

# Fighting the Global Social Media Infodemic: from Fake News to Harmful Content

Preslav Nakov Mohamed bin Zayed University of Artificial Intelligence

> MARI 2023 January 25, 2023

# Is Fact-Checking the Solution to Disinformation?

# The Fact-Checking Landscape





# How Much Has Been Manually Fact-Checked?

### **Statistics**

Property	Global	Snopes I	Politifact	Africa Check	Truth or Fiction	Check your Fact	FactsCan	Fullfact	AFP Factcheck	AFP Factuel (FR)
Number of claims	33,261	12,812	16,476	520	1,311	823	125	250	678	266

ClaimReview schema: 77K

Twitter Birdwatch: 12K

https://data.gesis.org/claimskg/

## Technology could help:

- make fact-checkers more productive (~translation memories)
- automate the entire process (~full machine translation)

# What Do Fact-Checkers Want?

- 1. To know what is the most important claim to check today
- 2. To know when somebody says something that they have already checked
- 3. To check claims as quickly as possible

IJCAI-2021: Automated Fact-Checking for Assisting Human Fact-Checkers. Preslav Nakov, David Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed, Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, Giovanni Da San Martino

# The CheckThat! Lab at CLEF 2018-2023

https://sites.google.com/view/clef2022-checkthat



# What is Worth Fact-Checking?

# **Claim Detection in News Articles**



**EMNLP-2022:** NewsClaims: A New Benchmark for Claim Detection from News with Attribute Knowledge. Revanth Gangi Reddy, Sai Chetan Chinthakindi, Zhenhailong Wang, Yi Fung, Kathryn Conger, Ahmed Elsayed, Martha Palmer, Preslav Nakov, Eduard Hovy, Kevin Small, Heng Ji

#### Donald Trump: (07:37) How are you doing? How are you?

#### Kristen Welker: (07:58)

And I do want to say a very good evening to both of you. This debate will cover six major topics. At the beginning of each section, each candidate will have two minutes, uninterrupted, to answer my first question. The Debate Commission will then turn on their microphone only when it is their turn to answer. And the Commission will turn it off exactly when the two minutes have expired. After that, both microphones will remain on. But on behalf of the voters. I'm going to ask you to please speak one at a time.

#### Kristen Welker: (08-27)

The goal is for you to hear each other and for the American people to hear every word of what you both have to say. And so with that, if you're ready, let's start. And we will begin with the fight against the coronavirus. President Trump, the first question is for you. The country is heading into a dangerous new phase. More than 40,000 Americans are in the hospital tonight with COVID, including record numbers here in Tennessee. And since the two of you last shared a stage, 16,000 Americans have died from COVID. So please be specific. How would you lead the country during this next stage of the coronavirus crisis? Two minutes, uninterrupted

#### Kristen Welker: (09:03)

... during this next stage of the coronavirus crisis. Two minutes uninterrupted.

#### Donald Trump: (09:04)

So as you know, 2.2 million people modeled out, were expected to die. We closed up the greatest economy in the world in order to fight this horrible disease that came from China. It's a worldwide pandemic. It's all over the world. You see the spikes in Europe and many other places right now. If you notice, the mortality rate is down B5%. The excess mortality rate is way down and much lower than almost any other country. And we're fighting it and we're fighting it hard. There is a spike. There was a spike in Florida and it's now gone.

#### Donald Trump; (09:41)

There was a very big spike in Texas. It's now gone. There w spike in Arizona. It's now gone. And there was some spikes other places, they will soon be gone. We have a vaccine th ready. It's going to be announced within weeks. And it's ge delivered. We have Operation Warp Speed, which is the mil distribute the vaccine

#### Donald Trump: (10:04)

I can tell you from personal experience, I was in the hospit got better. And I will tell you that I had something that the therapeutic. I guess they would call it. Some people could cure, but I was in for a short period of time. And I got bette

wouldn't be here tonight. And now they say i'm immune. whether its four months or a lifetime, nobody's been able to say that, but I'm immune. More and more people are getting better. We have a problem that's a worldwide problem. This is a worldwide problem, but I've been congratulated by the heads of many countries on what we've been able to do. If you take a look at what we've done in terms of goggles and masks and gowns and everything else, and in particular ventilators we're now making ventilators. all over the world, thousands and thousands a month distributing them all over the world, it will go away. And as I say, we're rounding the turn. We're rounding the corner. It's going away.

#### Kristen Welker: (11-04)

Okay. Former Vice President Biden to you. How would you lead the countryout of this crisis? You have two minutes uninterrupted.

#### Joe Biden: (11:13)

220.000 Americans dead. You hear nothing else I say tonight, hear this. Anyone who is responsible for not taking control. In fact, not saying I take no responsibility initially. Anyone is responsible for that many deaths should not remain as president of the United States of America. We're in a situation where there are a thousand deaths a day now. A thousand deaths a day. And there are over 70,000 new cases per day. Compared to what's going on in Europe as the New England Medical Journal said. they're starting from a very low rate. We're starting from a very high rate.

#### Joe Biden: [11-SA]

The expectation is we'll have another 200,000 Americans dead between now and the end of the year. If we just wore these masks, the president's own advisors have told him, we can save a 100,000 lives. And we're in a circumstance where the president thus far and still has no plan, no comprehensive plan

#### Joe Biden: (12:19)

What I would do is make sure we have everyone encouraged to wear a mask all the time. I would make sure we move into the direction of rapid testing, investing in rapid testing. I would make sure that we set up national standards as to how to open up schools and open up businesses so they can be safe and give them the wherewithal, the financial resources to be able to do that

#### Joe Biden: (12:41)

We're in a situation now where the New England Medical Journal, one of the serious, most serious journals in the whole world said for the first time. ever that the way this president has responded to this crisis has been absolutely tradic. And so folks, I will take care of this, I will end this, I will make sure we have a plan.

