



**WEVERIFY: ASSIST *WeVerify* OR ANALYSING FALSE
CONTENT, DISINFORMATION FLOWS, AND ONLINE INFLUENCE
CAMPAIGNS**

Kalina Bontcheva

University of Sheffield

K.Bontcheva@sheffield.ac.uk

@kbontcheva

How can digital technologies facilitate the detection and analysis of false information?



- **Automatically monitor multiple social platforms**
- **Detect fast-spreading content - candidate for verification**
- **Gather evidence about source trustworthiness**
- **Help verify images and videos**
- **Flag suspicious claims and textual content**
- **Help analyse the target recipients and disinformation spread**
- **Provide a shared repository of already debunked content**
- **Measure the impact of debunking strategies**



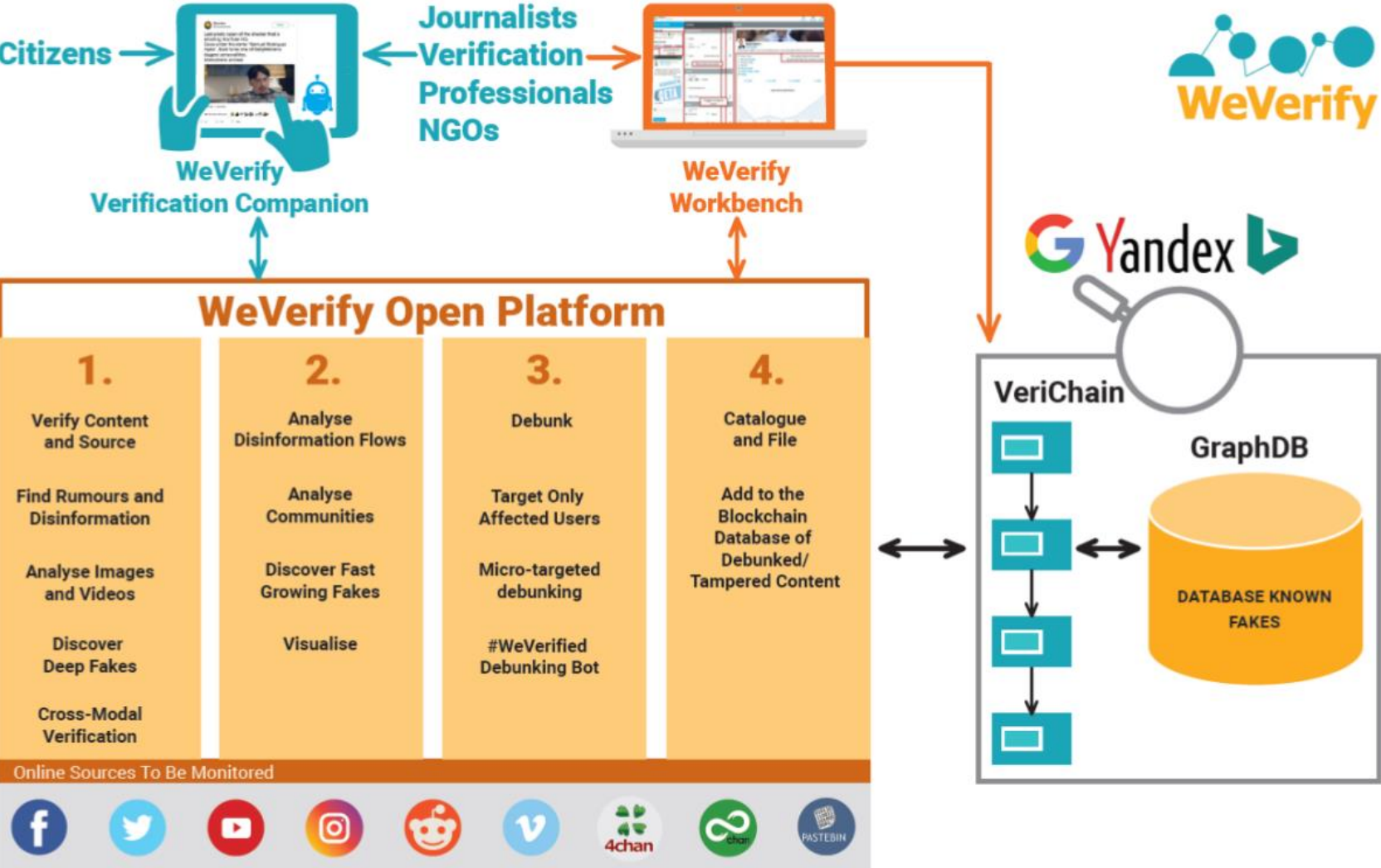
WeVerify: Forthcoming tools



- **Cross-modal and cross-platform approach**
 - Fringe + main platforms; Evidence from text, images, and videos; deep fakes
- **Overcome effort fragmentation**
 - A blockchain-based, authoritative database of already debunked fake items
- **Verification made simple**
 - A digital companion for citizens and an advanced version for journalists
- **Look beyond the content**
 - Analysis of disinformation source, spread strategies, and community networks
- **Build a community, not just technology**
 - Pilot with large stakeholder communities



WeVerify: Next Generation Verification Tools



@WeVerify

Chrome/Firefox Image & Video Verification Plugin



- **Verify Images and Videos**
- **Search and inspect manually past tweets around specific events**
- **Download: <https://weverify.eu/verification-plugin/>**
- **Open source & free-to-use**



Verifying Videos

InVID
IN VIDEO VERITAS

ANALYSIS KEYFRAMES THUMBNAILS TW. SEARCH MAGNIFIER METADATA VIDEO RIGHTS FORENSIC ABOUT SURVEY TUTORIAL

Video fragmentation and reverse image search

<https://www.youtube.com/watch?v=X7xq7Lq6d8Y>

Local file Submit

Video keyframes

With the contextual menu (right click) you can query multiple reverse image search engines or use the magnifier and forensic tools on each keyframe.

