

Fact-Checking, Fake News, Propaganda, and Media Bias:

Truth Seeking in the Post-Truth Era

Preslav Nakov, Qatar Computing Research Institute, HBKU

Giovanni Da San Martino, University of Padova

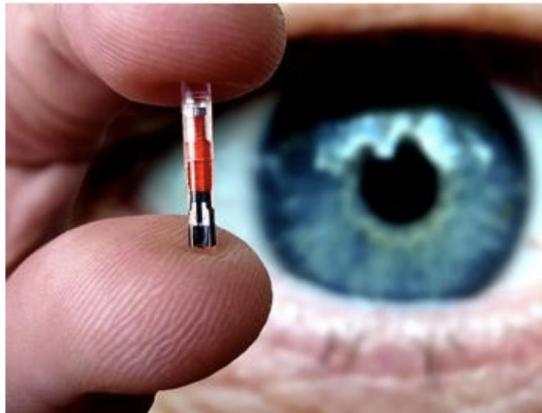
EMNLP'2020 tutorial

Is This “Fake News”?

It Is Coming The Agenda Behind Covid-19 Was Microchip Or As Bill Gates Says “Digital Tracking System” For Virus

MAR
26

👤 INTERNIST



<https://noesis2017.com/2020/06/17/it-is-coming-the-agenda-behind-covid-19-was-microchip-or-as-bill-gates-says-digital-tracking-system-for-virus-ella-cruz/>

SHARES

f Share

Tweet

Who would have thought it will come so soon? The same person who urged the world to **“reduce the population growth through vaccines” and insisted that through “reproductive health and vaccines we can reduce the population”** and again, the same

NEWS IN PHOTOS

Trump Delays Easter To July 15 To Keep Promise On Coronavirus

3/30/20 2:59PM • SEE MORE: DONALD TRUMP ▾



Recent Video



Tips For Parenting During Co

OnionNews |

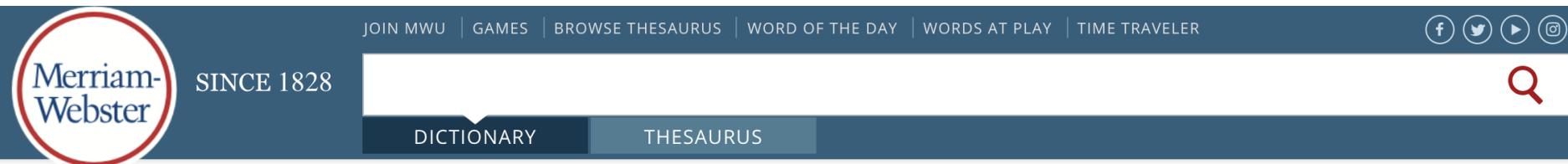
You may also like



<https://politics.theonion.com/trump-delays-easter-to-july-15-to-keep-promise-on-coronavirus-1842566559>

What is “Fake News”?

Merriam-Webster: Self-Explanatory, No Need to Define



Words We're Watching

The Real Story of 'Fake News'

The term seems to have emerged around the end of the 19th century

The reason *fake news* is unlikely to be entered in our dictionary anytime soon is that it is a self-explanatory compound noun — a combination of two distinct words, both well known, which when used in combination yield an easily understood meaning. *Fake news* is, quite simply, *news* ("material reported in a newspaper or news periodical or on a newscast") that is *fake* ("false, counterfeit").

Secretary Brunnell Declares **Fake News** About His People is Being Telegraphed Over the Country.

—*Cincinnati Commercial Tribune* (Cincinnati, OH), 7 Jun. 1890

Fake News. The following is handed to us for publication: Sunday's Enterprise says that I and a companion were run over by the Neptune and thrown into the water. As can be proved by more than one, we did not so much as get our feet wet, nor were we helped into the Neptune. Clarence Collins.

—*The Kearney Daily Hub* (Kearney, NE), 7 Jul. 1890

The public taste is not really vitiated and it does not in its desire for 'news' absolutely crave for distortions of facts and enlargements of incidents; and it certainly has no genuine appetite for **'fake news'** and 'special fiend' decoctions such as were served up by a local syndicate a year or two ago.

—*The Buffalo Commercial* (Buffalo, NY), 2 May 1891



WHAT GIVES?

Fake News Is An Oxymoron

Written By **Chitra Subramaniam** | Mumbai | Published: November 18, 2018 22:59 IST



Ever since the fountainhead of fake news moved into the White House two years ago, the news space for lesser mortals including media houses has taken a serious beating.

<https://www.republicworld.com/opinions/what-gives/fake-news-is-an-oxymoron>

Definition of 'fake news'

English: fake news

Example sentences

fake news in British

noun

false, often sensational, information disseminated under the guise of news reporting

Collins English Dictionary. Copyright © HarperCollins Publishers



<https://www.collinsdictionary.com/dictionary/english/fake-news>

Collins



- 2016** - BREXIT
- 2015** - BINGE-WATCH
- 2014** - PHOTOBOMB
- 2013** - GEEK

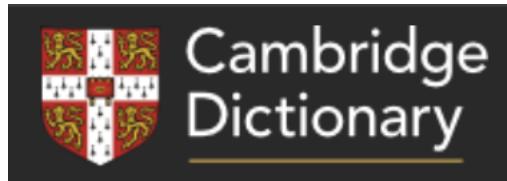


Oxford Dictionaries

WORD OF THE YEAR

post-truth

2016



Meaning of **fake news** in English



fake news

noun [U] • **UK** / ,feɪk 'nju:z / **US** / ,feɪk 'nu:z /

★ **false stories that appear to be news, spread on the internet or using other media, usually created to influence political views or as a joke:**

There is concern about the power of fake news to affect election results.

— Thesaurus: synonyms and related words

- Fake news is a neologism often used to refer to fabricated news. This type of news, found in traditional news, **social media or fake news websites**, has no basis in fact, but is presented as being factually accurate.
- Fake news is a type of yellow journalism or propaganda that consists of deliberate disinformation or hoaxes spread via traditional news media (print and broadcast) or **online social media**
- Fake news is written and published usually with the intent to mislead in order to damage an agency, entity, or person, and/or gain financially or politically, often using sensationalist, dishonest, or outright fabricated headlines to increase readership. Similarly, **clickbait** stories and headlines earn advertising revenue from this activity.