#### Kristen Welker: (13:01)

President Trump, I'd like to follow up with you and your comments. You talked about taking a therapeutic, I assume you're referencing Regeneron. You also said a vaccine will be coming within weeks. Is that a guarantee?

#### Donald Trump: (13:13)

No. it's not a guarantee, but it will be by the end of the year. But I think it has a good chance. There are two companies, I think within a matter of weeks and it will be distributed very guickly.

#### Kristen Welker: (13/22)

Can you tell us which companies?

#### Donald Trump: (13:23)

Johnson & Johnson is doing very well. Moderna is doing very well. Pfizer is doing very well and we have numerous others. Then we also have others that we're working on very closely with other countries, in particular Europe

#### Kristen Welker: (13:35)

#### Donald Trump: (15:24)

And frankly, he can the H1N1 swine flu and it was a total disaster. Far less lethal, but it was a total disaster. Had that had this kind of numbers. 700.000 people would be dead right now, but it was a far less lethal disease. Look, his own person who ran that for him, who, as you know, was his chief of staff said, "It was catastrophic. It was horrible. We didn't know what we were doing." Now he comes up and he tells us how to do this.

#### Donald Trump: (10:01)

Also, everything that he said about the way every single move that he said we should make, that's what we've done. We've done all of it, But he was way behind us

#### Kristen Welker: (16:10)

Vice President Biden, your response

#### Joe Biden: (16:12)

My response is he is xenophobic, but not because he shutdown access from China. And he did it late after 40 countries had already done that. In addition to that, what he did, he made sure that we had 44 people that were in there in China trying to get the Wuhan to determine what exactly the source was. What did the president say in January? He said, " No." He said. "He's being transparent. The president of China is being transparent. We owe him a debt of gratitude. We have to thank him."

#### Joe Biden: (16:38)

And then what happened was we started talking about using the Defense Act to make sure we go out and get whatever is needed out there to protect people. And again, I go back to this. He had nothing, he did virtually nothing. And then he gets out of the hospital and he talks about. "Oh, don't worry. It's all going to be over soon." Come on. There's not another serious scientist in the world who thinks it's going to be over soon.

#### Kristen Welker: (17:03)

President Trump, your reaction?

#### Donald Trump: (18:56)

Excuse me. I take full responsibility. It's not my fault that it came here. It's China's fault. And you know what? It's not Joe's fault that it came here either. It's China's fault. They kept it from going into the rest of China for the most part, but they didn't keep it from coming out to the world, including Europe and ourselves.

#### Kristen Welker: (19:16)

Vice President Biden.

#### Joe Biden: (19:17)

The fact is that when we knew it was coming, when it hit, what happened? What did the President say? He said. "Don't worry. It's going to go away. It'll be gone by Easter. Don't worry, Warm weather. Don't worry. Maybe inject bleach" He said he was kidding when he said that, but a lot of people thought it was serious. A whole range of things the President has said, even today, he thinks we are in control. We're about to lose 200,000 more people.

#### Kristen Welker: (19:43) President Trump.

#### Donald Trumo: (19:44)

Look, perhaps just to finish this, I was kidding on that, but just to finish this, when I closed he said I shouldn't have closed. And that went on for months. Nancy Pelosi said the same thing. She was dancing on the streets in Chinatown, in San Francisco, But when I closed, he said. "This is a terrible thing, you xenophobic." I think he called me racist even, because I was closing it to China. Now he says I should have closed it earlier. Joe, it doesn't

#### Joe Biden: (20:13)

didn't say either of those things.

Donald Trump: (20:14) You certainly did. You certainly did.

#### Kristen Welker: (22:04)

Let's talk about schools. President Trump-Donald Trump: (22:06)

I think we have to respond, if I might Kristen Welker: (22:07)

#### Please And then I have a follow-up

Donald Trump: (22:09)

Thank you, and I appreciate that, Look, all he does is talk about shut downs. But forget about him. His Democrat Governors. Cuomo in New York, you look at what's going on in California, you look at Pennsylvania. North Carolina, Democrats, Democrats all, they're shut down so tight and they're dying. They're dying. And he supports all these people. All he talks about is shut downs. No, we're not going to shut down. And we have to open our schools. And it's like, as an example, I have a young son, he also tested positive. By the time I spoke to the doctor the second time, he was fine. It just went away. Young people, I guess, it's their immune system.

#### Kristen Welker (22:47)

Let me follow up with you, President Trump, You've demanded schools open in person and insist they can do it safely. But just vesterday, Boston became the latest city to move its public school system entirely online after a coronavirus soike. What is your message to parents who worry that sending their children to school will endanger not only their kids but also their teachers and family/7

#### Donald Trump: (23:05)

want to open the schools. The transmittal rate to the teachers is very small. But I want to open the schools. We have to open our country. We're not going to have a country. You can't do this. We can't keep this country closed. This is a massive country with a massive economy. People are osing their jobs. They're committing suicide. There's depression, alcohol, drugs at a level that nobody's ever seen before. There's abuse, tremendous abuse. We have to open our country. I've said it often, the oure cannot be ening. And he wants

pre to trace, you need to be able to provide all the resources that are

to make sure that we're going to open safely. And by the way, all you

about it. So don't worry about it. Come on.

President Trump, let me follow up with you quickly.

vibrant, it's dying, Everyone's leaving New York.

Kristen Welker: (24:38)

Donald Trump: (24:40)

Kristen Welker; (25:15)

Vice President Biden.

Joe Biden: (25:15)

teachers out there, not that many of you are going to die, so don't worry

By the way, I will say this, If you go and look at what's happened to New

plexiglass, these are restaurants that had dying. These are businesses

not the answer. I mean, you're going to sit there in a cubicle wrapped

that to people. You just can't. Take a look at New York and what's

happened to my wonderful city. For so many years, I loved it, it was

with no money. Putting up plexiglass is unbelievably expensive, and it's

around with plastic. These are businesses that are dying. Joe. You can't dr.