Target: Belgian MEP Guy Verhofstadt

Agent: Hungarian government

Feedback

<https://twitter.com/dteyssou/status/1052075694694559745>

Chrome/Firefox Verification Plugin: Demo videos

- **Analyse Facebook, Twitter, or YouTube Videos (demo):** <https://www.youtube.com/watch?v=eEVTYq1LMxM>
- **Analysing keyframes in videos (demo):** <https://youtu.be/iNs8y5DoxP8>
- **Analyse YouTube video thumbnails (demo):** <https://youtu.be/Vn0XGI0kSRQ>
- **Searching for past tweets and other advanced features (demo):** <https://youtu.be/wzMCtpENuHA>
- **Image Magnifier:** <https://youtu.be/JljRs8p4aK4>
- **Image Metadata Analyzer + Geolocation (demo):** <https://youtu.be/Mg6Z-Jc54yo>
- **Image and Video Forensics (demo):** <https://youtu.be/de862rAPne4>
- **Other demos and tutorials: See Tutorial tab in the plugin**

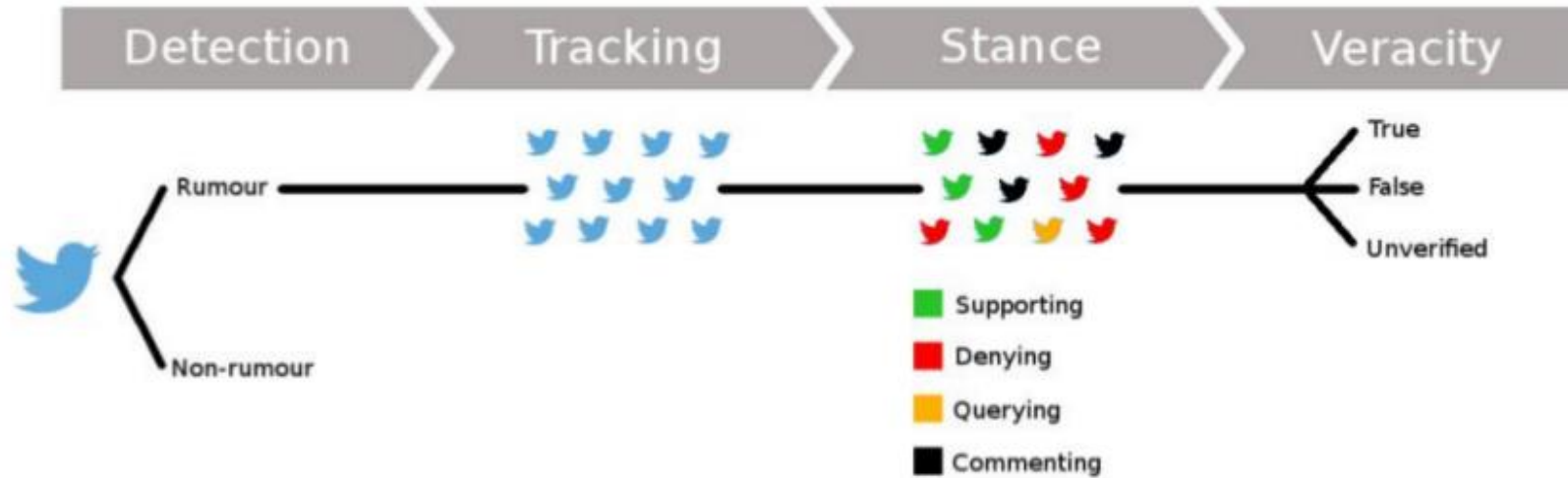
Verifying Rumours and Analysing Conversational Threads



- Rumour is “a circulating story of questionable veracity, which is apparently credible but hard to verify, and produces sufficient skepticism and/or anxiety”
- Key is not only whether it is true or false, but also:
 - Who is spreading it?
 - Are people believing it?
 - Rumour stance: support, deny, question, comment
 - Do we need to debunk it or not?
- Assist journalists through automated analysis embedded in a user-friendly tool
- The tool is undergoing testing, to be integrated within the Browser Plugin you saw earlier
- Demo link: <https://tweetveracity.gate.ac.uk/>
 - Try with: https://twitter.com/xoxo_mindy/status/1114002535059263488
 - Also with: <https://twitter.com/GargaC/status/1122045476889477120> to see the replies and their stance



PHEME: Analysing Rumours Automatically



Source: (PHEME project)

Verifying Rumours and Analysing Conversational Threads



GATE Tweet Verifier

[REST API](#) [About](#) [Login](#) [Register](#)

Twitter Rumour Verifier

https://twitter.com/xoxo_mindy/status/1114002535059263488

Verify!



Verifying Rumours: Demo Example 2

Results

Tweet

 **Garga Chatterjee** 
@GargaC

Believe it or not, this cartoon sitting on bonnet of jeep is Jagannath Sarkar, BJP MP candidate from Ranaghat. Its easy to laugh but thats not the point. This shows 1. How much alien BJP is to Bengalis incl Hindu Bengalis, 2. Bengali-born BJP ppl are Dhimmis of Hindi imperialism.



689 3:51 AM - Apr 27, 2019

389 people are talking about this

 **Veracity of tweet**

False Unverified True

[Provide feedback](#)

User profile



User	Garga Chatterjee (GargaC)
User verified	Verified
Location	Kolkata - Cambridge
Profile Description	Brain scientist. Columnist. Bangla Pokkho. Bengali Nationalist. Harvard PhD. Professor. Federalist. Sometimes I give back as good as I get.
Account Created	20/04/2012 06:37:37
Account Age (days)	2699 days
Followers	22067
Friends	563
Number of tweets	35344
Average tweet per day	13.095

Tweet Metadata

External Links

No external links

Media



Image - https://pbs.twimg.com/media/D5JOE_2U8AAQ62N.jpg

Veracity Responses

No responses

Verifying Rumours and Analysing Conversational Threads (2)



Replies

Sort from Show stance

ayesha ramjoo
@ayesharamjoo2

Replying to @GargaC and 7 others
what are we turning our Indian soil into Ram Leela maidaan ?