Maybe It Is Just “News I Don’t Like”?

The FAKE NEWS media (failing @nytimes, @NBCNews, @ABC, @CBS, @CNN) is not my enemy, it is the enemy of the American People!

RETWEETS

51,272

LIKES

162,294



3:48 PM - 17 Feb 2017

78K

51K

162K

Maybe it Cannot be Formally Defined, but People Recognize It When They See It

A black and white portrait of US Supreme Court Justice Potter Stewart. He is wearing a dark suit and tie, and has a serious expression. He is standing in front of a bookshelf filled with books.

I shall not today attempt further to define the kinds of material but I know it when I see it.

(Potter Stewart)

izquotes.com

I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description ["hard-core pornography"], and perhaps I could never succeed in intelligibly doing so. But I know it when I see it.

Actually, Most People Do Not Recognize the “Fake News”



The screenshot shows a news article from BNT News dated 08.03.2019. The header includes the logo of Bulgarian National Television (БНТ) and links to watch online channels 1, 2, and 3. The main headline reads: "Mariya Gabriel: 75% of young people do not recognize fake news". Below the headline, a subtext states: "Europe is taking small steps to fight against fake news, but the road is very long. Mariya Gabriel, European Commissioner for Digital Economy and Digital Society told ...". There is a "Share with friends" section with icons for Twitter, Facebook, and Google+. At the bottom, there is a video thumbnail showing a woman in a pink blazer smiling, with the BNT 1 logo in the top left corner of the thumbnail.

БЪЛГАРСКА НАЦИОНАЛНА ТЕЛЕВИЗИЯ · Б·Н·Т· BULGARIAN NATIONAL TELEVISION

WATCH BNT ONLINE

·Б·Н·Т· 1 ·Б·Н·Т· 2 ·Б·Н·Т· 3

08.03.2019 BNT News

Mariya Gabriel: 75% of young people do not recognize fake news

Europe is taking small steps to fight against fake news, but the road is very long. Mariya Gabriel, European Commissioner for Digital Economy and Digital Society told ...

Share with friends

Twitter Facebook G+

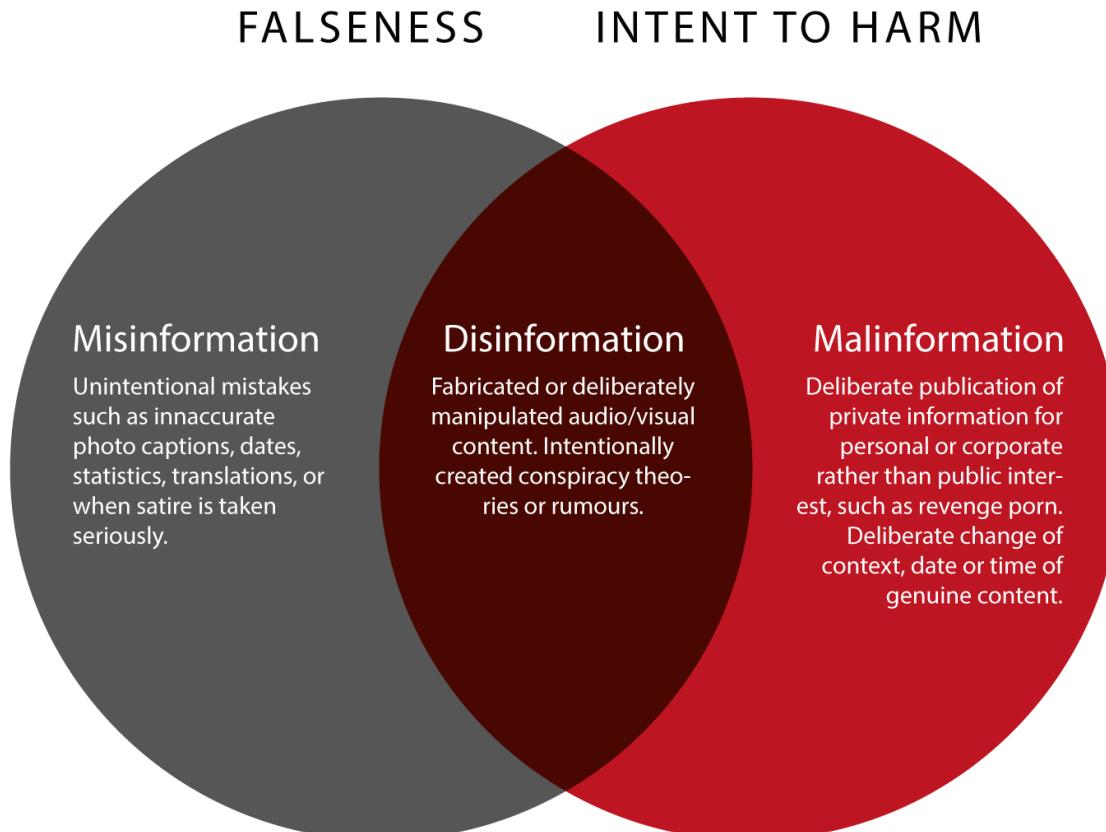
BNT 1

Varna 6° 8:17

<https://www.bnt.bg/en/a/mariya-gabriel-75-of-young-people-do-not-recognize-fake-news>

Disinformation

TYPES OF INFORMATION DISORDER



**SATIRE OR PARODY**

No intention to cause harm but has potential to fool

**MISLEADING CONTENT**

Misleading use of information to frame an issue or individual

**IMPOSTER CONTENT**

When genuine sources are impersonated

**FABRICATED CONTENT**

New content is 100% false, designed to deceive and do harm

**FALSE CONNECTION**

When headlines, visuals or captions don't support the content

**FALSE CONTEXT**

When genuine content is shared with false contextual information

**MANIPULATED CONTENT**

When genuine information or imagery is manipulated to deceive

**Why Should We Fight
the “Fake News”?**

 NEWSWEEK MAGAZINE

How Big Data Mines Personal Info to Craft Fake News and Manipulate Voters

BY NINA BURLEIGH ON 6/8/17 AT 1:01 PM



<https://www.newsweek.com/2017/06/16/big-data-mines-personal-info-manipulate-voters-623131.html>