Take a look at what New York has done in terms of turning the curve down.

n terms of the number of people dving. And I don't look at this in terms of

They're the red states, they're the states in the Midwest, they're the states

in the upper Midwest. That's where the spike is occurring significantly. But

the way he does, blue states and red states. They're all the United States.

And look at the states that are having such a spike in the coronavirus.

they're all Americans. They're all Americans. And what we have to do is

say, wear these masks, number one. Make sure we get the help that the

York, it's a ghost town, it's a ghost town. And when you talk about

needed to do this. And that is not inconsistent with saying that we're going

rson in our massive

v gum at the same

open a business.

cople cannot infect

can take testing

ed. You need to be

ed resources to open"

to have, if you have a

https://www.rev.

com/blog/transc

ripts/donald-

trump-joe-

biden-final-

debate-

presidential-

transcript-2020

9

# The debate has 1000+ sentences! What is worth fact-checking?

would definitely agree to take a compavinus vancine if it was approved by the government. What steps would you take to give Americans confidence in a vaccine if it were approved?

#### Joe Biden: [14:26]

Make sure it's totally transparent. Have the scientists of the world see it. know it, look at it, go through all the processes. And by the way, this is the same fellow who told you, "This is going to end by Easter" last time. This is the same fellow who told you that. "Don't worry, we're going to end this by the summer." We're about to go into a dark winter, a dark winter and he has no clear plan. And there's no prospect that there's going to be a vaccine available for the majority of the American people before the middle of next

#### Kristen Welker: (14:54)

President Trump, your reaction. He says, you have no plan.

#### Donald Trump: (14:56)

I don't think we're going to have a dark winter at all. We're opening up our country. We've learned and studied and understand the disease, which we didn't at the beginning. When I closed and banned China from coming in heavily infected and then ultimately Europe, but China was in January. Months later, he was saying I was xenophobic. I did it too soon. Now he's saying, "Oh, I should have moved guicker," but he didn't move guicker. He was months behind me, many months behind me.

#### Donald Trump: (15:24)

And frankly, he ran the H1N1 swine flu and it was a total disaster. Far less lating) but it use a total disactor Light that had this kind of numbers 700.000 people would be dead right now, but it was a far less lethal disease. Look, his own person who ran that for him, who, as you know, was and takin old address and it with other tests for the the balance ad

rameu a loc méameu a loc éreas doctors, greas hospitais, And now recovered. 99.9 of young people recover. 99% of people recover. We have to recover-

#### Donald Trump: (18:03)

99% of people recover. We have to recover, We can't close up our nation. We have to open our school and we can't close up our nation, or you're not going to have a nation

#### Kristen Welker: (18:11)

And, of course, the CDC has said young people can get sick with COVID-19 and can pass it. Vice President Biden. I want to talk broadly about strategy though

#### Joe Biden: (18:20)

Can I respond to that?

#### Kristen Welker: (18:21)

30 seconds, please. Then I have a question.

#### Joe Biden: (18:23)

Number one, he says that we're learning to live with it. People are learning to die with it. You folks home will have an empty chair at the kitchen table this morning. That man or wife going to bed tonight and reaching over to try to touch, there out of habit, where their wife or husband was, is gone. Learning to live with it. Come on: We're dying with it, because he's never said. See, you said, "It's dangerous." When's the last time? Is it really dangerous still? Are we dangerous. You tell the people it's dangerous now. What should they do about the danger? And you say, 'I take no responsibility

#### Kristen Welker: (18:56)

Latima talk should your Tisseriou

Disau Let's talk about your different stratenies toward dealing with this. Mr Vice President you suggested you would support new shutdowns if scientists recommended it. What do you say to Americans who are fearful that the cost of shutdowns, the impact on the economy, the higher rates of hunger, depression, domestic, and substance abuse outweighs the risk

#### of exposure to the virus?

#### Joe Biden: (20.54)

What I would say is, I'm going to shut down the virus, not the country. It's his ineptitude that caused the country to have to shut down in large part, why businesses have gone under, why schools are closed, why so many people have lost their living, and why they're concerned. Those other concerns are real. That's why he should have been, instead of in a sand trap at his golf course, he should have been negotiating with Nancy Pelos and the rest of the Democrats and Republicans about what to do about the acts they were passing for billions of dollars to make sure people had the capacity.

Oh, well, no. I'm not shutting down today, but there are ... Look, you need

standards. The standard is, if you have a reproduction rate in a community

Sistancing. Do not open bars and do not open gymnasiums. Do not open

until you get this under control, under more control." But when you do

capacity to do it safely. For example, schools, Schools, they need a lot of

money to be open. They need to deal with ventilation systems. They need

to dasi with smaller riscoss more teachers more and a confused

open, give the people the capacity to be able to open and have the

that's above a certain level, everybody says, "Slow up. More social

#### Kristen Welker: (21:23) But you haven't ruled out more shut downs?

Joe Biden: (21-25)

# We Can Learn From Human Fact-Checks

been told through investigative reporting that he owes about \$650 million to Wall Street and foreign banks. Or maybe he doesn't want the American people, all of you watching tonight, to know that he's paid nothing in federal taxes, because the politifact poly's ever seen were a couple of years when he had to the politifact of years when he was trying to get a casino license, and they showed he didn't pay any federal income tax.

Sep 27, 2016

We rated this claim Mostly False. <u>http://www.politifact.com/truth-</u> <u>o-meter/statements/2016/sep</u> /27/hillary-clinton/debate-hillaryclinton-tax-claim-ignores-years-whe/

TRUMP: That makes me smart.