♥ 3 10:40 AM - Apr 28, 2019

[See ayesha ramjoo's other Tweets](#)

Question 0 0 0 0

SARANIK GHOSH
@Saranikg

Replying to @GargaC and 5 others
Here is another blue tick handle spreading fakery . The likes of @GargaC are blinded by Modi hate .

♥ 6:41 AM - Apr 28, 2019

[See SARANIK GHOSH's other Tweets](#)

Comment 0 0 0 0

Udayan Bose
@boseudayan

Replying to @GargaC and 6 others
Usual vile from fat and ugly Garga

♥ 4 5:45 AM - Apr 28, 2019 · Mumbai, India

[See Udayan Bose's other Tweets](#)



Analysing Memes and Adverts



- Above are adverts circulating on Facebook during the UK EU membership referendum
 - disclosed by Facebook to the DCMS Parliamentary Committee
- Demo link: <http://demos.gate.ac.uk/weverify/ocr/>
 - Try with the different demo memes provided on the demo page
 - Highly colourful text benefits from B&W or Grayscale conversion prior to processing
 - OCR tech usually trained on black text on white background

Analysing Memes and Adverts: Example Ad 1

Settings: Convert to black & white and Invert; No post-processing

Output

```
w . , ' TU»  
' , - '\ "h "V114" r .  
if v fix "fif  
"43,155." 3:5";  
fl , kt" "I; W , TURKEY, ALBANIA,  
"k * SERBIA, MACEDONIA,  
' ' 'r &MONTENEGRO  
'* "i * ARE JOINING THE EU  
V, . .  
;_ tn; " w j ' Iv u ' wmmishumhe UK?  
- " if K l' i
```

Source

Uploaded file: [brexit-eu-disinfo.png](#)

Original image:



Preprocessed image:

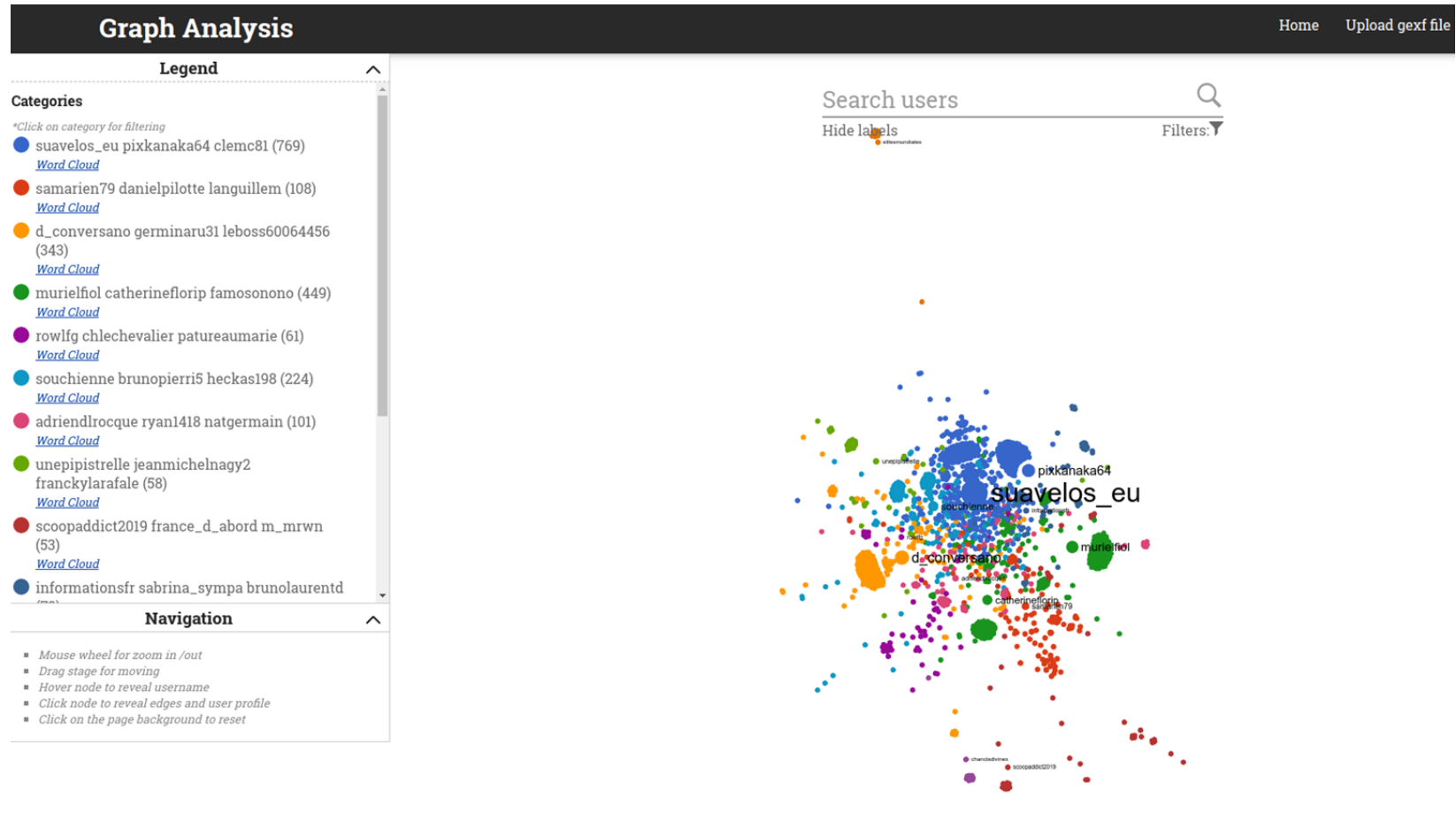


Suavelos: WeVerify SNA Visualisation Tool



Suavelos Study (Le Monde)

<https://www.disinfo.eu/2019/09/11/suavelos-white-supremacists-funded-through-facebook/>