Hoteles en Tiflis
desde \$9

Reserve ahora



Hoteles en Portland
desde \$60

Reserve ahora



Hoteles en Mascate
desde \$25

Reserve ahora

KAYAK

KAYAK: Buscador Hoteles

Fake news handed Brexiteers the referendum – and now they have no idea what they're doing



'Would we have won without immigration? No. Would we have won without...the NHS? All our research and the close result strongly suggests no. Would we have won by spending our time talking about trade and the single market? No way'

Andrew Grice | @IndyPolitics

Wednesday 18 January 2017 16:45 | 157 comments



Click to follow
The Independent Voices

0 Notifications
<https://www.facebook.com/theindependentvoices>



European Elections

Polling from across Europe. Updated daily.



Half of European voters may have viewed Russian-backed ‘fake news’

Kremlin-backed campaigns are promoting extremist views and amplifying them to sow discord, says cyber firm.

By MARK SCOTT | 5/7/19, 12:45 PM CET | Updated 5/8/19, 7:00 PM CET

<https://www.politico.eu/article/european-parliament-russia-mcafee-safeguard-cyber/>

Web creator Tim Berners-Lee blasts Facebook, saying it makes his invention easy to 'weaponize'

Published: Mar 19, 2018 4:40 a.m. ET



Aa

On World Wide Web's 29th birthday, Tim Berners-Lee criticizes its "gatekeepers" such as Facebook and suggests more regulation



Computer scientist Tim Berners-Lee, the "father of the World Wide Web."

<https://www.marketwatch.com/story/webs-creator-blasts-the-tech-giants-that-make-his-invention-easy-to-weaponize-2018-03-12>

WORLD



4:13

Viral WhatsApp Messages Are Triggering Mob Killings In India

July 18, 2018 · 9:12 AM ET

[+ QUEUE](#)[DOWNLOAD](#)[EMBED](#)

<https://www.npr.org/2018/07/18/629731693/fake-news-turns-deadly-in-india>

U.N. Fact Finders Say Facebook Played a 'Determining' Role in Violence Against the Rohingya



A girl looks out of her tent in the Balukhali refugee camp in Cox's Bazar, Bangladesh on Jan. 17, 2018. U.N. fact finders have pointed to the role that hate speech on social media has played in fueling anti-Rohingya sentiment and violence

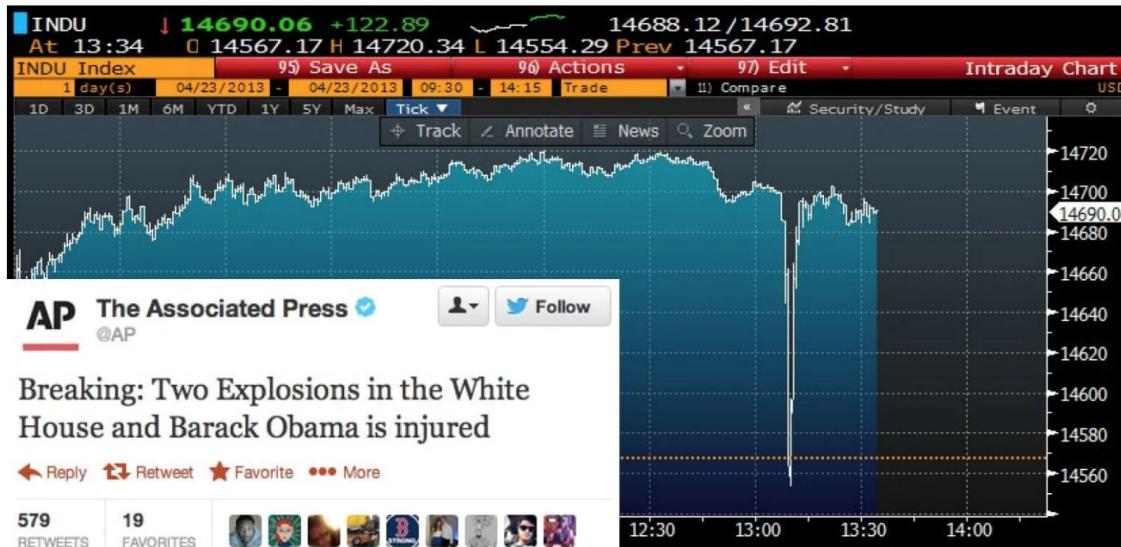
<https://time.com/5197039/un-facebook-myanmar-rohingya-violence/>

Syrian hackers claim AP hack that tipped stock market by \$136 billion. Is it terrorism?

<https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/>

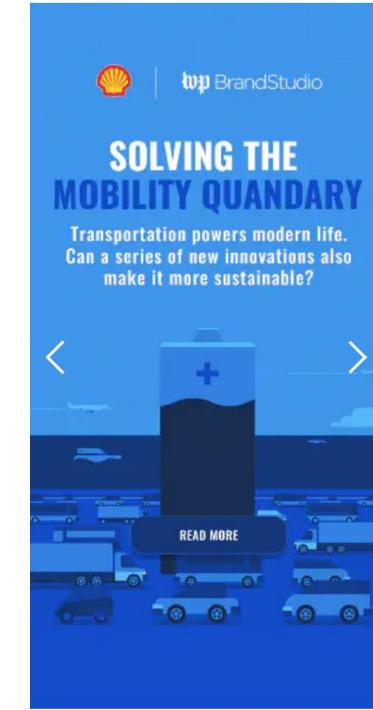
By Max Fisher

April 23, 2013



This chart shows the Dow Jones Industrial Average during Tuesday afternoon's drop, caused by a fake A.P. tweet, inset at left.

