33de26045>

52 118

TechCrunch @TechCrunch 1:17AM - 10 Jun 2020

Twitter starts putting fact-checking labels on conspiracies about 5G and COVID-19 <https://tcrn.ch/3dMx3m0> by @tayhatmaker

46 108

Bloomberg QuickTake News 3:01AM - 10 Jun 2020

Coronavirus Updates: One Blood Type May Protect Against Virus; Covid-19 Is Fauci's 'Worst Nightmare'

Live

410.1k 2.1k 1.8k

24horas.cl 9:50PM - 4 Jun 2020

Covid-19: cobertura multiplataforma del Departamento de Prensa de TVN por coronavirus

New feed

Fill out a name for the new feed

COVID-19 YouTube

Select a Social Media Platform

Twitter Facebook YouTube VKontakte LinkedIn

Search criteria

coronavirus, COVID-19

Translate to

Add feed





Truly Media: Collaborative Cross-Modal Verification Workbench

COVID-19 & 5G ^① SHARED

Type URL to add... Add CSV CSV Comment

Any status ▼ false-claim ▼ Any source ▼ Any assignee ▼ ★ 🔍 Date added (newest) ▼ List Grid 1 2 > >>

Any tag ▼

- false-claim
- fact-check
- consequence

lesmoutonsense 5:30PM - 6 May PENDING

[MàJ] 5G-Nocide – Première corrélation repérée entre la 5G et une Mutation Virale ou Bactérienne ?! – Les moutons enragés

Lorsque je suis tombé sur cet article de Sputnik la semaine dernière : Ce que l'on sait du nouveau virus qui a déjà fait un mort (2 à l'heure où j'écris ces lignes) en Chine et se propage à l'...

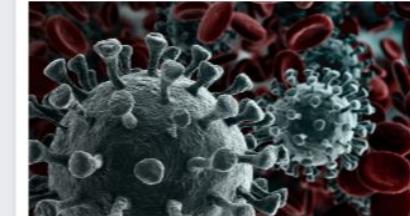


false-claim Add tag

worldtruth.tv 11:56AM - 6 May 2020 RAW

Does 5G Exposure Alter The Structure and Function of Hemoglobin, Causing Coronavirus Patients To Die From Oxygen Deprivation? | WorldTruth.Tv

Is 5G to partially blame for coronavirus deaths? Can 5G cause the blood that's circulating in your body to be unable to carry oxygen? What's especially horrifying about the critical care of coronavirus ... [Read more](#)



false-claim Add tag

kerberos-hellas.blogspot.com 6:06PM - 4 May 2020 RAW

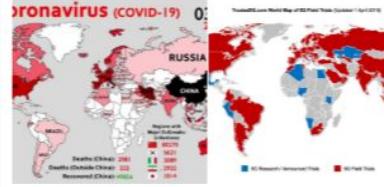
ΚΕΡΒΕΡΟΣ: 5G ΚΑΙ ΚΟΡΟΝΟΙΟΣ ΑΠΟΤΕΛΕΙΩΝΟΥΝ ΤΟΝ ΑΝΘΡΩΠΙΝΟ ΟΡΓΑΝΙΣΜΟ! Στερητικό σύνδρομο οξυγόνου από το δίκτυο 5G Δρ Μέτσης: «Η τεχνολογία δεν μεταδίδει τη νόσο, αλλά η ενέργεια της συχνότητας (60 GHz) απορροφάται από τον οργανισμό Την ώρα πο...



false-claim Add tag

Diane Alexander @dalexan909 6:23AM - 7 Apr 2020 PENDING

@ChristinePolon1 I just now started researching 5G. I found 5G coverage worldwide to shockingly mirror COVID-19 worldwide. The most concentrated areas of 5G = the most concentrated areas of corona. Countries that don't have 5G? Few corona issues. Could this be mere coincidence? Statistical odds?



Comment 0

85 110 false-claim Add tag Collection Notes





Truly Media: Collaborative Cross-Modal Verification Workbench

Verification Tools

< Back to "Coronavirus"

Checklist

- Latitude Longitude
- Preservation link
- Related Links**
- Content
 - Reverse image search results are related to:
 - No results found
 - Same event
 - Other event
 - Is this the original?
 - Yes
 - No

Comments:
It was fact-checked and found fake at
<https://gbrews.com/url.html?>
<https://www.boomlive.in/fake-news/video-shows-chinese-policemen-killing-coronavirus-patients-factcheck-6885>
 - Earliest version
 - Date
 - Time
 - Time zone

Similar content (add url, etc.)
<https://gbrews.com/url.html?>
<https://www.boomlive.in/fake-news/video-shows-chinese-policemen-killing-coronavirus-patients-factcheck-6885>
 - Is metadata available?
 - Yes
 - No
 - Add location for weather information
- Source
- Crediting & Permission
- Attachments
 - Upload attachment

Content **Source** **Location** **Context**

Name: Viber video
Description: Video of chinese people being shot

Multimedia

VIDEO

00:00 00:01 00:02 00:03 00:04 00:05 00:06 00:07 00:08 00:09 00:10 00:11

KEYFRAMES

ANNOTATIONS

IMAGES

SHARED IMAGES

David Icke
@davidicke View mentions

Official Twitter of ground breaking Author & Public Speaker, David Icke. Watch David's Brand New Film 'Renegade' Here - <http://radi.al/Renegade> #davidicke

Translate

Isle of Wight, England

<http://www.davidicke.com>

Joined August 18, 2009

GMT+0300

Followers you follow

Tweets 43,951

Following 269

Followers 333,122

Listed 2,301

ITEM NOTES





WeVerify / InVID plugin



TOOLS

ASSISTANT

TUTORIAL

DEMO

CLASSROOM

ABOUT

English



Advanced tools

The advanced tools are unlocked

LOGOUT



Video



Image



Search



Data Analysis



Image analysis

NEW

It brings you contextual information of an image posted on Facebook or Twitter



Magnifier

It provides a magnifying lens and a photo editor to help you examine an image thoroughly



Metadata

It extract metadata for jpeg images and videos (in mp4 or m4v format)



Forensic

↑

It provides an enhanced toolkit to detect image forgeries



OCR

NEW

It provides optical character recognition for texts in images



CheckGif

NEW



It compares an original image and a tampered one (including if the image has been cropped) and generates a GIF as output

Feedback

“Swiss Army knife” toolset → designed with end users



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



WeVerify / InVID plugin: ASSISTANT



The assistant will help you to analyse a webpage, an image or a video file and suggest which WeVerify tools are useful for each case

Choose what you want to analyse

Webpage link
Insert the link of a webpage, and the assistant will suggest the most useful tools for the contents of the given page.