Suavelos: Focusing on a Specific Community

Legend

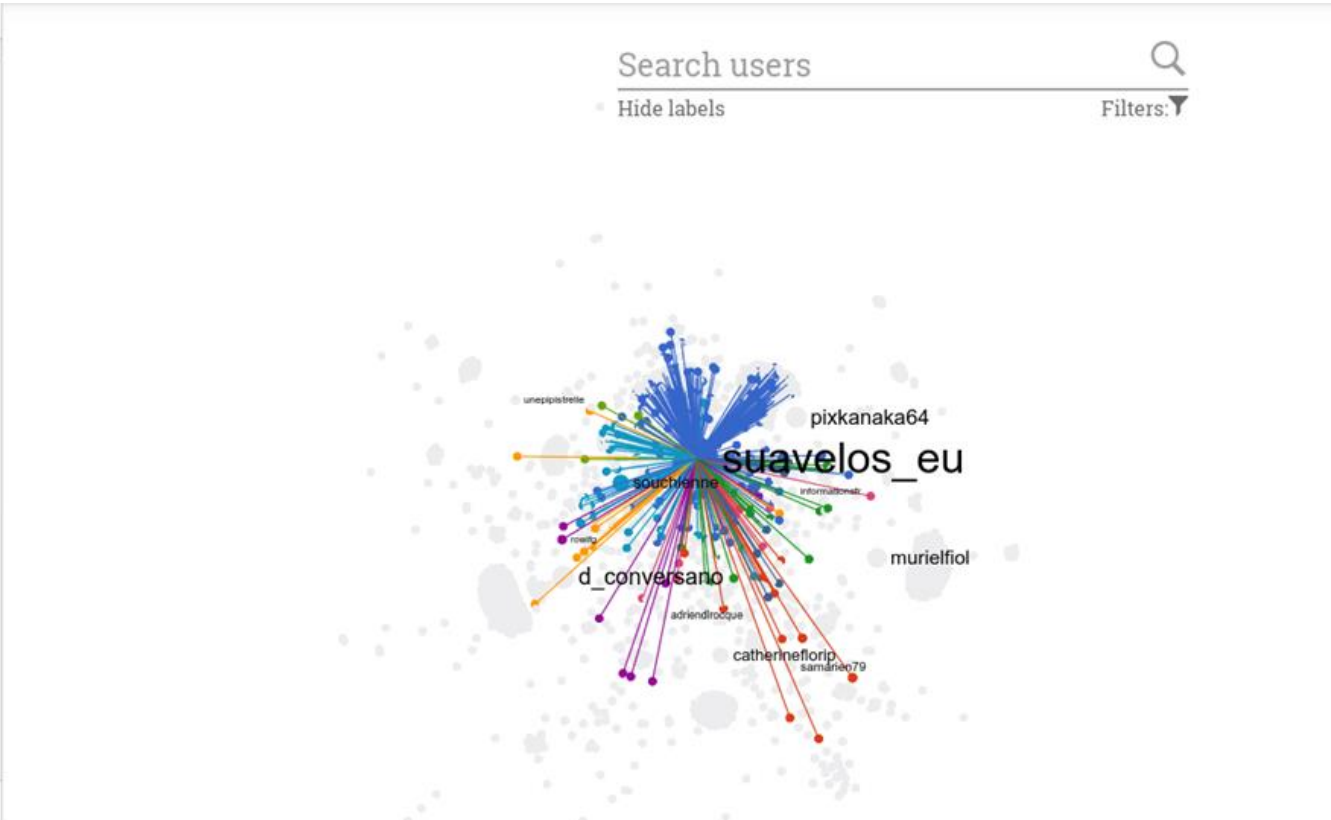
Categories

**Click on category for filtering*

- suavelos_eu [Word Cloud](#)
- samarien79 [Word Cloud](#)
- d_conversano (343) [Word Cloud](#)
- murielfiol [Word Cloud](#)
- rowlfg chle [Word Cloud](#)
- souchienn [Word Cloud](#)
- adriendrocque ryan1418 natgermain (101) [Word Cloud](#)
- unepipistrelle jeanmichelnagy2 franckylarafale (58) [Word Cloud](#)
- scoopaddict2019 france_d_abord m_mrwn (53) [Word Cloud](#)
- informationsfr sabrina_sympa brunolaurentd [Word Cloud](#)



Navigation




suavelos_eu
[@suavelos_eu](#) [Follow](#)

Tweets: 0 | Following: 1 | Followers: 2431

Mostly connected with:

- 746 (65%) | ● 126 (11%) | ● 76 (7%)

 pixkanaka64 Degree: 417	 souchienne Degree: 271
 samarieo79 Degree: 131	 danielpilotte Degree: 71
 clemc81 Degree: 60	 unepipistrelle Degree: 54