At 1:07 p.m. on Tuesday, when the official Twitter account of the Associated Press sent a [tweet](#) to its nearly 2 million followers that warned, "Breaking: Two Explosions in the White House and Barack Obama is injured," some of the people who momentarily panicked were apparently on or near the trading floor of the New York Stock Exchange.



UNICEF blames anti-vaxxers for the 300% spike in global measles outbreaks

Published: Apr 25, 2019 3:49 p.m. ET



Aa

'The ground for the global measles outbreaks we are witnessing today was laid years ago,' report says



<https://www.marketwatch.com/story/unicef-blames-anti-vaxxers-for-the-300-spike-in-global-measles-outbreaks-2019-04-25>



Measles cases have spiked 300% around the world over the past year.

Four European countries stripped of 'measles-free' status: WHO

Disease no longer considered eliminated in UK, Greece, the Czech Republic and Albania as WHO warns of spike in cases.

29 Aug 2019



Nearly 365,000 measles cases were reported globally this year, the highest since 2006 [Lindsey Wasson/Reuters]

<https://www.aljazeera.com/news/2019/08/european-countries-stripped-measles-free-status-190829071033053.html>

“We’re not just fighting a pandemic;
we’re fighting an infodemic.”



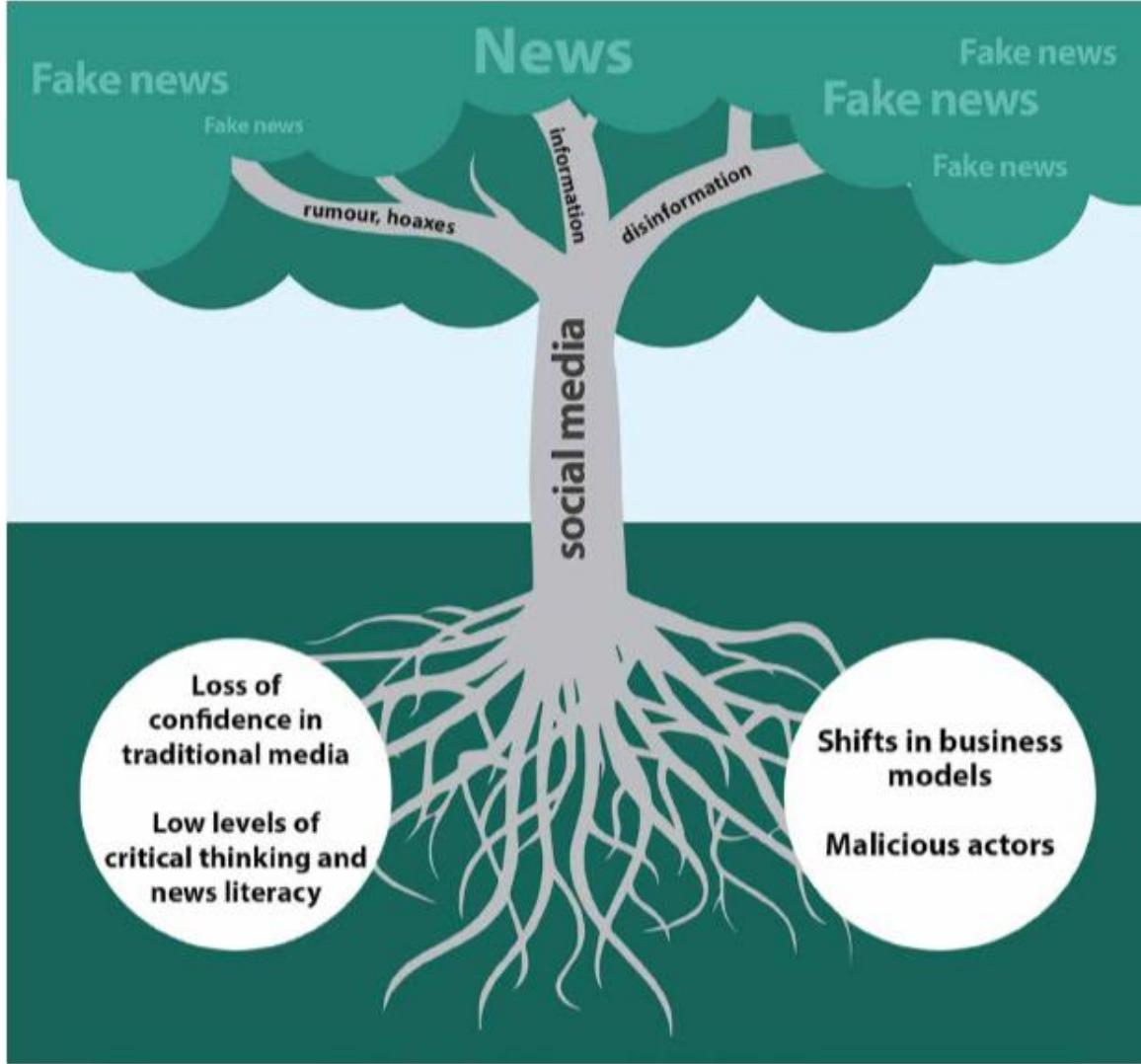


To help alleviate suffering and save lives, WHO has been working night and day in five key ways:

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.

<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>

Why Fake News?



By UNESCO - World Trends in Freedom of Expression and Media Development Global Report 2017/2018, CC BY-SA 3.0-igo, <https://commons.wikimedia.org/w/index.php?curid=70286494>

Fake news 2.0: personalized, optimized, and even harder to stop

Artificial intelligence will automate and optimize fake news, warns a technology supplier to US intelligence agencies.

by **Will Knight**

Mar 27, 2018

Fake news may have already influenced politics in the US, but it's going to get a lot worse, warns an AI consultant to the US government.

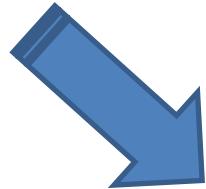
"The automation of the generation of fake news is going to make it very effective," Gourley told the audience at EmTech Digital, organized by *MIT Technology Review*.