Local file
The assistant will suggest the most useful tools depending on if the file is a video or image

Feedback

Users are not always clear which tool(s) should be used for their specific use case





WeVerify / InVID plugin: ASSISTANT



TOOLS ASSISTANT GUIDE INTERACTIVE CLASSROOM FACT CHECK ABOUT

Assistant

The assistant will help you to analyse a webpage, an image or a video file and suggest which WeVerify tools are useful for each case

Choose what you want to analyse

Webpage link

Insert the link of a webpage, and the assistant will suggest the most useful tools for the contents of the given page.

Local file

The assistant will suggest the most useful tools depending on if the file is a video or image

Give the URL of the page to analyse

Paste the URL here

<https://ghanafact.com/2021/06/fact-check-breath-holding-exercise-prescribed-in-viral-video-is-not-safe/>

CLEAN

Fact Checker

Some elements of this content have been highlighted

Warning

Some elements of this content have been flagged

URL Media

Media

Potential Tools

Below are the tools you can use on this media type

Magnifier

Examine the image thoroughly

Metadata

Retrieve image/video metadata



The WeVerify project has

under grant agreement No 825297



WeVerify / InVID plugin: ASSISTANT



Give the URL of the page to analyse

Paste the URL here

<https://ghanafact.com/2021/06/fact-check-breath-holding-exercise-prescribed-in-viral-video-is-not-safe/>

CLEAN



Fact Checker

Some elements of this content have been highlighted

The domain has been found as part of a domain credibility check by: Duke Reporters' Lab

It has been labelled as: "fact checker"



It has the following comment against it: "A volunteer effort that focuses on misleading statements on social media and messaging platforms. It also monitors public discourse and political promises. The non-profit social enterprise is operated by FactSpace West Africa. It launched without funders and pledged to avoid accepting future grants or contributions that exceed 35% of its budget in order to preserve the project's editorial independence."



Warning

Some elements of this content have been flagged



The text from this URL has matched against the following from the database of known fakes

: Doctors offer advice for preventing COVID-19, symptoms like coughing and fever indicate pulmonary fibrosis, fibrosis is detectable by holding your breath for 10 seconds, drinking water every 15 minutes repels coronavirus



The text from this URL has matched against the following from the database of known fakes

To test for the new coronavirus, take a deep breath and hold for more than 10 seconds. If you complete it successfully without coughing, discomfort, stuffiness or tightness it proves there is no fibrosis in the lungs, basically indicating no infection.



The text from this URL has matched against the following from the database of known fakes

Memes With Advice For Preventing Coronavirus Are NOT Accurate; Gargling With Salt Water & Vinegar Does NOT Kill The Virus

DOMAIN AND CONTENT CREDIBILITY WARNINGS



WeVerify / InVID plugin: ASSISTANT



Assistant

Enter URL

URL mode

URL



Warning

Some elements of this content have been flagged



The text from this URL has matched against the following

[A picture shows a horse that was found living in a flat in Preston. Lar](#)



The image from this URL has matched against the follow

[http://weverify.eu/resource/Claim/67cce9538786d2945161ae2de806](#)



The image from this URL has matched against the follow

[http://weverify.eu/resource/Claim/2b4a4a6dbfa7c8952040410fede80](#)

URL Content

Text

” Horse Kept in a Preston Apartment, Police and animal welfare officers were called to Centenary Mill, where a horse was found living in a flat on his own, officers say they don't know how the horse got there or who it belongs to, but it had food and water and looked to be in good health. The horse is now been kept in a field until someone claims him

Media



?

Potential Tools

Below are the tools you can use on this media type



Magnifier

Examine the image thoroughly



Forensic

Detect any image manipulation



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



WeVerify / InVID plugin: ASSISTANT



Analysed Image

NEW IMAGE

Filters

COMPRESSION TRACES DEEP LEARNING CLONING

Zero GHOST CAGI
Double Quantization DCT BLOCK

No detection Detection

Mouse over the filters to see a transparent mask with the results on the image.

Those filters detect anomalies in previous processes of creation and compression of the image. If a new element has been removed or added to the image, it can be detected if it has a different compression than the rest of the image.

Combination of multiple filters that outline in white or green-light a common area of the image.

Complex textures of an object or saturated areas of the photo can generate false positives.

Enhancers

Error Level Analysis Laplacian Median Noise

IMAGE FORENSICS



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



WeVerify / InVID plugin: ASSISTANT



IMAGE



BOUNDING BOXES



BLOCK DETECTION

● Yes ● No

LANGUAGE DETECTION

Yes No

OB SPECIEY LANGUAGE

Select ×

Clear

Submit

EXTRACTED WORDS

1 : Отвсичко На снета друидите Нил Много Мризека смяната На часовото преме

EXTRACTED TEXT

Отвсичко На снета друидите Нил Много Мризека смяната На часовото преме

LANGUAGE OF TEXT

Predicted Script : Cyrillic

Predicted Language : BG

GOOGLE TRANSLATION LINK

<https://translate.google.co.in/?sl=auto&tl=en&text=Отвсичко%20На%20снета%20друидите%20Ни%20Много%20Мризека%20смяната%20На%20часовото%20преме&op=translate>

IMAGE OCR





WeVerify / InVID plugin: ASSISTANT



IMAGE



BLOCK DETECTION

Yes No

LANGUAGE DETECTION

Yes No

OR SPECIFY LANGUAGE

Select ▾

Clear

Submit

[Direct link to Google translate](#)

BOUNDING BOXES



EXTRACTED WORDS

- 1 : तनम T ऊवगारे
- 2 : नहों होओग पन गानापनरेद
- 3 : mङोiK{ 1]#084" roिमाना arllan Doul
- 4 : 4]]0"80
- 5 : किसान बना बैरागी भारतीय किसान यूनियन जिला जीन्द
- 6 : माउर Uमतमन

EXTRACTED TEXT

तनम ऊवगारे नहों होओग पन गानापनरेद mङोiK roिमाना arllan Doul किसान बना बैरागी भारतीय किसान यूनियन जिला जीन्द
माउर Uमतमन

LANGUAGE OF TEXT

Predicted Script : Devanagari
Predicted Language : HI

GOOGLE TRANSLATION LINK

<https://translate.google.co.in/?sl=auto&tl=en&text=तनम%20ऊवगारे%20नहों%20होओग%20पन%20गानापनरेद%20mङोiK%20roिमाना%20arllan%20Doul%20किसान%20बना%20बैरागी%20भारतीय%20किसान%20यूनियन%20जिला%20जीन्द%20माउर%20Uमतमन&op=translate>