**Click on user to see activity*



Suavelos: Mapping Interactions over Time

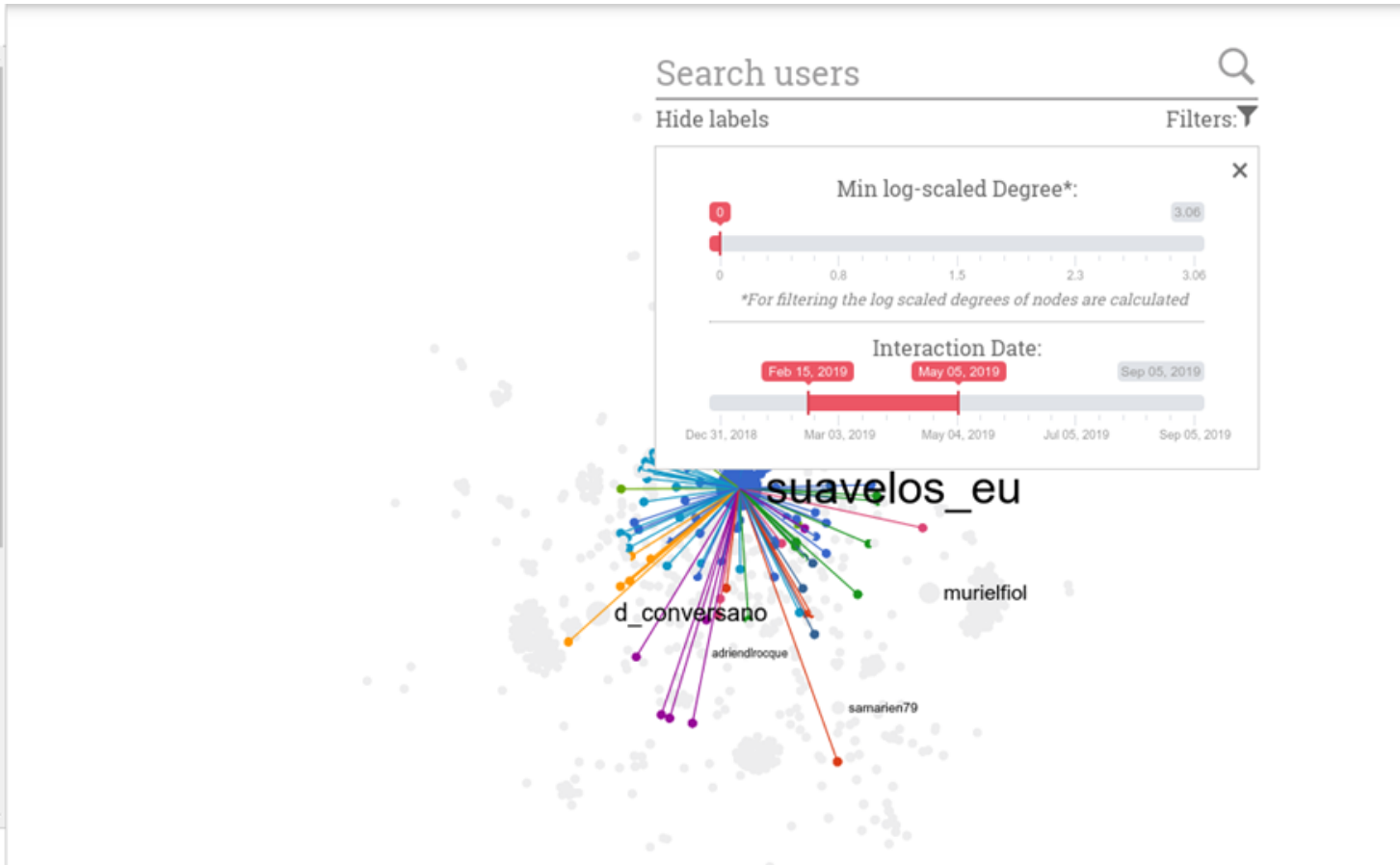
Legend

Categories

*Click on category for filtering

- suavelos_eu
[Word Cloud](#)
- samarien79
[Word Cloud](#)
- d_conversano (343)
[Word Cloud](#)
- murielfiol
[Word Cloud](#)
- rowlfg chle
[Word Cloud](#)
- souchienne
[Word Cloud](#)
- adriendrocque ryan1418 natgermain (101)
[Word Cloud](#)
- unepipistrelle jeanmichelnagy2 franckylarafale (58)
[Word Cloud](#)
- scoopaddict2019 france_d_abord m_mrwn (53)
[Word Cloud](#)
- informationsfr sabrina_sympa brunolaurentd
[Word Cloud](#)

Navigation









suavelos_eu
@suavelos_eu

Follow

Tweets 0 | Following 1 | Followers 2431

Mostly connected with:

- 746 (65%)
- 126 (11%)
- 76 (7%)

 pixkanaka64 Degree: 417	 souchienne Degree: 271
 samarien79 Degree: 131	 danielpilotte Degree: 71
 clemc81 Degree: 60	 unepipistrelle Degree: 54

*Click on user to see activity



TrulyMedia: Collaborative Verification Tool



Search collections

+ Add collection

- DEMO: Amazon ...**
updated a few seconds ago
6 · 9
- test**
test collection
updated 2 months ago
1 · 0
- Propagandalive**
updated 2 months ago
1 · 0

DEMO: Amazon fire

Type URL to add... Add

Any status ▾ Any tag ▾ Any source ▾ Any assignee ▾ ☆ Search in collection Date added ▾

Emmanuel Macron @EmmanuelMacron
3:15PM - 22 Aug 2019 **FAKE**

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



58503 ❤️ 162041

viral x Add tag

Checklist 10% complete

Ernesto Araújo @ernestofarajujo
8:32PM - 4 Sep 2019 **RAW**

First they fabricate a crisis. Then they try to sell its "solution", which is what they wanted in the first place. Fake crisis: "The Amazon is being burned". Fake solution: "Let's have international control over the Amazon." Thanks @camanpour for the opportunity to expose that.
<https://twitter.com/camanpour/status/1169302930710597632>

camanpour@camanpour
The world has watched fire engulf the Amazon with shock. But Brazil's Foreign Minister @ernestofarajujo insists "the Amazon is not burning, not burning at all. We had fires this year a little bit more than last year, but a little bit below average for the last 20 years." <https://t.co/3ynXS4qQrA>



1958 ❤️ 7295

Instagram
7:08AM - 5 Sep 2019 **FAKE**

Amazon Image




<https://www.instagram.com/p/B1bXxLnBmML/>
by Julia Bayer

viral x Add tag

Checklist 0% complete

Samuel TECH Official
12:56PM - 26 Aug 2019 **RAW**

Amazon Forest Fire 2019 | Please Pray For Amazon



79 👍 8 🗨️ 0

Checklist 7% complete

Pawan Durani @PawanDurani
10:40PM - 3 Sep 2019 **FAKE**

Highway in Brazil thru Amazon forest fire. <https://t.co/An7Rqjdg7a>

<https://www.facebook.com/?fbid=1819276798217898&s>
7:35AM - 4 Sep 2019 **FAKE**

FB Sven Bräsel
Brand

Amy Jackson @IamAmyJackson
3:35AM - 4 Sep 2019 **RAW**

Food For Thought
- 3,000 animals are killed per second for human greed and consumption
- The Amazon was on fire for an entire month, with one of the main factors being cattle farming
- The meat & dairy industry are to blame for leading causes of cancer, heart disease & obesity