<https://www.technologyreview.com/s/610635/fake-news-20-personalized-optimized-and-even-harder-to-stop/>

Manual Fact-Checking

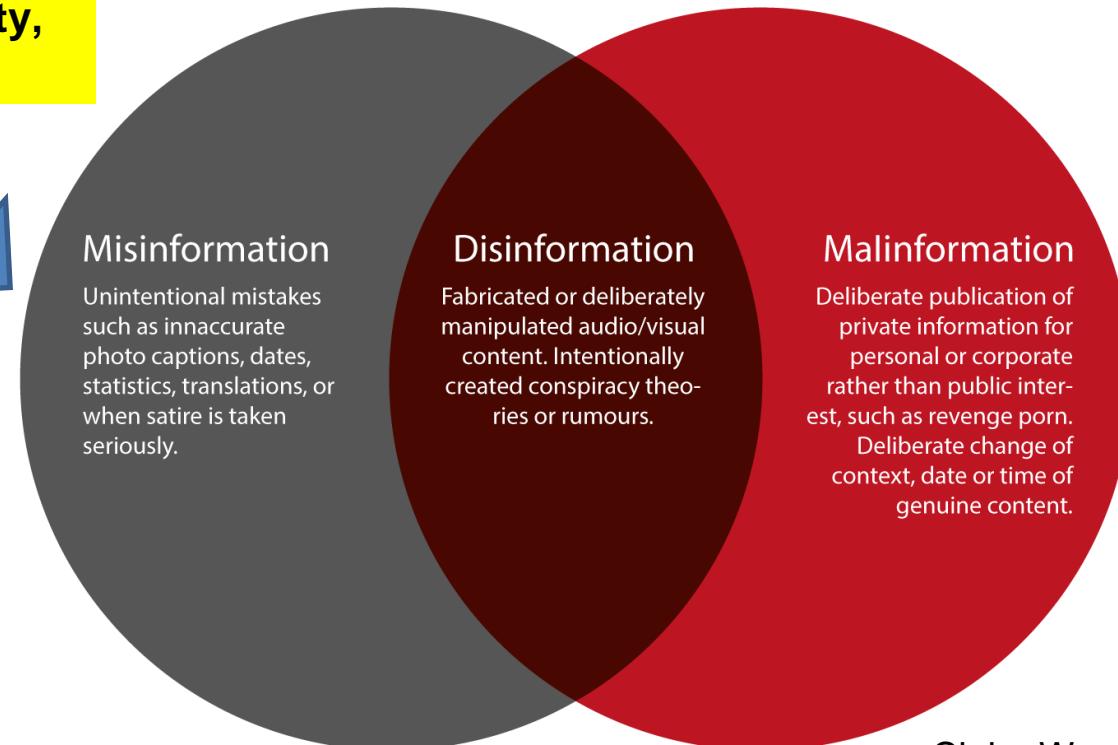
TYPES OF INFORMATION DISORDER

(Manual) fact-checking focuses on factuality, ignoring harm.



FALSENESS

INTENT TO HARM



The Fact-Checking Landscape

 FACTCHECK.ORG A Project of the Democracy Public Policy Group

HOME ARTICLES ASK A QUESTION VISUAL SPOTLIGHT ARCHIVES ABOUT US SEARCH



Timeline of Russia Investigation

June 7, 2017

Key moments in the FBI probe of Russia's efforts to influence the 2016 presidential election

 WINNER OF THE PULITZER PRIZE

My Comment  

Search 

SECTION: [TRUTH-O-METER](#) | [PEOPLE](#) | [PROMISES](#) | [PARTS OF FIRE](#) | [ABOUT US](#) | [LOG IN](#)





 Snopes

CONTACT US | GET THE NEWSLETTER

What's New | Most Read | Facebook | Home | User | About | Help | Log In

Limits on Travel to U.S. Issued After Supreme Court Ruling

People from six Muslim-majority countries will be allowed to enter the United States if they are proven family relationships.



 Snopes Facebook Group

Like Us | Follow Us | RSS Feeds

Full Fact is the UK's independent factchecking charity

Get the facts you need straight to your inbox

email address

Subscribe



Autism prevalence is up but scientists have consistently found that vaccines don't cause it



Are assaults on police officers rising?

Most Recent



Full Fact publishes first report on Facebook's Third Party Fact Checking programme

28TH AUG 2019

Can refugees enter the UK illegally?

27TH AUG 2019

Does a no deal Brexit mean no divorce bills?

About Us | Verification Plugin | Results | News And Events | Blog



Partners/ Consortium

HOME | PARTNERS/ CONSORTIUM

 ontotext

ONTOTEXT will coordinate the overall project as well as lead WP3, where, together with ATC, they will implement the WeVerify open platform for decentralised content verification. Building on its expertise in solution development, platforms targeting media and publishing, and development of open-source tools, ONTO will thus orchestrate and integrate the contributions of the project's technology providers.



The University Of Sheffield.

USFD will lead WP5 which will develop the WeVerify citizen-oriented digital companion. In WP4 the project team will develop human-in-the-loop tools, based on state-of-the-art machine learning-based algorithms from the PHEME project (including rumour detection, stance analysis, and veracity classification). USFD will also engage actively with the definition and implementation of the open-source WeVerify platform (WP3), as well as contribute to the use case pilots, dissemination, and community building.

 PHEME

Computing Veracity – the Fourth Challenge of Big Data

About PHEME | News | Events | Software & Data Downloads | Publications | Deliverables | Consortium

About PHEME

PHEME builds technology for striking how true claims made online are

FAKED OUT EVIDENCE GONE TO AIR IF IT'S BEEN FAUCED

I HEARD A RUMOUR

Checking whether rumours on social media are true

SWI

ON Twitter

The Fact-Checking Landscape

<https://reporterslab.org/fact-checking/>



FACT-CHECKING NEWS

FACT-CHECKING NEWS

Fact-checking count tops 300 for the first time

October 13, 2020

[SEE ALL FACT-CHECKING NEWS](#)

GLOBAL FACT-CHECKING SITES

The Reporters' Lab maintains a database of global fact-checking sites. You can use the map to explore sites around the world or use the menu below. (Here's more [how we identify fact-checkers](#).)