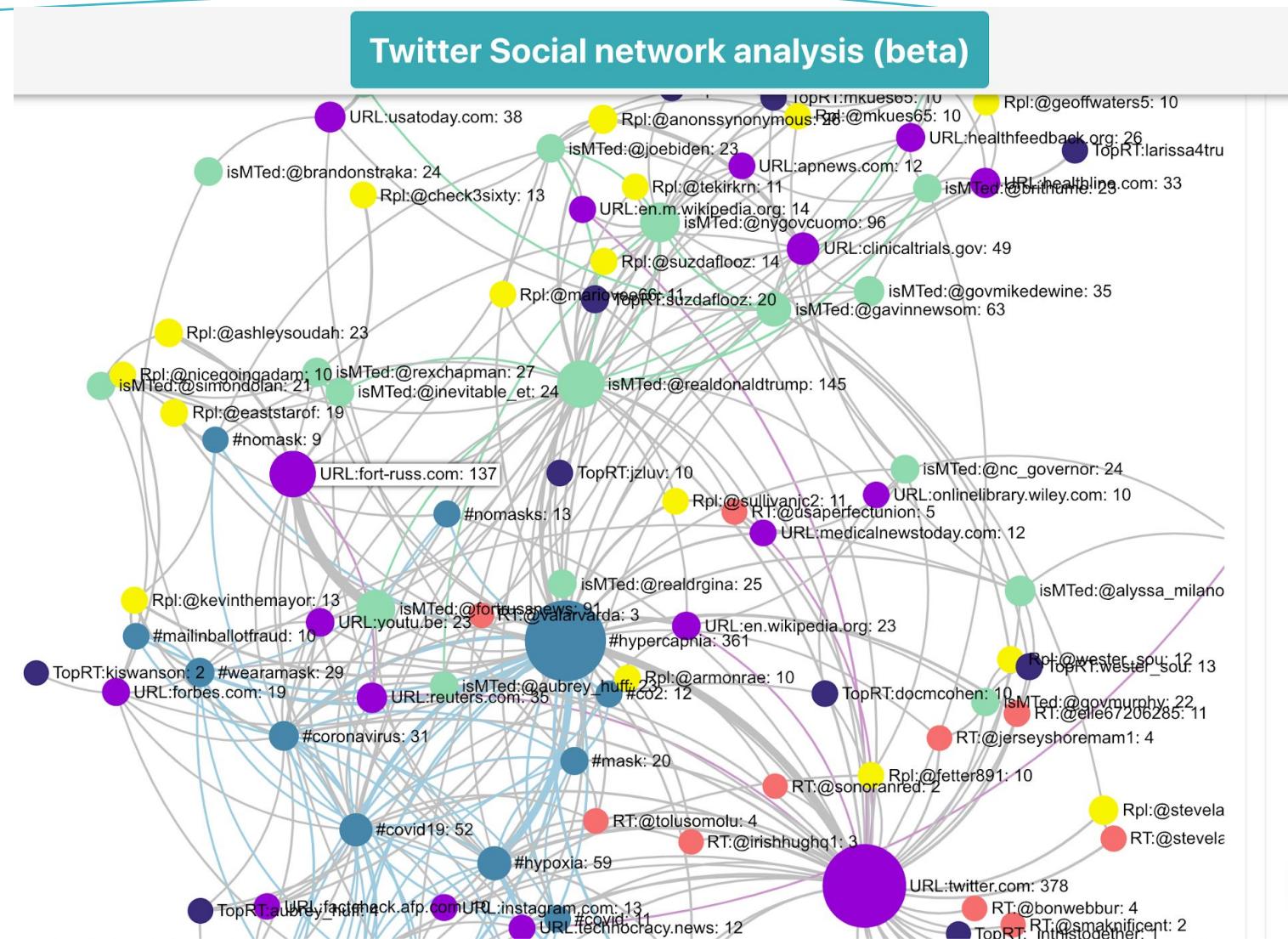
IMAGE OCR



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



WeVerify / InVID plugin: ASSISTANT





Conversation analysis



Twitter Conversation

Analysis of a single Twitter Conversation

Conversation Starter

URL of the Tweet

<https://twitter.com/EmmanuelMacron/status/1164617008962527232>

SUBMIT

Tweet Summary



Emmanuel Macron  @EmmanuelMacron

Officiel du gouvernement - France



Our house is burning. Literally. The Amazon rain forest - the lungs which produces 20% of our planet's oxygen - is on fire. It is an international crisis. Members of the G7 Summit, let's discuss this emergency first order in two days!

#ActForTheAmazon



Emmanuel Macron 

@EmmanuelMacron

Président de la République française.

Joined Twitter 2,836 days ago on the 20th October 2013 and averages 3.67 tweets a day.

724 Following

7,125,662 Followers

10,400 Tweets

11,651 Lists

In here there should be some description of what the tool is showing and how it works, incuding details on how the timeline view works if we are showing the range selector. This would be essentially static

Feedback

Li and Scarton 2020: Revisiting Rumour Stance Classification: Dealing with Imbalanced Data. RDSM 2020

Scarton and Li 2021: Cross-Lingual Rumour Stance Classification: a First Study with BERT and Machine Translation. TTO 2021



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



Conversation analysis

>

8:15 PM · Aug 22, 2019

148.9K 27.8K Copy link to Tweet

We have currently processed 21,257 replies to the selected tweet. The following charts show the overall stance breakdown as well as the timeline of these replies.

Stance	Percentage
comment	66.9%
deny	22.8%
query	5.44%
support	4.89%

Date	Replies
Aug 21	~100
Aug 22	~5,000
Aug 23	~10,000
Aug 24	~3,000
Aug 25	~500
Aug 26	~100
Aug 27	~50
Aug 28	~20
Aug 29	~10

Feedback





Conversation analysis



Replies Explorer

There are 4,844 replies with a stance label of **deny** ✓

 alexandre f. sá
@alecsfusa 

Replies to @EmmanuelMacron

First: it is our house. Not yours.
Second: do take care of French forests that rests. Third: Nasa says Burning in Amazon this year is below the average in the last ten years. Stop lying, Macron

8:27 PM · Aug 22, 2019 

   Copy link to Tweet

The top 100 hashtags appearing in the replies:

#FAKENEWSNO #BrasilAcimaDeTodos
#AmazonNotYouHouseMacron #ClimateWar #AmazonFreeOfOngs #BolsonaroSociopata #DoSomething
#AmazoniaWithoutNGO #NotreDame #macronidiot #mindyourownbusiness
#ActForTheAmazon #Brexit #G7Biarritz #YellowVests #PrayforAmazonia
#BrazilsNotForYou #SomosTodosRicardoSalles #AmazonFires #PrayForTheAmazon #Colonialism
#AmazoniaÉdoBrasil #Amazonisnotyours #MacronOBrasilNaoESuaCasa
#ClimateChange #BJP
#ShutUpMacron
#Europe #fakenews #MacronFake #Bolsonaro #Mercosur
#AmazoniaSemONG #AmazonRainforest #FakeNewsAmazonia #MacronLiar #AmazoniaSOS
#AmazonIsOurs #macroLies #G7 #AmazoniaSemONGs #Brazil #FakePicture
#MacronLeGrandMenteur #BolsonaroTemRazao #G7 #G7France #AmazoniaENossa #ClimateHoax
#AmazoniaSemFakes #Amazon #FakeNews #CalaBocaMacron #MacronIdiota
#FakeNewsNao #AAA #AmazoniaSemFake #EU #ClimateChangelsReal #Amazônia
#EmmanuelMacron #France #ActForAmazonia #AmazoniaEdoBrasil #ActForTheAmazo
#AmazonFire #macronlies #AmazonForest #ActOnClimate
#AmazoniaÉnossa #ClimateCrisis #Macron #BolsonaroPresidenteAte2026 #Amazonia
#PriePourNotreDame #Brasil #ClimateEmergency #MacronFakeNews
#RespeitemASoberaniaDoBrasil #G7Summit #GiletsJaunes #MacronIdiot
#California #ZarifIsAnApologist #amazoniaedobrasil #fake
#AmazoniaFires #giletsjaunes #2020Election #AMAZONIASEMONGS
#EmmanuelFakemacron #CarbonTaxScam #ClimateBreakdown
#FORABOLSONARO #EmmanuelMacronFakeNews
#Bolsonaro2022 #CurrentOccupant