CLINTON: So if he's paid zero, that means zero for troops, zero for vets, zero

# **Detecting Check-Worthy Claims**

This tool identifies automatically w	hich claims in a given document a	re most worthy and shou	ld be prioritized	for fact-checking, e.g., by investigative journalists.		
,		,,	1			
0.0 (least	check-worthy)			(most check worthy) 1.0		
-						
<sup>0.46</sup> : We have an awful lot of vacant houses i	n this city.					
0.05: They're in good shape and structurally s	sound.					
0.55: In the last three years we've taken 3,000	) vacant houses, taken them away from	m the homeowners who neg	plected them, sold	them on the web site BuildingDetroit.org, and moved families		
back in, Three thousand families occupie	d.					
0.15: Senator Young says he wants to sell sid	e lots.					
0.44: He's a little late to the party.						
0.40: We've sold 8,200 side lots to next -door	neighbors who are maintaining them s	so they could add swing sets	s and slides and g	jardens.		
0.20: And then what we are doing is taking a s	strategy that works so well in Midtown,	and we call it the Strategic I	Neighborhood Fu	nd.		
0.59: We are starting to go to 10 neighborhoo	ds across the city and investing all kin	ds of money.				
0.52: We've lined up now \$60 million of private	e money and this week in West Village	we just opened the Coe.				
0.46 with the autiful new apartments and	retail.		and the second second second			
0.45. We're doing the same thing in the Fitzgerald neighborhood at McNichols and Livernois, where we're renovating 115 houses and then out in Old Redford.						
0.43: We're about to kick off a program: we're going to renovate 50 houses around the old Holcomb school.						
0.02 we re going from neighborhood to neighborhood and then for the storefronts.						
0.49. And I'm user need to create beautiful areas where businesses want to invest.						
0.46: We're geing te put \$90 million inte it						
0.37: Twenty three neighborhood quarters your shopping Court is making the heautiful flowers and walkways and the like and recreate those old chooping districts						
0.44. That's the plan we have for the next four vears						
That's the plan we have for the nextfour years.						
	Enter new text Min	nic source: General	~ S	Sort		

**NAACL'2018 (demo):** ClaimRank: Detecting Check-Worthy Claims in Arabic and English. sraa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Màrquez, Preslav Nakov

11

# **Check-Worthiness**

V



Since this will never get reported by the media, I wanted to share a copy of this check. @realDonaldTrump is once again donating his salary back to the United States Government — This quarter, it will be donated to @HHSGov to confront, contain, and combat #Coronavirus.

DONALD J. TRUMP c/o The Mar-a-Lago Club APITAL ONE, N.A. 2 WEST S7TH STREET 50-791/214 1100 S Ocean Blvd 003573 Palm Beach, FL 33480 CHECK DATE CHECK AMOUNT 01/29/20 \*\*\*\*\$100.000.00\*\* PAY OFFICE OF THE ASSISTANT SECRETARY FOR TO THE ORDER OF HEALTH

D. Trump is donating his salary to combat COVID19?

# This is check-worthy!

Donald J. Trump

11:11 PM · Mar 3, 2020 from Manhattan, NY · Twitter for iPhone

**ECIR-2020:** CheckThat! at CLEF 2020: Enabling the Automatic Identification and Verification of Claims in Social Media. Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari

# What Makes a Tweet Check-Worthy?

- Q0: viral?
- Q1. contains verifiable factual daim?
- Q2. likely to be false?
- Q3. of public interest?
- Q4: harmful?
- Q5. worth fact-checking?

	One-task	Multi-task
Q2	69.2	72.9
Q3	82.5	71.6
Q4	56.0	67.9
Q5	62.0	76.8

**EMNLP'2021 (Findings):** Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, Abdulaziz Al-Homaid, Wajdi Zaghouani, Tommaso Caselli, Gijs Danoe, Friso Stolk, Britt Bruntink, Preslav Nakov.

## Detecting Check-Worthy Tweets

**Tweet Classifier** 

Masks don't stop transmission. Vaccines don't stop transmission. Don't allow anyone to coerce you into thinking these

**Tweet Analytics** 

liculous measures work. Stop the allowance of mandates for ludicrous reasons.	Classify Tweet
Summarv	
Does the tweet contain a verifiable factual claim?	
Yes	
Does the tweet appear to contain false information?	
No, probably contains no false info	
Will the tweet have an effect on or be of interest to the general public?	
Yes, probably of interest	
Is the tweet harmful to the society/person(s)/company(s)/product(s)?	
Yes, probably harmful	
Do you think that a professional fact-checker should verify the claim in the tweet?	

# Detecting Previously Fact-Checked Claims

# **Automatic Fact-Checking**

- **Problem:** people do not trust automatic methods
  - Yet, they trust human fact-checkers
- **Proposal:** Look for previously fact-checked claims
  - people tend to repeat the same claims!









## **Detecting All Previously Fact-Checked Claims in a Document**



**EMNLP'2022:** Assisting the Human Fact-Checkers: Detecting All Previously Fact-Checked Claims in a Document. Shaden Shaar, Nikola Georgiev, Firoj Alam, Giovanni Da San Martino, Aisha Mohamed and Preslav Nakov

# **Detecting Previously Fact-Checked Claims**



# Not an Easy Task: Need for Contextual Understanding!



**NAACL'2022 (Findings):** *The Role of Context in Detecting Previously Fact-Checked Claims.* Shaden Shaar, Firoj Alam, Giovanni Da San Martino, Preslav Nakov

# Is Automatic **Fact-Checking** the Solution to **Disinformation?**

# Fact-Checking Using Tables



**EMNLP'2022:** *PASTA: Table-Operations Aware Fact Verification via Sentence-Table Cloze Pre-training.* Zihui Gu, Ju Fan, Nan Tang, Preslav Nakov, Xiaoman Zhao and Xiaoyong Du

## WhatTheWikiFact: Fact-Checking Claims Against Wikipedia

#### You are now in the debug mode!

This system is a fact-cheking pipeline demonstration system and is intended to demonstrate the results of testing some "toy" statements. It is not a system for checking real (political) claims.