244 ❤️ 1519

Checklist 7% complete

Feeds

TrulyMedia: Verify Content

[Back to Collection](#)

Verification status

Raw Pending Verified Fake

by Julia Bayer - 8:24AM - 5 Sep 2019

Publication information

Graphic License issue Profanity

Checklist

General Information

Keywords

Publication date: 22-08-2019, Publication time: 3:15PM, Time zone: UTC-0400

Event date: 22-08-2019, Event time: --:--:--, Time zone:

Is geolocation possible? Yes No

Location Landmarks/reference points

Latitude Longitude

Preservation link

Related Links [Add](#)

[Upload attachment](#)

Content

Reverse image search results are related to [Clear](#)

No results found

Same event

Other event

Verification Tools

Source Content Source History

Emmanuel Macron [@EmmanuelMacron](#)

Président de la République française. [Translate](#)

France

<http://www.elysee.fr>

Joined October 20, 2013

GMT-0400

Followers you follow

Tweets near France

Analyze user in TruthNest

Tweets: 8,378 Following: 708

Notes

Type your notes here...


[Switch item](#)

Multilingual Disinformation

inews 24 24 marzo alle ore 10:12 · €

ULTIM'ORA - Dottoressa e OSS aggredite in Pronto Soccorso - GUARDATE COSA FA QUESTO UOMO !!!!

Condividi per Denunciare



1,4 mln visualizzazioni

Mi piace Commenta Condividi

5,6 mila Commenti più in vista

26.440 condivisioni 3,1 mila commenti

EL PAÍS

La verdad detrás del bulo sobre la agresión de un musulmán en un centro médico

La historia, difundida en Facebook, fue protagonizada por un hombre ebrio y ocurrió en R

EL PAÍS · Madrid · 26 MAR 2017 · 16:53 CEST



Imagin del vídeo difundido por Facebook

La supuesta historia de un supuesto musulmán que agredió a un médico en un centro de salud español ha sobrepasado los tres millones de visitas en Facebook. Pero solo la mitad del relato es verdadero: Es cierto que hubo una agresión en un centro de salud pero el protagonista era un ruso en estado de embriaguez y ocurrió en la localidad rusa de Nóvgorod.

En las imágenes, que Juan Manuel González Sosa publicó en su cuenta de Facebook el pasado 25 de marzo, se aprecia cómo un hombre zarandea y golpea a una persona vestida con una bata blanca. Segundos después aparece una

WHAT? METHODOLOGY AND PRINCIPLES TEAM NOTICE CONTACT

teyit.org

Home page Política History Education Science Health Urban Legends

Syrians refugee allegedly beat a doctor from Russia shared video

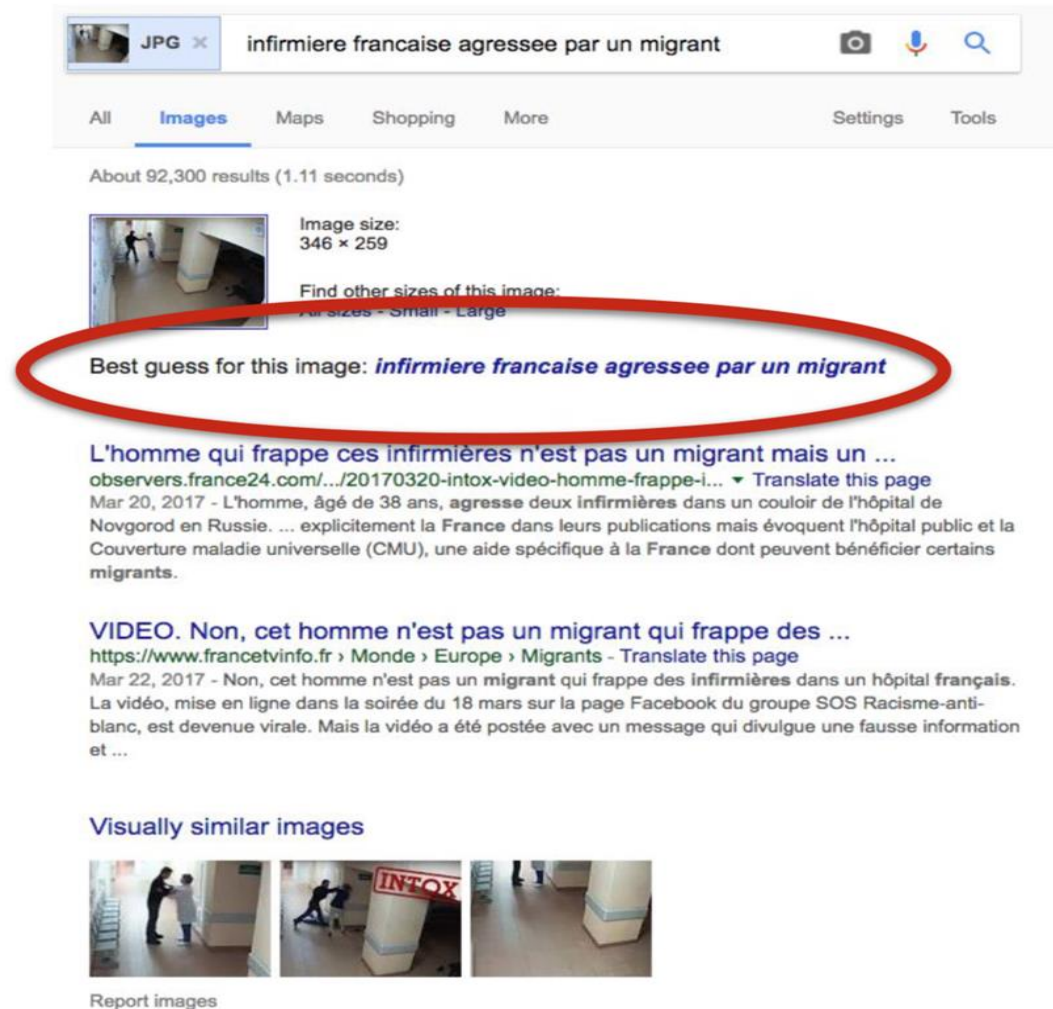
12/09/2017 · 12:39

CLAIM / Video: Syrian refugee beat female doctor in a hospital. **FALSE**

In a video shared on social media, a Syrian refugee women doctors in the hospital beating was alleged.