Conversation analysis

>



The ten most active users contributing to this conversation are:

Screen Name	Tweets
ReciteSocial	50 (1.03%)
Marcelajdg	9 (0.19%)
f_a_monteiro	7 (0.14%)
Beliza43623416	5 (0.10%)
flavioemel	5 (0.10%)
AnaMariaRezen11	4 (0.08%)
BRUTO1991	4 (0.08%)
BrVerde_Amarelo	4 (0.08%)
MayDgony	4 (0.08%)

The ten most frequently occurring URLs within the replies are:

URL
https://earthobservatory.nasa.gov/images/145464/fires-in-brazil
https://support.twitter.com/articles/20169199
https://www.alamy.com/stock-photo-amazon-rain-forest-afire-30695050.html
https://www.oantagonista.com/brasil/macron-usa-imagem-de-fotografo-que-morreu-em-2003-para-falar-de-queimadas-sob-bolsonaro/
https://fb.ru/news/environment/2018/11/16/35125
https://twitter.com/dimacgarcia/status/1164352140992794626?s=20
https://www.carbonbrief.org/carbon-emissions-from-amazon-wildfires-could-counteract-deforestation
https://www.jornaldacidadonline.com.br/noticias/16071/macron-apoiou-projeto-de-desmatamento-de-milhares-de-hectares-com-mais-de-120-especies-protedidas-na-amazonia
https://www.science20.com/robert_walker/nasa_say_the_amazon_is_burning_at_below_average_rates_yet_many_news_stories_say_record_rates-240959
http://alamy.com/stock-photo-am

Feedback





Blockchain Database of Known Fakes



WeVerify Dashboard Search Similarity Search New Debunk Zlatina Marinova Logout

Search

Search for Debunks by filters and/or keyword

Filters Reset

Date
08/01/2020 Filter 08/31/2020

Author
All Authors

Organization
All Organizations

Language
All Languages

Has Appearance
All

Has Review
All

Search Result For: COVID

AUTHOR: ALL AUTHORS ORGANIZATION: ALL ORGANIZATIONS FROM: 08/01/2020 TO: 08/31/2020

About : 96 Debunks Order By: Score

DEBUNK 39 days ago

Flu is killing many more people than Covid-19 at the moment. Overview

Full Fact

DEBUNK 55 days ago

A viral Facebook post makes the false claim that the nasal swab test for COVID-19 damages the “blood-brain barrier” and can therefore lead to brain infections. Medical experts say the swab test is completely safe. Overview

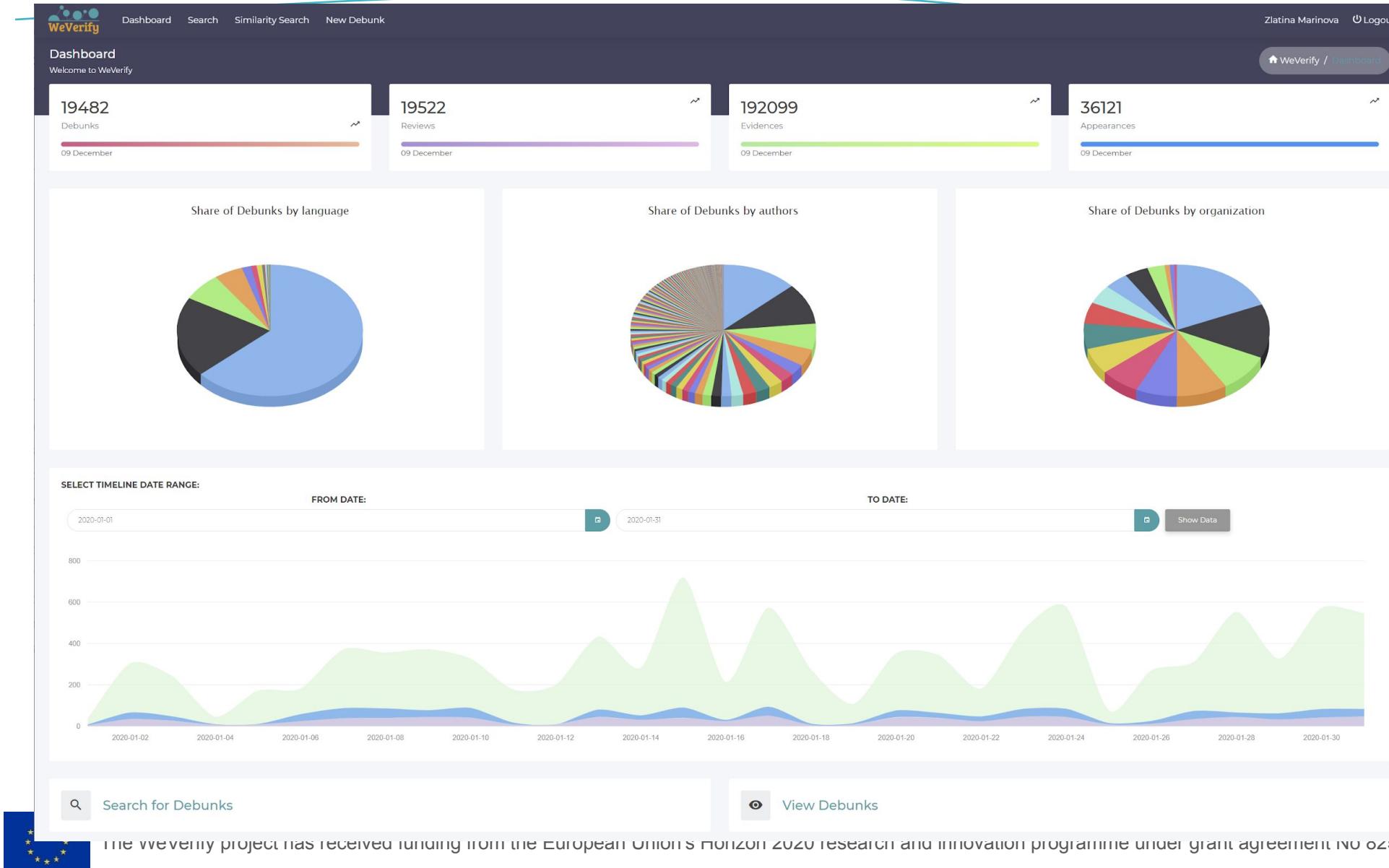
Bala Thenappan

DEBUNK 42 days ago

The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820201.