BROWSE IN LIST ▶



How Much Has Been Manually Fact-Checked?

Statistics

Property	Global	Snopes	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

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

Technology could help:

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

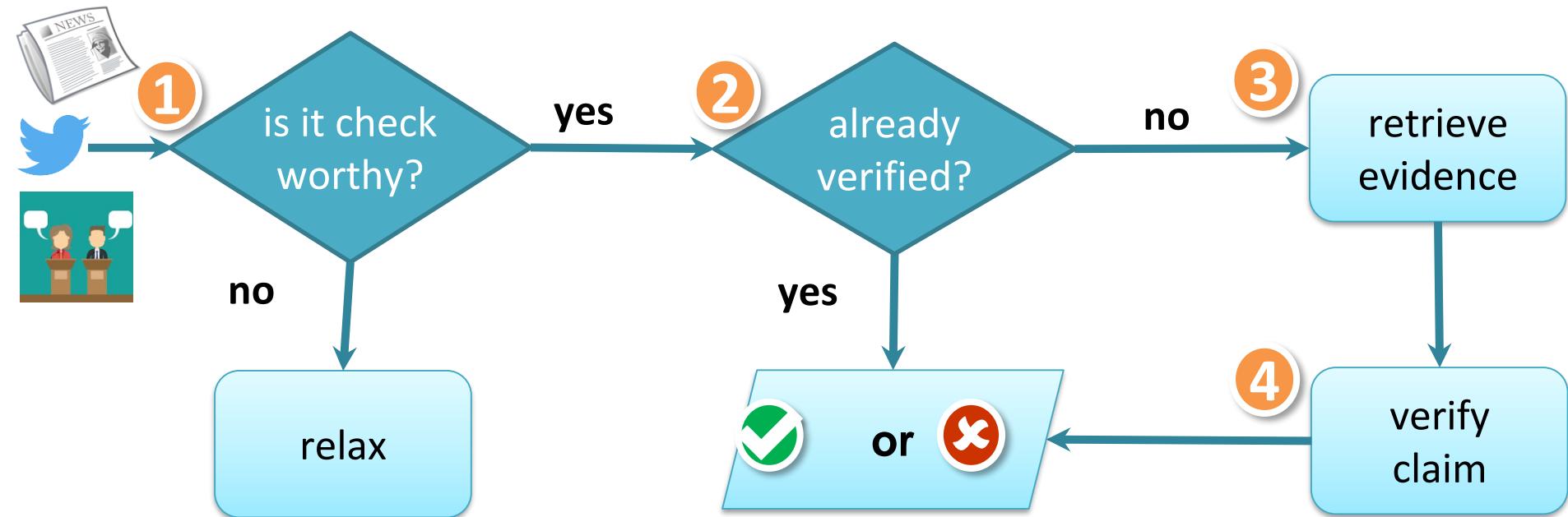
*Round Table at the CLEF-2020 CheckThat! Lab on September 22, 2020:
David Corney, Full Fact*

What do fact checkers want?

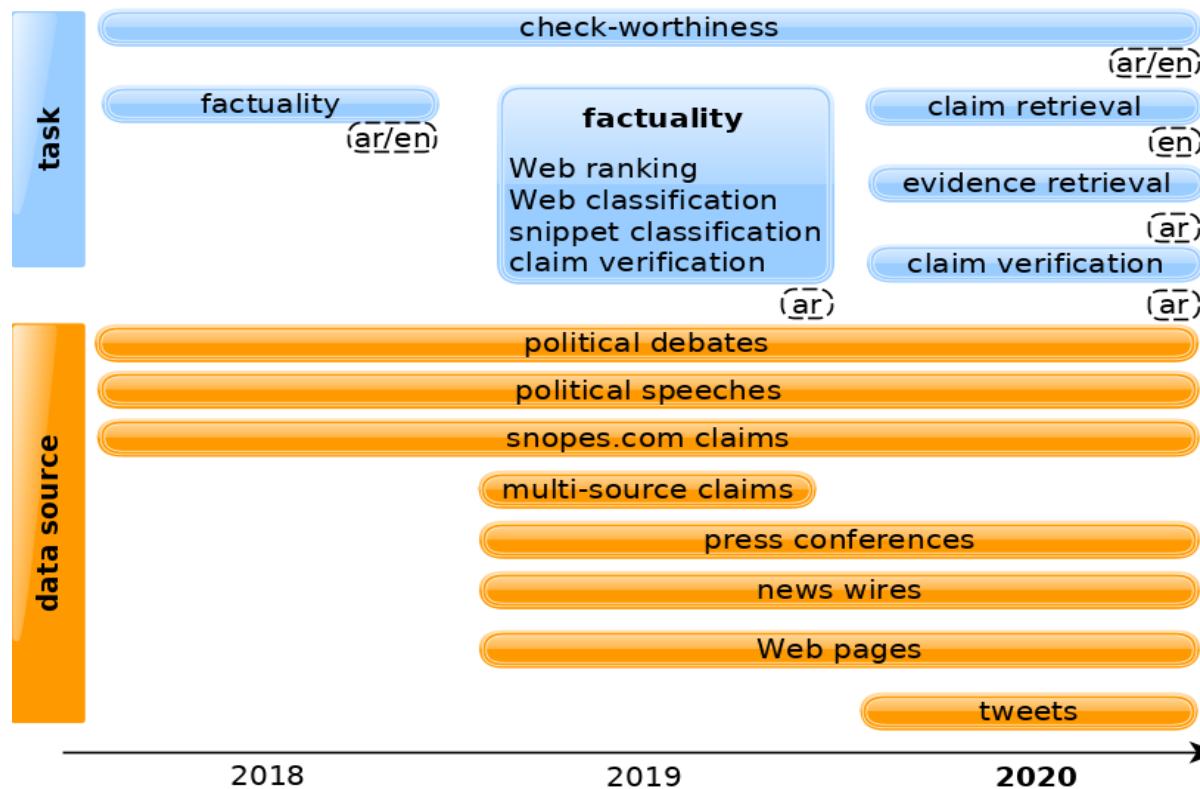
1. To know what is the most important claim to check today
Check-worthiness
2. To know when someone says something that they've already checked
Verified claim retrieval
3. To check claims as quickly as possible
Evidence retrieval / claim verification



The Fact-Checking Flow-Chart



The CheckThat! Lab at CLEF 2018-2021





- Can we fact-check every claim in the world?