This tool performs fact checking by comparing information with Wikipedia data.

Select the language		
English -		
Enter claims here (one per row, max 5)		
Abu Dhabi and Dubai are in the same country.		
Output + options +		
Verify		
Claim	Verdict	Confidence
Abu Dhabi and Dubai are in the same country.		

**CIKM-2021:** *WhatTheWikiFact: Fact-Checking Claims Against Wikipedia.* Anton Chernyavskiy, Dmitry Ilvovsky, Preslav Nakov

https://www.tanbih.org/whatthewikifact

# Heterogeneous Graph for Automatic Fact-Checking



**CIKM-2020 (best paper award):** FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, Min-Yen Kan

# FbMultiLingMisinfo: Large-Scale Multilingual Benchmark for Misinformation Detection

PolitiFact

- from Facebook Privacy-Protected Full URLs Data Set (all URLs shared on ٠ Facebook at least 100 times January'2017 –July'2019
- ٠ fact-checking label
- ٠
- Multilingual, larger, real-world, much more challeging

2500

2000

1500

1000

500

count



IJCNN'2022: FbMultiLingMisinfo: Challenging Large-Scale Multilingual Benchmark for Misinformation Detection. Giorgio Barnabo, Federico Siciliano, Carlos Castillo, Stefano Leonardi, Preslav Nakov, Giovanni Da San Martino, Fabrizio Silvestri

# Crowd Fact-Checking



**Verifying** Article



**Does Ivermectin Cause Sterility in Men?** 

One study purportedly found that 85% of men who were given the antiparasitic were sterile following the research period.

By Madison Dapcevich Published 8 September 2021, Updated 10 September 2021

0 🕤 🖂 G 🥣



Context

The study in question was not published in a credible journal, nor was it hosted by an accredited, reputable institution. In the decade since the study's supposed 2011 publication, there has been little - if any - related research to confirm its findings. Furthermore, a spokesperson for the U.S. Food and Drug Administration told Snopes that infertility in men is not a known side effect of ivermectin and, as such, is not included in U.S. labeling requirements.

AACL-2022: CrowdChecked: Detecting Previously Fact-Checked Claims in Social Media. Momchil Hardalov, Anton Chernyavskiy, Ivan Koychev, Dmitry Ilvovsky, Preslav Nakov



Can we fact-check every claim in the world?

# This is Very Hard !



# of the spread of "fake news" on Twitter happens in less than 10 minutes

Zaman, T., Fox, E. B., & Bradlow, E. T. (2014). A bayesian approach for predicting the popularity of tweets. *The Annals of Applied Statistics*, *8*(3), 1583-1611.

# faster is how "fake news" travel in Twitter, compared to the truth

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146-1151.

## Need for Early Fake News Detection: **Focus on the Source!**

# So, Better to Go After the Source

#### https://here.comes.url.com



# Thus, We Can Detect "Fake News" Before It Was Even Written!





**ACL'2020:** What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, Preslav Nakov

# Media Profiling Based on Audience Overlap in AlexaRank



**EMNLP'2022:** *GREENER: Graph Neural Networks for News Media Profiling.* Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel and Preslav Nakov



# Multitask Ordinal Regression for Factuality & Bias

- Ordinal regression
- Learn jointly
  - Factuality
  - Left-center-right bias
    - 7-point
    - 5-point
    - 3-point
- Centrality
  - Hyper-partisanship



**NAACL-HLT 2019:** *Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media* Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James R. Glass, Preslav Nakov

# Propaganda

**Propaganda** is a communication tool that is **deliberately** designed to **influence** the opinions and the actions of other people in order to achieve a predetermined goal.



## Propaganda Techniques

Repetition Black-and-white Fallacy/Dictatorship **Obfuscation, Intentional vagueness, Confusion Thought-terminating cliché** Emotional **Causal Oversimplification** Flag-waving Slogans appeals Appeal to authority Transfer **Presenting Irrelevant Data (Red Herring)** Smears **Exaggeration/Minimisation** Whataboutism Logical Misrepresentation of Someone's Position (Straw Man) fallacies Name calling/Labeling Appeal to fear/prejudice Reductio ad hitlerum Bandwagon Glittering generalities (Virtue) Loaded Language Douht Appeal to (Strong) Emotions


**EMNLP-2019:** *Fine-Grained Analysis of Propaganda in News Article.* Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, Preslav Nakov (Natural News) The Covid crimes are years in the making and more than a handful of vaccine manufacturers and promoters could be on trial soon for mass-murder<sup>9</sup> and conspiracy to commit<sup>9</sup> genocide.

- Senator Rand Paul puts Dr.

Anthony Fauci at the forefront of the whole plandemic brigade, since Fauci lied and DID help fund gain-of-function research that led directly to the (Wuhan Flu) virus outbreak in humans across planet Earth.

Already those who planned and propagated the scamdemic are begging for forgiveness, knowing that Nuremberg II trials are coming, sooner or later, for them.

There are too many people involved in this mass murder scheme, but the ones at the top should be tried first.

- Anthony "Fraudulent" Fauci is the King of Misinformation<sup>10</sup> when it<sup>6</sup> comes to anything related to Covid-19

- Fauci flip-flopped<sup>9</sup> on mask wearing, social distancing, lockdowns, medications, vaccinations, and just about every piece of "advice" that spewed from his mouth for the past two years.

Back and forth he went, talking about mandates that were vital to 'flattening the curve' and saving lives, when, come to find out, most of his advice is responsible for killing millions, if not billions, of people.

Then he wrote a book on it all so he can further pad his rich pockets from the whole scamdemic,<sup>9</sup> that's ongoing and may never<sup>9</sup> end.