Blockchain Database of Known Fakes





Blockchain Database of Known Fakes



WeVerify Dashboard Search Similarity

Debunk

Image Claims To Show Antifa Activists Holding An 'Antifa Against Bikers' Banner

Appearances

01/11/2020, 18:11 PM



15/06/2020, 06:06 AM



Trevor Schakohl
External URL: <https://www.facebook.com/p/1281020645>



This is a fight they don't want to start. One they won't win.

1 like 7 comments 7 shares

Verdict: False

The image has been digitally altered. The original banner actually has the words "Free All Antifa Prisoners" on it.

Login

WeVerify / Dashboard / Search Results / Image Claims To Show...

See Evidences

shared on Facebook allegedly shows antifa activists holding a banner that reads "Antifa Against Bikers."

has been digitally altered. The original banner actually has the words "Free All Antifa Prisoners" on it.

have recently circulated online that members of the Hells Angels motorcycle clubs were traveling together to confront antifa, short for anti-fascist, activists.

Being shared appears to reference that debunked claim, showing a crowd of people in front of what appears to be a stage. Some people in front of the photo hold a banner, purportedly bearing the words "ANTIFA AGAINST BIKERS."

"This is a fight they don't want to start," the accompanying post reads. "One they won't win." (RELATED: Video Claims To Show Hells Angels Bikers Riding To Seattle To Fight Antifa Activists) The image appears to have been digitally manipulated to





<https://cloud.gate.ac.uk>

TwitIE Named Entity Recognizer for Tweets

Named entity recognition service for Twitter data. Identifies person, location, organization etc. and also performs normalization of abbreviations and common shorthands ("brb", "gr8", "2day", etc.).

1,200 free requests / day
Larger batches £0.80 / CPU hour

Twitter Collector

Collect tweets, view tweet statistics, and store results in your dashboard for further analysis.

£0.05 / CPU hour

English Tweet Tokenizer

Identifies words and punctuation in tweets

1,200 free requests / day
Larger batches £0.80 / CPU hour

Social Media Tokenizer: French

A tokenizer for French social media text that handles hashtags and user mentions etc.

1,200 free requests / day
Larger batches £0.80 / CPU hour

Social Media Tokenizer: German

A tokenizer for German social media text that handles hashtags and user mentions etc.

1,200 free requests / day
Larger batches £0.80 / CPU hour

Rumour veracity classifier

Classifies tweets based on whether the claims they make are true, false or unverifiable.

1,200 free requests / day
Larger batches £0.80 / CPU hour

Hyperpartisan scorer

Estimates the probability that a document could be classed as "hyper-partisan".

600 free requests / day
Batch processing not available

WeVerify OCR Service

This pipeline identifies URLs in the input text and tries to process them with optical character recognition to identify text in images.

150 free requests / day
Batch processing not available

Stance Classification: English

Classifies a reply to a tweet based on its stance (support, deny, query, or comment) to the original

1,200 free requests / day
Batch processing not available

COVID-19 Claim Categoriser

Assigns claims about COVID-19 to one of several categories, including government actions, impact on community, medical advice, conspiracy theories, vaccine, etc.

1,200 free requests / day
Larger batches £0.80 / CPU hour

Domain Credibility [DEPRECATED]

Annotates URLs to highlight the credibility of the domain at which they ultimately point.

1,200 free requests / day
Larger batches £0.80 / CPU hour

Stance Classification: Multilingual

Classifies a reply to a tweet based on its stance (support, deny, query, or comment) to the original

1,200 free requests / day
Batch processing not available

URL Domain Analysis

Annotates URLs to highlight potential credibility information on the domain at which they ultimately point.

1,200 free requests / day
Batch processing not available

Multi-Lingual OCR Service

A multi-lingual OCR service that detects areas of text, extracts the text from each area, and then determines the language of each block of text.

150 free requests / day
Batch processing not available





The
University
Of
Sheffield.

Specialised tools



- Tools tailored to the **Irish (politics) scenario**
- New data / new models / model adaptation
- **Collaboration with journalists** to fit their needs



<https://edmohub.ie/>



journal**media**group



NEWS
WHIP



Institute for Future
Media & Journalism



The
University
Of
Sheffield.



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



● COVID-19 misinformation classifier

 Label Studio

Labeling

Annotation Guidelines

Claim: An “open letter” purportedly written by US billionaire and philanthropist Bill Gates about the novel coronavirus pandemic has been shared in English and Chinese on Facebook, Twitter and various websites. The Chinese-language posts state the letter was translated from its original publication in British newspaper The Sun.

Explanation: But the letter is fake; Gates’ philanthropic organisation, the Bill & Melinda Gates Foundation, clarified on Weibo that the letter presents “false information” and urged netizens to stop sharing it. The Sun has reportedly removed the letter from its website.

<https://factcheck.afp.com/gates-foundation-urges-netizens-stop-sharing-fake-bill-gates-coronavirus-letter>

Choose one of the following (required):

- Public authority^[1]
- Community spread and impact^[2]
- Medical advice, self-treatments, and virus characteristics^[3]
- Prominent actors^[4]
- Conspiracies^[5]
- Virus transmission^[6]
- Virus origins^[7]
- Public preparedness^[8]
- Vaccines and medical treatments^[9]
- Protest and civil disobedience^[0]
- Social alarm and divisions^[a]
- None of the above / Cannot determine^[q]

Confidence (required):

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

Optional Remarks (optional, press enter to add remark)

Further details:

<https://arxiv.org/abs/2006.03354>





- IFCN dataset:
 - Claim → rephrased by a fact-checker
 - Explanation → why a claim is false
 - Source link → page of the debunk
 - Date → of publication in the IFCN Poynter website
 - Type of media
 - Original website