Google is getting confused



The screenshot shows a Google Images search interface. The search bar contains the text "infirmiere francaise agressee par un migrant". Below the search bar, the "Images" tab is selected. The search results show "About 92,300 results (1.11 seconds)". A red oval highlights the "Best guess for this image: infirmiere francaise agressee par un migrant" text. Below this, there are two search results. The first result is a link to "observers.france24.com" with the headline "L'homme qui frappe ces infirmières n'est pas un migrant mais un ..." and a sub-headline "Mar 20, 2017 - L'homme, âgé de 38 ans, agresse deux infirmières dans un couloir de l'hôpital de Novgorod en Russie. ... explicitement la France dans leurs publications mais évoquent l'hôpital public et la Couverture maladie universelle (CMU), une aide spécifique à la France dont peuvent bénéficier certains migrants." The second result is a link to "francetvinfo.fr" with the headline "VIDEO. Non, cet homme n'est pas un migrant qui frappe des ..." and a sub-headline "Mar 22, 2017 - Non, cet homme n'est pas un migrant qui frappe des infirmières dans un hôpital français. La vidéo, mise en ligne dans la soirée du 18 mars sur la page Facebook du groupe SOS Racisme-anti-blanc, est devenue virale. Mais la vidéo a été postée avec un message qui divulgue une fausse information et ...". Below the search results, there is a section titled "Visually similar images" with three small image thumbnails. The first thumbnail shows a person in a white coat (likely a nurse) being pushed or hit by another person in a hallway. The second thumbnail shows a person in a white coat being pushed or hit by another person in a hallway, with a sign that says "INTOX" visible in the background. The third thumbnail shows a person in a white coat being pushed or hit by another person in a hallway. Below the thumbnails, there is a "Report images" link.

JPG x infirmiere francaise agressee par un migrant

All **Images** Maps Shopping More Settings Tools

About 92,300 results (1.11 seconds)

Image size: 346 x 259

Find other sizes of this image: [Full size](#) - [Small](#) - [Large](#)

Best guess for this image: ***infirmiere francaise agressee par un migrant***

L'homme qui frappe ces infirmières n'est pas un migrant mais un ...
observers.france24.com/.../20170320-intox-video-homme-frappe-l... Translate this page
Mar 20, 2017 - L'homme, âgé de 38 ans, agresse deux infirmières dans un couloir de l'hôpital de Novgorod en Russie. ... explicitement la France dans leurs publications mais évoquent l'hôpital public et la Couverture maladie universelle (CMU), une aide spécifique à la France dont peuvent bénéficier certains migrants.

VIDEO. Non, cet homme n'est pas un migrant qui frappe des ...
<https://www.francetvinfo.fr> Monde Europe Migrants - Translate this page
Mar 22, 2017 - Non, cet homme n'est pas un migrant qui frappe des infirmières dans un hôpital français. La vidéo, mise en ligne dans la soirée du 18 mars sur la page Facebook du groupe SOS Racisme-anti-blanc, est devenue virale. Mais la vidéo a été postée avec un message qui divulgue une fausse information et ...

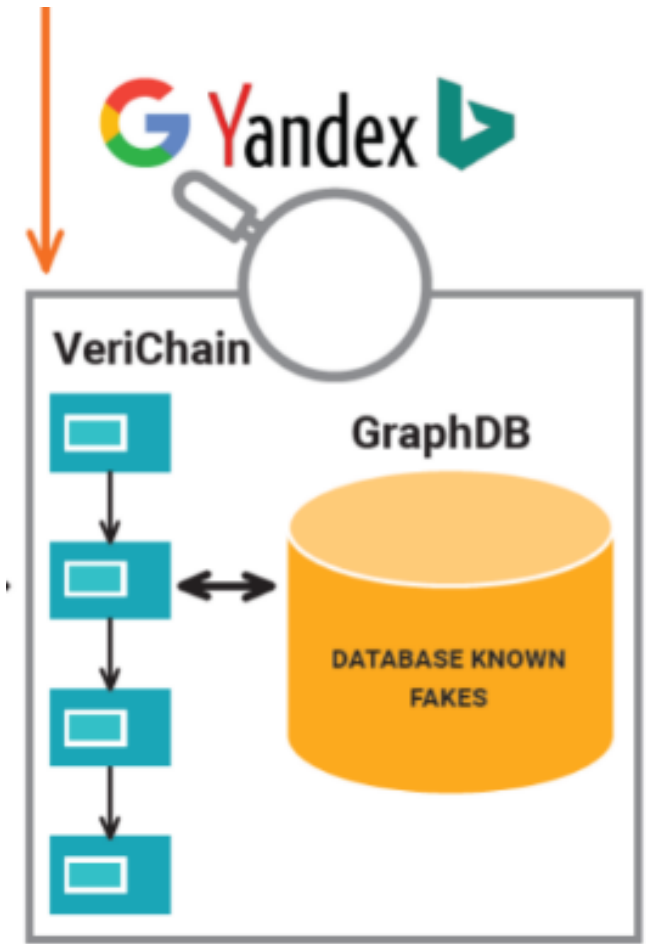
Visually similar images

Report images



Database of Known Fakes: Overview

- Store previous debunked content
- Useful for professionals, non-professionals and data scientists
- Quickly find known fake images, videos, and false narratives
- Main challenges include technology (scalability) and ethics (right to be forgotten)