#### That's why Fauci is the King of Misinformation.

2 - Appeal to fear prejudice (?)
5 - Causal/Oversimplification (?)
6 - Doubt (?)
8 - Flag Waving (?)
9 - Loaded Language (?)
10 - Name Calling, Labeling (?)
14 - Repetition (?)
15 - Slogans (?)
1 - Appeal to Authority
3 - Bandwagon
4 - Black and White Fallacy

7 - Exaggeration, Minimisation

11 - Obfuscation, Intentional Vagueness, Confusion

12 - Red Herring

13 - Reduction ad hitlerum

ACL-2020 (best demo award, honorable mention)

https://www.naturalnews.com/2022-11-15-nuremberg-ii-mass-murder-trials-could-commence-after-midterms.html

ACL-2020 (best demo award, honorable mention): Prta: A System to Support the Analysis of Propaganda Techniques in the News. Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barrón-Cedeño, Preslav Nakov



**Propaganda Detection in Arabic** 

Home Results

More 🕚

~ Q

# SHARED TASK ON PROPAGANDA DETECTION ation which is false or explaining just part of a situ by the government of ARCA by us to make people

co-located with EMNLP 2022 in Abu Dhabi, UAE

### SemEval'2023 Task 3 Data Persuasion Techniques and Logical Fallacies

Gallia est omnis divisa in partes tres, guarum unam incolunt Belgae, aliam Aguitani, tertiam gui ipsorum lingua Celtae, nostra Galli appellantur. Hi omnes lingua, institutis, legibus inter se differunt. Gallos ab Aguitanis Garumna flumen, a Belgis Matrona et Seguana dividit. Horum omnium fortissimi sunt Belgae, propterea guod a cultu atque humanitate provinciae longissime absunt, minimegue ad eos mercatores saepe commeant atque ea quae ad effeminandos animos pertinent important, proximique sunt Germanis, qui trans Rhenum incolunt, quibuscum continenter bellum gerunt. Qua de causa Helvetii quoque reliquos Gallos virtute praecedunt, quod fere cotidianis proeliis cum Germanis contendunt, cum aut suis finibus eos prohibent aut ipsi in eorum finibus bellum gerunt. Eorum una pars, guam Gallos obtinere dictum est, initium capit a flumine Rhodano, continetur Garumna flumine, Oceano, finibus Belgarum, attingit etiam ab Seguanis et Helvetiis flumen Rhenum, vergit ad septentriones. Belgae ab extremis Galliae finibus oriuntur, pertinent

Manipulative Wording: Loaded Language

Attack on Reputation: Smears

Manipulative Wording: Exaggeration

Justification: Appeal to Values



#### Would your Artificial Intelligence system be able to detect them ? Take part in the SemEval 2023 Shared Task 3 to find out !





### Fake News Generation/Detection – Propaganda Connection (work in progress)

AJDABIYAH, Libya | Thu Apr 7, 2011 6:34 pm EDT AJDABIYAH, Libya -LRB- Reuters -RRB- - Rebels fighting to overthrow Muammar Gaddafi said five of their fighters were killed when NATO planes mistakenly bombed a rebel tank column near the contested port of Brega in eastern Libya... "In rebel-held eastern Libya, wounded rebels being brought to a hospital Ajdabiyah said their trucks and tanks were hit on Thursday by a NATO air strike outside Brega. NATO-said it was investigating an attack by its aircraft on a tank column in the area along the Mediterranean coast on Thursday, saying the situation was " unclear and fluid . " Rebels said at least five of their fighters were killed when NATO planes mistakenly bombed a rebel tank column near the contested port. "A number of vehicles were hit by a NATO strike ", officers from UN concluded. The fighting for Brega, the only active front, has dragged on for a week and has entered a daily pattern of advances back and forth with neither side making major gains ...

Table 1: An example of our generated fake news. Given an authentic news article, our approach first identifies <del>a</del> salient sentence, which it then replaces with a plausible but disinformative sentence that is coherent to the context. Finally, it generates a propaganda sentence to make the article resemble human-written fake news.



### Coordination – Propaganda Connection in the UK Elections



**ACM WebSci'2022 (best paper award):** *The Spread of Propaganda by Coordinated Communities on Social Media* Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, Preslav Nakov

# Multimodality

### Propaganda Detection in Memes: COVID-19





ACL-2021: Detecting Propaganda Techniques in Memes. Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov and Giovanni Da San Martino

### Propaganda in Memes: Political



ACL-2021: Detecting Propaganda Techniques in Memes.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, Giovanni Da San Martino

### **Detecting Harmful Memes and Their Targets**







THERE IS A MUCH MORE SINISTER REASON WHY CHINA LET THE VIRUS GO AND WITHHELD INFORMATION.

AS PELOSI SAID... CHINA PREFERS BIDEN TO WIN THE ELECTION. WHERE'S HUNTER? WILL HUNTER BE SUICIDED LIKE EPSTEIN



#### Type:

. . .

- Fake news
- Hate speech
- Attacks, sarcasm

#### Harmfulmess:

- very harmful
- partially harmful
- harmless

#### Target:

- individual
- organization
- community/country
- society/general public/others

ACL'2021 (Findings): Detecting Harmful Memes and Their Targets.

Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty

### **Detecting Harmful Memes and Their Targets**



**EMNLP'2021 (Findings):** *MOMENTA: A Multimodal Framework for Detecting Harmful Memes and Their Targets* Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty

### Detecting the Specific Target of Harmful Memes

In 1993, Bill appointed Hillary to head the White House Health Care Reform Committee



Candidates: bill clinton. hillary clinton, white house, donald trump, democrat Harmful: bill clinton. hillary clinton, democrat Not-harmful: white house. donald trump

**NAACL-HLT'2022 (Findings):** *DISARM: Detecting the Victims Targeted by Harmful Memes* Shivam Sharma, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty

Detecting the Specific Target of Harmful Memes



**NAACL-HLT'2022 (Findings):** *DISARM: Detecting the Victims Targeted by Harmful Memes* Shivam Sharma, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty

### Who is the Hero, the Villain, the Victim?



Figure 1: Examples of heroes, villains and victims, as portrayed within memes.