This is Very Hard !

50%

**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.

6X

**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!**

What is Worth Fact-Checking? *(In Political Debates)*

Fact-checking Donald Trump, Joe Biden in the final presidential debate

NATIONAL

DEBATES



Donald Trump and Joe Biden at the final 2020 presidential debate. (AP)

<https://www.politifact.com/article/2020/oct/23/fact-checking-donald-trump-joe-biden-final-preside/>



Donald Trump

stated on October 22, 2020 in a presidential debate:

"We are rounding the turn (on coronavirus). We are rounding the corner."

PUBLIC HEALTH

CORONAVIRUS

DONALD TRUMP



A mailbox containing free masks for pedestrians stands in the center of Tenafly, N.J., on Oct. 22, 2020.



The 28 Hottest Gifts of 2020

Get "Black Friday" deals now on these brilliant inventions.

ZangDeal

<https://www.politifact.com/fact-checks/2020/oct/23/donald-trump/no-fight-against-coronavirus-isnt-rounding-corner-/>

Donald Trump: (07:37)

How are you doing? How are you?

Kristen Welker: (07:58)

And I 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 your first question. The Debate Committee 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, 10,000 Americans have died from COVID. So please be specific. How would you lead the country through 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 85%. 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 was a spike in Arizona. It's now gone. And there was some spike in other places, they will soon be gone. We have a vaccine that's ready. It's going to be announced within weeks. And it's going to be delivered. We have Operation Warp Speed, which is the mission to distribute the vaccine.

Donald Trump: (10:04)

I can tell you from personal experience, I was in the hospital and 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 better, wouldn't be here tonight. And how they say I'm immune. Whether it's four months or a lifetime, nobody's been able to say, 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 country out 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: (11:56)

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, in rapid testing. I would make sure that we set up national standards as to how to open 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 this way this president has responded to this crisis has been absolutely tragic. And so folks, I will take care of this. I will end this. I will make us 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 quickly.

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)

President Trump, your reaction?

Let me follow up with you and because this is your information. You have

Donald Trump: (13:41)

..

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

Donald Trump: (10:04)

I can tell you from personal experience, I was in the hospital and 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 better, wouldn't be here tonight. And how they say I'm immune. Whether it's four months or a lifetime, nobody's been able to say, 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 country out 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: (11:56)

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.

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 quicker," but he didn't move quicker. 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 lethal, but it was a total disaster. Had that 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: (16:01)

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: (16:58)

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 scientist in the world who thinks it's going to be over soon.

Kristen Welker: (17:03)

President Trump, your reaction?

Donald Trump: (17:03)

..

Donald Trump: (17:03)

Pioneering Work: *ClaimBuster*

<https://idir.uta.edu/claimbuster/>



ClaimBuster: Automated Live Fact-checking

Data Collection Press Ackno



Fact-check Your Own Text



U.S. Presidential Debates

Live Event

ClaimPortal

ClaimBuster API

Tweets by @ClaimBusterTM

 ClaimBuster
@ClaimBusterTM



From #PresidentialDebate: Hispanic, women, asian, people with diplomas, with no diplomas, mit graduates, number one in their class, everybody had the best numbers.
#FactCHAT #Elections2020



Oct 23, 2020

 ClaimBuster
@ClaimBusterTM



From #PresidentialDebate: We had the best black unemployment numbers in the history of our country. #FactCHAT #Elections2020



Oct 23, 2020

 ClaimBuster
@ClaimBusterTM



From #PresidentialDebate: 11.4 million jobs in a short period of time, et cetera.

Naeemul Hassan, Chengkai Li, Mark Tremayne:

Detecting Check-worthy Factual Claims in Presidential Debates. CIKM 2015: 1835-1838

ClaimBuster: Sample Output

- 0.21 I'm chris wallace of fox news.
- 0.43 I welcome you to the third and final of the 2016 presidential debates between secretary of state hillary clinton and donald j. Trump.
- 0.27 This debate is sponsored by the commission on presidential debate.
- 0.19 The commission has designed the format.
- 0.40 Six roughly 15-minute segments with two-minute answers to the first question, then open discussion for the rest of each segment.
- 0.37 Both campaigns have agreed to those rules.
- 0.43 For the record, i decided the topics and the questions in each topic, none of those questions has been shared with the commission or the two candidates.
- 0.35 The audience here in the hall is promised to -- has promised to remain silent.
- 0.12 No cheer, boos or other interruptions so we and you can focus on what the

The *ClaimBuster* Manual Annotation Schema

SE: Actually, we've increased funding for dealing with nuclear proliferation about 35 percent since I've been the president.

More Context

Will the general public be interested in knowing whether (part of) this sentence is true or false?

- There is **no** factual claim in this sentence.
- There is a factual claim but it is **unimportant**
- There is an **important** factual claim.

Submit

Skip this sentence

Modify My Previous Responses

The *ClaimBuster* Manual Annotation Schema

Non-Factual Sentence (*NFS*): Subjective sentences (opinions, beliefs, declarations) and many questions fall under this category. These sentences do not contain any factual claim. Below are some examples.

- *But I think it's time to talk about the future.*
- *You remember the last time you said that?*

The *ClaimBuster* Manual Annotation Schema

Unimportant Factual Sentence (*UFS*): These are factual claims but not check-worthy. In other words, the general public will not be interested in knowing whether these sentences are true or false. Fact-checkers do not find these sentences as important for checking. Some examples are as follows.