<https://www.poynter.org/ifcn-covid-19-misinformation/>

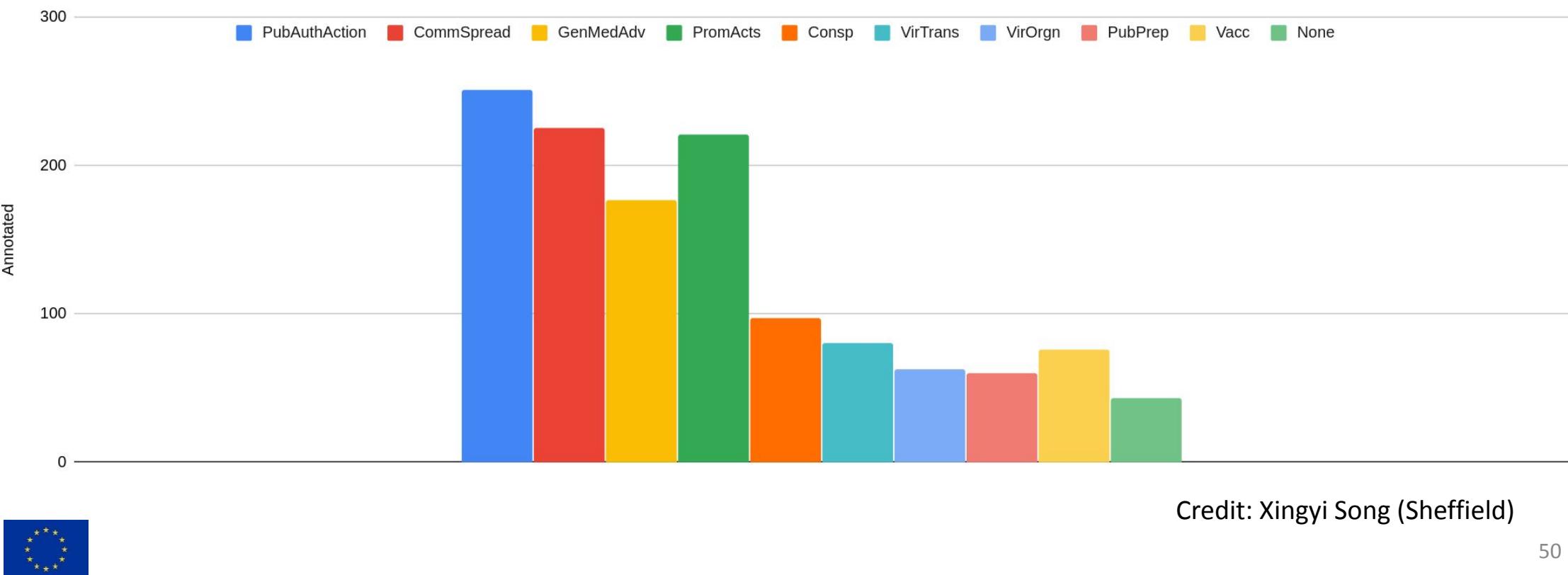




COVID-19 Classifier

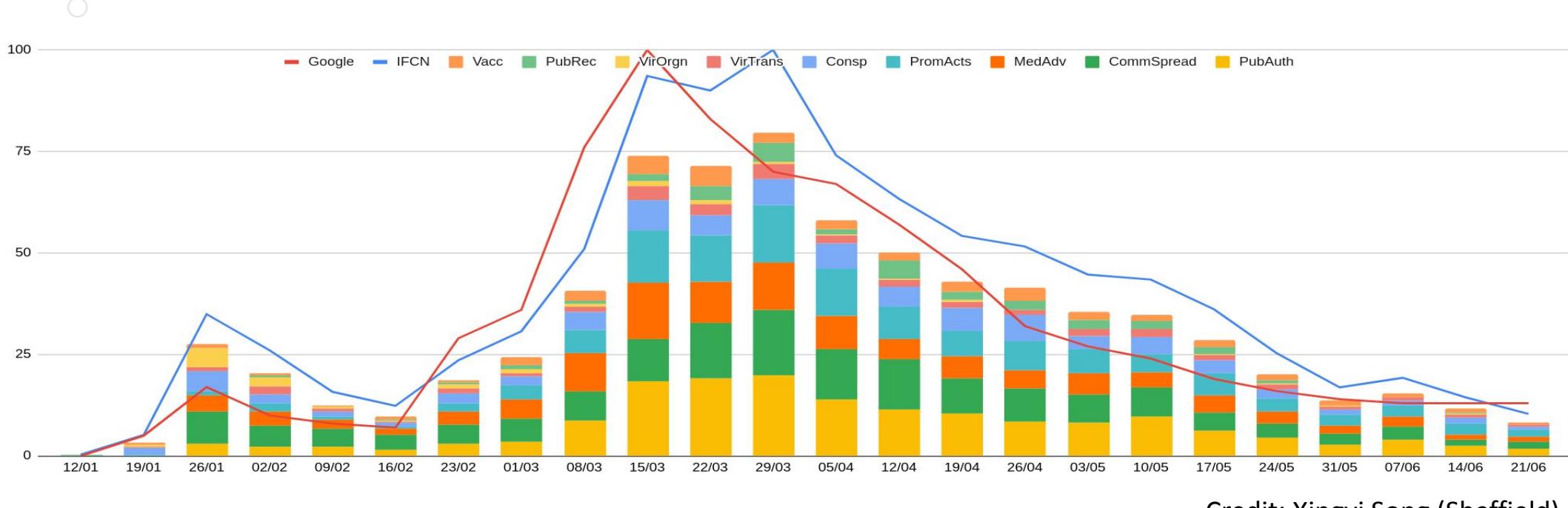
- **Data annotation:**

- 27 volunteers (English claims only)
- Assign most relevant class + a confidence score
- After removing low quality annotations → 1,293 debunks annotated (Cohen's $k= 0.70$)



COVID-19 Classifier

- Large-scale analysis:
 - Start with **Conspiracy theories**
 - Then **Community spread** and **Virus origin** in mid February
 - **Public authority action** dominate in March
 - **Community Spread** and **Prominent actors** are top topics in the later period