**EACL-2023:** Characterizing the Entities in Harmful Memes: Who is the Hero, the Villain, the Victim? Shivam Sharma, Atharva Kulkarni, Tharun Suresh, Himanshi Mathur, Preslav Nakov, Md. Shad Akhtar, Tanmoy Chakraborty

### Who is the Hero, the Villain, the Victim?



**EACL-2023:** Characterizing the Entities in Harmful Memes: Who is the Hero, the Villain, the Victim? Shivam Sharma, Atharva Kulkarni, Tharun Suresh, Himanshi Mathur, Preslav Nakov, Md. Shad Akhtar, Tanmoy Chakraborty

### Explaining Why a Hero, a Villain, a Victim



**AAAI'2023:** What do you MEME? Generating Explanations for Visual Semantic Role Labelling in Memes Shivam Sharma, Siddhant Agarwal, Siddhant Suresh, Preslav Nakov, Shad Md. Akhtar, Tanmoy Chakraborty

### Explaining Why a Hero, a Villain, a Victim



**AAAI'2023:** What do you MEME? Generating Explanations for Visual Semantic Role Labelling in Memes Shivam Sharma, Siddhant Agarwal, Siddhant Suresh, Preslav Nakov, Shad Md. Akhtar, Tanmoy Chakraborty

# From "Fake News" to "Infodemic"

### "We are not just fighting a pandemic; we are fighting an infodemic."





To help alleviate suffering and save lives, WHO has been working night and day in five key ways:

1. Helping build countries' capacity to prepare and respond

2. Providing accurate information and fight the infodemic, together with numerous partners

3. Ensuring supplies of essential medical equipment for frontline health workers.

- 4. Training and mobilizing health workers.
- 5. Accelerating research and development.

# The coronavirus is the first true social-media "infodemic"

Social media has zipped information and misinformation around the world at unprecedented speeds, fueling panic, racism .. and hope.

by Karen Hao and Tanya Basu

February 12, 2020

#### On January 19—a week before the Lunar New Year—Tommy Tang left

Shenzhen with his girlfriend to visit her family in Wuhan for the holiday. They had heard of the novel coronavirus (now officially known as COVID-19), but as far as they knew, it was localized to a small area. The local government had assured people that it would only affect those who visited a specific food market and contracted it directly from wild animals. https://www.technologyreview.com/ 2020/02/12/844851/thecoronavirus-is-the-first-true-socialmedia-infodemic/

### TYPES OF INFORMATION DISORDER

#### FALSENESS INTENT TO HARM

#### Misinformation

Unintentional mistakes such as innaccurate photo captions, dates, statistics, translations, or when satire is taken seriously.

#### Disinformation

Fabricated or deliberately manipulated audio/visual content. Intentionally created conspiracy theories or rumours.

#### Malinformation

Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn. Deliberate change of context, date or time of genuine content.

> https://medium.com/1stdraft/information-disorderpart-3-useful-graphics-2446c7dbb485 59

Claire Wardle of First Draft News

### **Evolution of the Terminology**

- 1. "Fake news": focus on factuality
- 2. Disinformation: factuality + harm
- 3. "Infodemic": focus on harm

### **COVID-19 Infodemic: Holistic Approach**

✓ Mar 16 · Mar 16 · Mar 16 ② خالد جاسم ■ قاعد في البيت غير صحيح مايتم تداوله حالياً عن قيام مروحيات عسكرية برش مبيدات ضد قايروس كورونا في سماء جميع أنحاء البلاد ، هذى اشاعة ،،

onovinua

Military helicopters spray pesticides against Coronavirus. This is rumor!

(a) rumor in Arabic and English

AND SO THE ZOMBIE APOCALYPSE BEGINS..., MAKE READY YOUR WEAPONS PEOPLE...

Coronavirus: Huge queues as Americans panic buy guns amid virus outbreak | US News | Sky News



Coronavirus: Huge queues as Americans panic buy guns amid virus outbreak People buying more guns and ammunition have reportedly said they are preparing for a breakdown in public order over COVID-19.  $\mathscr{D}$  news.sky.com

3:00 PM - Mar 17, 2020 - Twitter for Android

#### (b) spreading panic

Nameste 🙏 Twitter 😍

Today I wanna share something useful & effective for all of us -

GOOD NEWS!!! Wuhan's Corona virus can cure itself by a bowl of freshly boiled garlic water. The old Chinese doctor proved its effectiveness. Many patients have also proven it to be effective.

9:20 AM · Mar 11, 2020 · Twitter for Android

#### (d) bad cure

Along with Chinese food, Italian food is on the list of things I do not order anymore #coronavirus #MoshUp

3:41 AM · Mar 9, 2020 · Twitter for Android

#### (e) xenophobic/racist/prejudices

After weeks of the Federal Government dragging its feet on COVID-19 testing, Trump just admitted in front of the cameras he wants to manipulate the measurement of the outbreak for his personal political benefit.

1:40 AM · Mar 7, 2020 · Twitter Web App

#### (f) blaming/accusing the authorities

#### Replying to @spectatorindex

If the Corona virus enters Spain He will enter from the Barcelona defense side

11:05 PM - Mar 14, 2020 - Twitter for Android

(c) joke

Germany is conducting 160k Covid-19 tests a week. It has a total 35k ventilators, 10k ordered to be made by the govt. It has converted a new 1k bed hospital in Berlin.

It's death rate is tiny bcos it's mass testing allows quarantine and bcos it has fewer non reported cases.