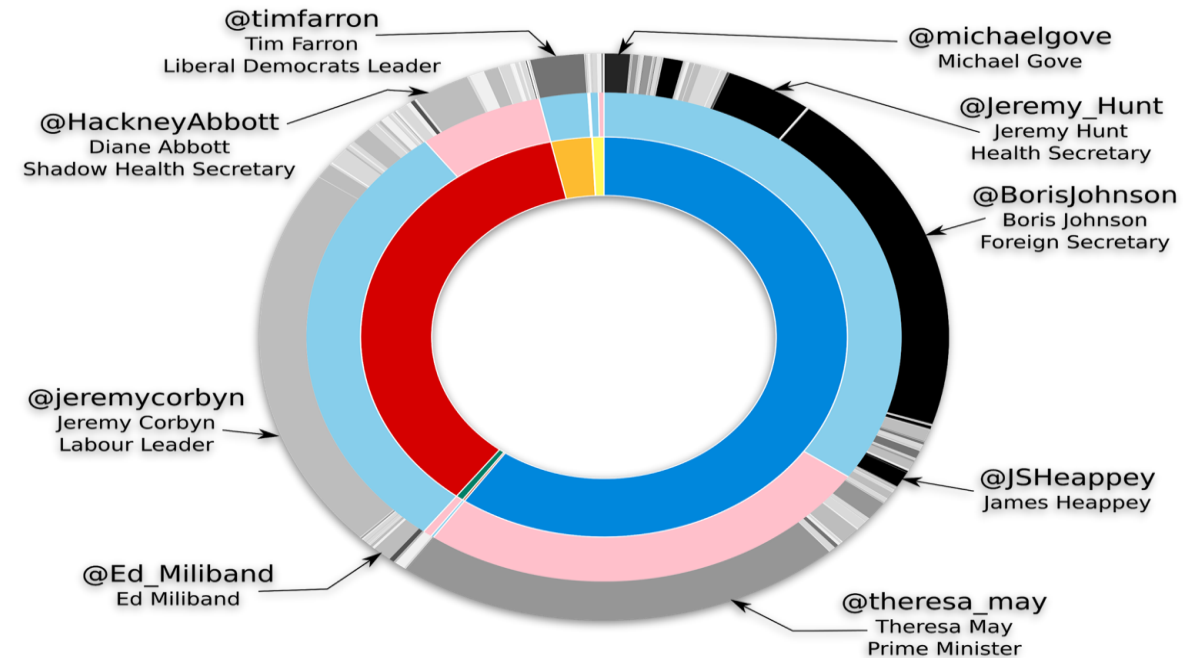


Elections: Online Abuse & Misinformation

Abuse of Politicians in UK elections 2015 & 2017 (& 2019)

- Abuse quantity doubled and now doubled yet again
- Abuse is spiky in nature

Abuse and misinformation are hard to research - ephemeral evidence on social platforms



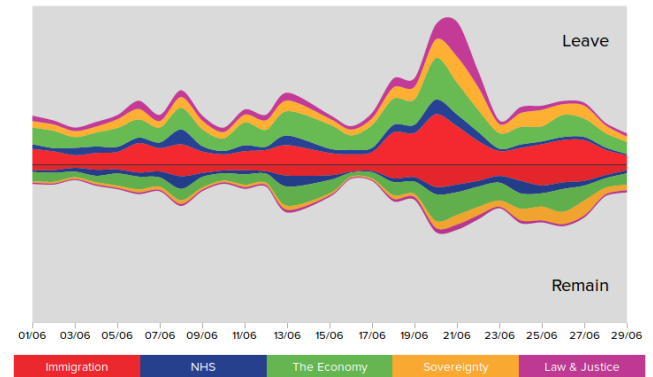
Misinformation in UK EU membership referendum 2016

Post-truth politics: "£350 million" figure widely shared, false

Propaganda and fake news: state-owned propaganda sites and fake news sites widely shared, mostly by leavers

Partisan media: role of the hyper-partisan sites in particular

Polarised crowds: studied the leave/remain camps



Can we trust the platforms to police themselves?

Twitter gave US Congress a set of around 2k Internet Research Agency linked accounts

We helped **BuzzFeed** journalists uncover **45 new bot accounts** that Twitter missed

Demonstrated that platforms need some independent oversight

Major Challenges Ahead

- **Researching elections and disinformation online is very hard**
 - **Data and evidence is continuously deleted**
- **Preserving important social media content for future studies**
 - **Regulate social platforms to be obligated to do it**
- **Establish policies for ethical, privacy-preserving research and data analytics**
- **Independent measurement of the effectiveness of technological solutions implemented by social media platforms**

A Final Thought

- **Freedom of expression and AI tools for disinformation**
 - **Focus on limiting the spread, not banning the false content as this introduces risks of restricting or censoring free speech**
 - **Making platforms responsible to remove disinformation could lead to collateral censorship - platforms over-censor to avoid liability**
 - **AI algorithms contain implicit bias and thus need oversight**



Project Contacts



- **Dr. Milena Yankova-Doseva**

Project Coordinator, Ontotext AD

email: milena.yankova@ontotext.com

- **Zlatina Marinova**

Administrative Coordinator, Ontotext AD

email: zlatina.marinova@ontotext.com

- **Prof. Kalina Bontcheva**

Scientific Director, University of Sheffield

email: k.bontcheva@sheffield.ac.uk

- **Nikos Sarris**

Exploitation Director, ATC

email: n.sarris@atc.gr

- **Denis Teyssou**

Innovation Manager, AFP

email: denis.teyssou@afp.com

- **Jochen Spangenberg**

Dissemination Lead, DW

email: jochen.spangenberg@dw.com

•



Thank you for your attention!

Further information in this STOA report:

Alaphilippe, A., Bontcheva, K., Gizikis, A. Automated tackling of disinformation: Major challenges ahead.

[http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU\(2019\)624278](http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2019)624278)

www.weverify.eu

<https://twitter.com/WeVerify>