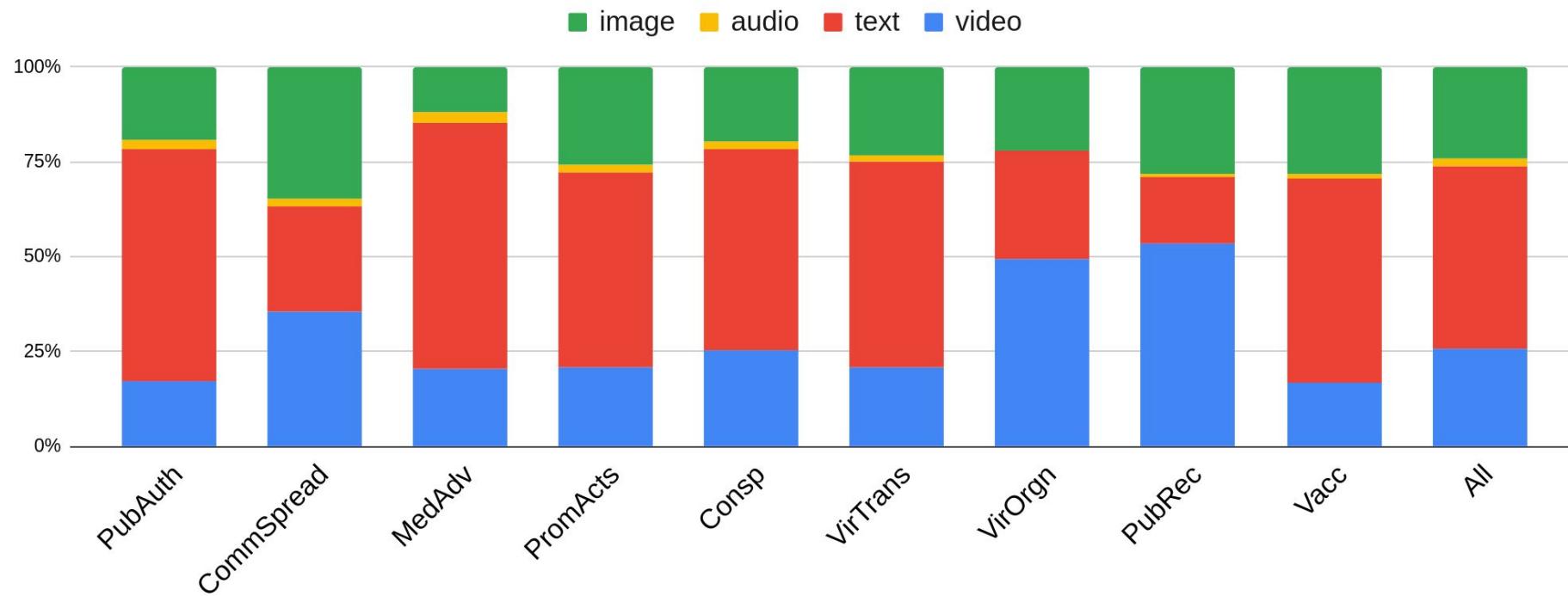
Credit: Xingyi Song (Sheffield)





COVID-19 Classifier

- Large-scale analysis:
 - 50% text
 - ~ 50% of Virus origin and Public preparedness → videos



Credit: Xingyi Song (Sheffield)

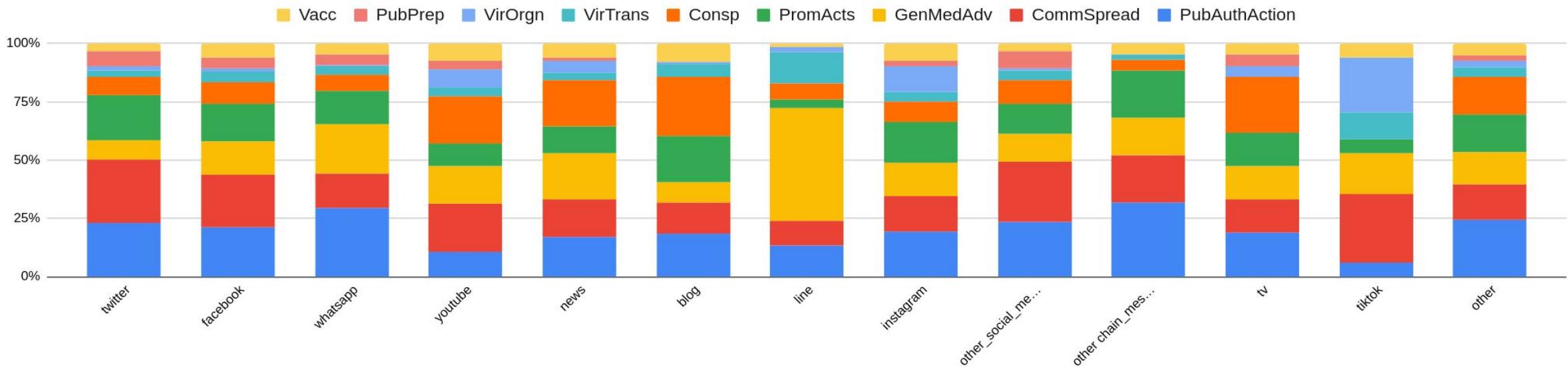




COVID-19 Classifier

- Large-scale analysis:

- Most spread in **social media** → Public Authority action and Community spread
- **Instagram and Tiktok** → Virus origin
- **News, Youtube, Blog and TV** → Conspiracy theory
- **Messages apps** → General medical advise



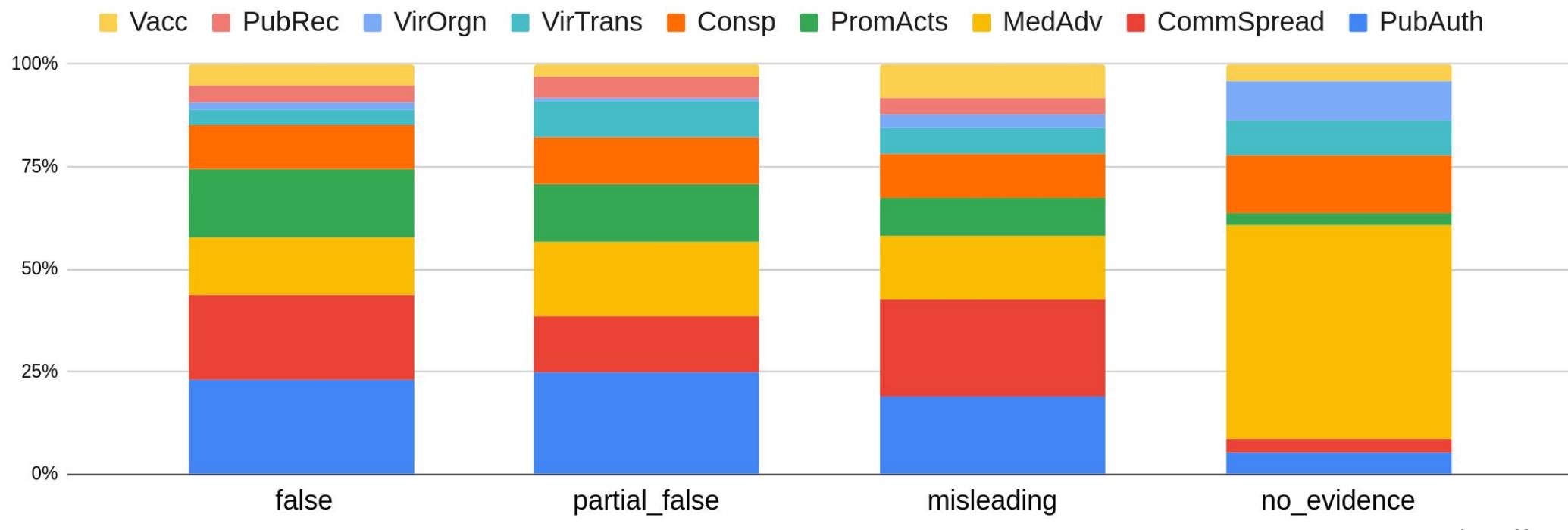
Credit: Xingyi Song (Sheffield)