10:43 AM · Mar 20, 2020 · Twitter for iPhone

#### (g) advice/discussion of action taken

My body, my health. We don't need a killer like BillGates telling us what's healthy for us. This is the absolute insult to our intelligence. #NWO #5G #Agenda21 #ID2020 #coronavirus #COVID19

#### Think my Coronavirus is killer? Wait till I roll out the Mandatory Vaccine!



4:47 PM - May 12, 2020 - Twitter for Android

(h) conspiracies

EMNLP'2021 (Findings): Fighting the COVID-19 Infodemic: Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society. Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, Abdulaziz Al-Homaid, Wajdi Zaghouani, Tommaso Caselli, Gijs Danoe, Friso Stolk, Britt Bruntink, Preslav Nakov.

## Can We Win the War on "Fake News"?

### **Complex Problem, No Easy Solution**

- Need for cooperation
  - social media / technology companies
  - governments
  - international organizations
  - journalists, fact-checkers, media, NGOs, etc.
  - researchers: academia, industry

### Finland is winning the war on fake news. What it's learned may be crucial to Western democracy

By Eliza Mackintosh, CNN Video by Edward Kiernan, CNN



https://edition.cnn.com/interactive/201 9/05/europe/finland-fake-news-intl/

#### Media literacy across Europe

Finland ranked first out of 35 countries in a study measuring resilience to the post-truth phenomenon



#### https://edition.cnn.com/interactive/2019/05/europe/finland-fake-news-intl/

# Conclusion



e losto

अफवाह सुनना नहीं, सुनना तो मानना नहीं।

DO NOT LISTEN TO RUMOUR, BUT, IF YOU DO, DO NOT BELIEVE IT.

HT-20 JIM



### MSc & PhD PROGRAMS & SCHOLARSHIP OPPORTUNITIES



mbzuai.ac.ae

### ""

### MBZUAI is the world's first Al university, specializing in research and graduate-level study

MBZUAI is an open invitation from Abu Dhabi and the UAE to the world to unleash the full potential of AI



### WHAT ISMBZUAI?

#### **GREAT FACULTY**

- World-class academic leadership and international faculty in CV,ML, and NLP- top-40 in world in all three areas, along with only 10 other universities (csrankings.org)
- World-renowned faculty including Eric Xing, Michael Jordan, Le Song, Tim Baldwin
- Currently 4 full-time NLP faculty, with aim to reach 14 by end of 2023
- New departments of HCI, Robotics, and CSto open in 2022-23

#### **GREAT FACILITIES**

- Top-class student experience on a brand-new campus in an international environment
- World-leading computing infrastructure

#### **GREAT FINANCIAL SUPPORT**

 Outstanding student support package through its scholarship program so students can focus on learning and research



### SURPRISING STATISTICS

- Age = 2
- Total faculty = 40
- Total researchers = 120
- Total students = 250
- Student gender mix = 31% female
- Total nationalities represented among students = 40
- Language for all instruction (and most of everything) = English
- On-campus pools, gyms, games rooms = 1 x 1 x 1
- On-campus restaurants = 8
- Green energy = 100% (from our own solar farm)
- Major international collaborations: Weizman Institute of Science, IBM



mbzuai.ac.ae

### MBZUAI PROGRAMS

MBZUAI offers MSc and PhD programs:

#### MSc Programs (2 years)

- Master of Science programs in in Computer Vision, Machine Learning, and NLP (with Robotics and CSto be added in 2023)
- First year of coursework, internship over the summer semester, and second year largely research

#### PhD Programs (4 years)

- PhD programs in in Computer Vision, Machine Learning, and NLP (with Robotics and CSto be added in 2023)
- First year of coursework, internship over the summer semester, and second to fourth years largely research

Also faculty and post-doc opportunities


## MBZUAI SCHOLARSHIPS

All full-time students admitted to MBZUAI receive the following support:

- Full tuition waiver
- Competitive stipend (= approx. USD\$2200 [MSc] or USD\$2800 [PhD] per month)
- Free on-campus accommodation (studio apartment)
- Annual travel to and from home country
- Health insurance



### MBZUAI ADMISSION CRITERIA

#### For MSc Programs

<b>e</b>			
Degree	Academic Grades	English Language	Graduate Record Examination (GRE)
Bachelor's degree in a <b>STEM field</b> such as Computer Science, Electrical Engineering, Computer Engineering, Mathematics, Physics and other relevant Science and Engineering majors	Student should have a minimum undergraduate <b>CGPA of 3.2 (on a</b> <b>4.0 scale)</b>	Valid certificate copies of either a standard TOEFL iBT with a minimum total score of 90; IELTS Academic with a minimum overall score of 6.5; or EmSAT English with a minimum score of 1550	General score is <b>optional</b> and would be considered in the evaluation of the applicants
For PhD Programs			
rarnerrogiane			
Degree	Academic Grades	English Language	Graduate Record Examination (GRE)
Bachelors' and Master's degrees in a <b>STEM field</b> such as Computer Science, Electrical Engineering, Computer Engineering, Mathematics, Physics and other relevant Science and Engineering majors	Student should have a minimum CGPA of 3.2 (on a 4.0 scale) for both Bachelors and Masters degrees	Valid certificate copies of either a standard TOEFL iBT with a minimum total score of 90; IELTS Academic with a minimum overall score of 6.5; or EmSAT English with a minimum score of 1550	General score is <b>optional</b> and would be considered in the evaluation of the applicants 3 Recommendation letters
		Contract I and I and the	

mbzuai.ac.ae

### APPLY NOW

https://mbzuai.ac.ae/study/admissions

MBZUAI is accepting applications now for Fall 2023 – early application recommended

For PhD applications, best to work with potential supervisor in developing research proposal

# QUESTIONS?

Do you have questions about the programs or how to apply? Help with grade equivalence from your degree? Help finding the right supervisor?

Contact Prof Tim Baldwin

timothy.baldwin@mbzuai.ac.ae

mbzuai.ac.ae