COVID-19 Classifier

- Large-scale analysis:
 - Community spread → highest in misleading claims
 - ~ 50% of No evidence → General advise



Credit: Xingyi Song (Sheffield)





● COVID-19 misinformation classifier

Fact checked claim	Factchecker	Topic	Engagement ↓
→ The coronavirus “snuck up on us,” adding that it is “a very unforeseen thing.”	PolitiFact	Conspiracies	32,312
→ The Covid-19 was invented in 2004, as shown in a patent filed by the Institute Pasteur	20 Minutes	Conspiracies	30,960
→ This is a phishing scam.	Fullfact.org	Conspiracies	25,771
→ Trump's claims about how quickly WHO acted to announce human-to-human transmission	Washington Post Fact-Checker	Conspiracies	19,546
→ American scientist Judy Mykowitz exposes facts about the US administration regarding coronavirus and accuses the US military of fabricating it in conjunction with the Wuhan Chinese laboratory.	Misbar	Conspiracies	18,682
→ Claim that the coronavirus was created and patented in a French institute (Institut Pasteur).	AFP	Conspiracies	16,679
→ South Korean government secretly requested for certain factory to produce one million facial masks to be sent to North Korea.	JTBC news	Conspiracies	12,517
→ “The ‘biological’ lab in Wuhan where the COVID-19 virus was created was ‘funded’ by President Barak Hussein Obama in 2015 to the tune of \$3,800,000 American dollars.	FactCheck.org	Conspiracies	12,389
→ “There was no real scientific basis for believing that” social distancing would be necessary, “since it had never been studied.”	PolitiFact	Conspiracies	12,332
→ Riots in the American streets, after the collapse of the American economy due to the Covid-19 pandemic.	Misbar	Conspiracies	12,192

Joint work with First Draft:

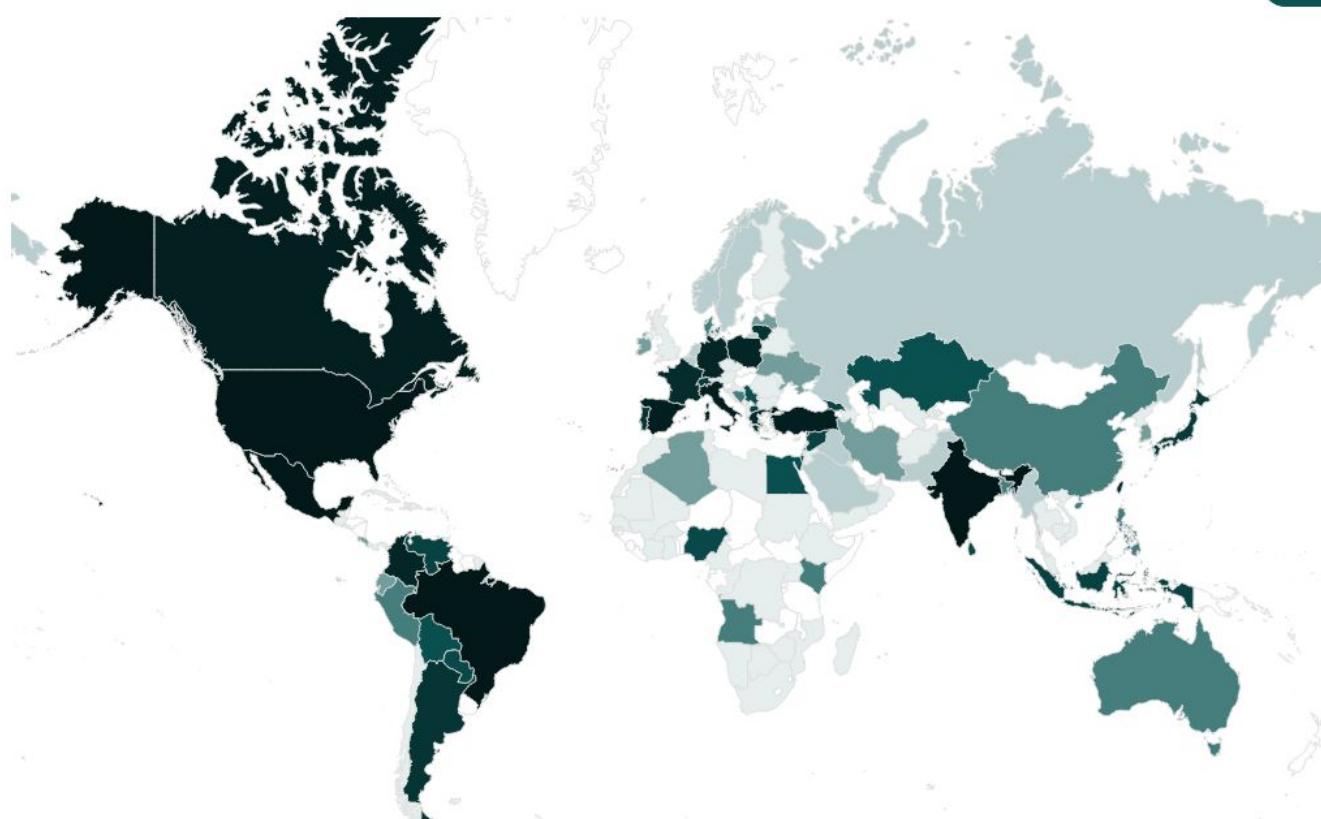
<https://datadeficits.firstdraftnews.org/>

<https://firstdraftnews.org/long-form-article/data-deficits/>



Specialised tools

- COVID-19 misinformation classifier



Data deficits Engagement deficits Search interest Total engagements Average engagements No. of fact checks

All topics
Community spread and impact
Conspiracies

No. of fact checks by Country





Summary

- We need to **listen to end users**
- **Fully automatic approaches** may not be ideal / desirable
- More research in **domain adaptation** is needed
- Disinformation is **multilingual!**





Challenges

- **Engaging with end users** is not easy
- **Identifying requirements** is also not trivial
- Sometimes **less is more** and we need to deal with it!
- Very difficult to provide **highly effective tools** for disinformation analysis without support from platforms





The
University
Of
Sheffield.

Thank you for your attention!



Iknor Singh



Yue Li



Kalina Bontcheva



Xingyi Song

Thanks to the WeVerify, SoBigData++ and Irish EDMO teams!!!
Special thanks to Denis Teyssou, Muneerah Patel, Ian Roberts, Mark Greenwood, Mehmet Bakir



www.weverify.eu
[@WeVerify](https://twitter.com/WeVerify)



Disclaimer: This presentation reflects the views only of the author(s), and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297



The
University
Of
Sheffield.

Thank you for your attention!



We are hiring!!!

Post-doc and PhD positions available to start from September!



c.scarton@sheffield.ac.uk

 **@carolscarton**



The WeVerify project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825297