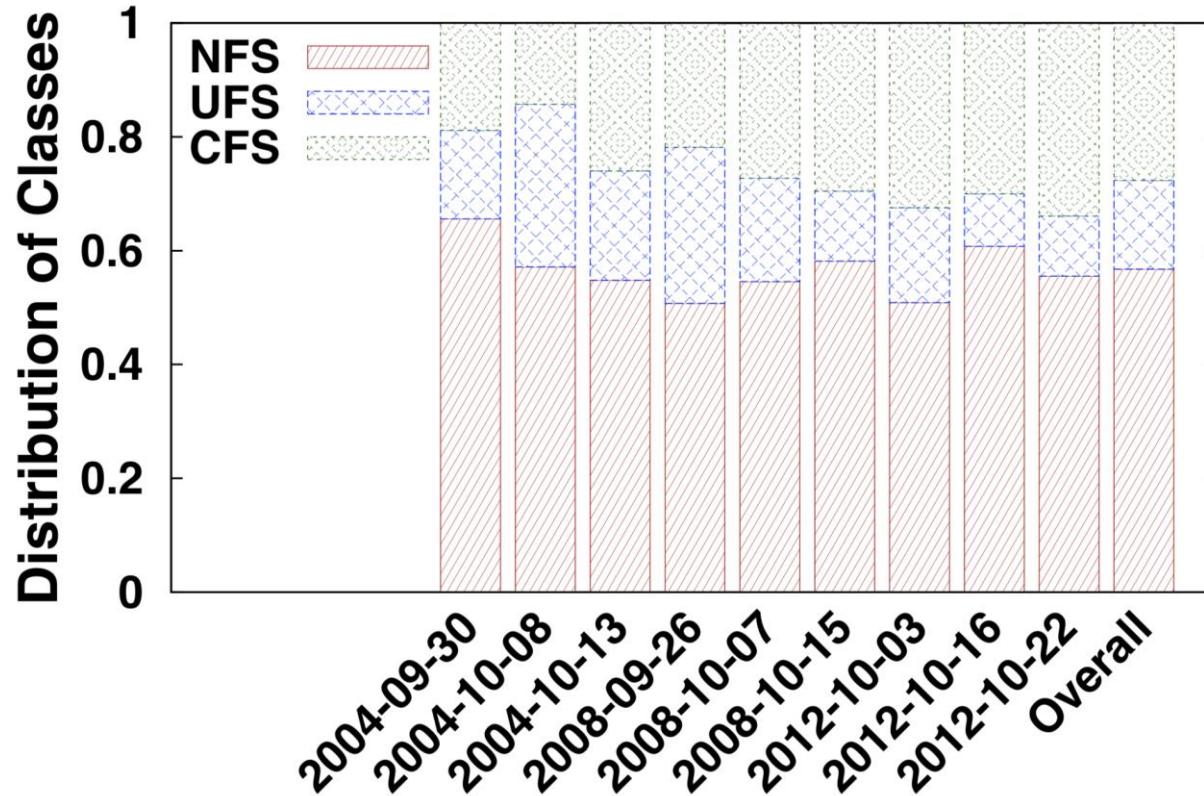
- *Next Tuesday is Election Day.*
- *Two days ago we ate lunch at a restaurant.*

The *ClaimBuster* Manual Annotation Schema

Check-worthy Factual Sentence (CFS): These sentences contain factual claims and the general public will be interested in knowing whether the claims are true or false. Journalists look for these type of claims for fact checking. Some examples are:

- *He voted against the first Gulf War.*
- *Over a million and a quarter Americans are HIV-positive.*

ClaimBuster Data Distribution



Feature-Based Classifier

Category	Type	# of Features	Example
Sentiment	continuous	1	-0.5, 0.0, 0.5
Length	discrete	1	5, 10, 50
Word (W)	continuous	6130	<i>debt, five</i>
POS Tag (P)	discrete	43	<i>VBD, CD</i>
Entity Type (ET)	discrete	26	<i>Person</i>

Donald Trump: (07:37)

How are you doing? How are you?

Kristen Welker: (07:58)

And I 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 your first question. The Debate Committee 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. We're here for 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, 10,000 Americans have died from COVID. So please be specific. How would you lead the country through 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 world. You see the spikes in Europe and many other places right now, the mortality rate is down 85%. The excess mortality is way down and much lower than almost any other country. And fighting it and we're fighting it hard. There is a spike. There was a spike, and it's now gone.

Donald Trump: (09:41)

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

Donald Trump: (10:04)

I can tell you from personal experience, I was in the hospital, got better. And I will tell you that I had something that they gave therapeutic, I guess they would call it. Some people could save, but I was in for a short period of time. And I got better we wouldn't be here tonight. And how they say I'm immune. Who months or a lifetime, nobody's been able to say that, but I'm

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 country out 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: (11:56)

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 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 responded to this crisis has been absolutely tragic. 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 quickly.

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

Donald Trump: (15:24)

And frankly, he ran 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: (16:01)

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

Donald Trump: (16:58)

Everything. 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.

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 wide range of things the president said, said, even today, we think he is control. We're about to lose 200,000 more people.

Kristen Welker: (19:43)

President Trump.

Donald Trump: (19:44)

Look, perhaps just to finish this, I was kidding on that, but just to finish this, when I closed school I 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 xenophob." 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.

Kristen Welker: (22:04)

Let's talk about schools, President Trump--

Donald Trump: (22:04)

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, all they're shut down so tight and 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 yesterday, Boston became the latest city to move its public school system entirely online after a coronavirus spike. 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?

Donald Trump: (23:03)

I want to open the schools. The transmission rate to the teachers is very small. But I want to open the schools. We have to open our country. We're doing this. We can't keep this country in a massive economy. People are suicide. There's depression, alcohol, ten before. There's abuse, tremendous. I've said it often, the cure cannot be at's what's happening. And he wants country if one person in our massive down.

to walk and chew gum at the same time, but they need resources to open? You're going to open a business, illness. You need to have. If you have a as dividers so people cannot infection where you can take testing is in fact infected. You need to be provide all the resources that are consistent with that we're going to make sure that we're going to open safely. And by the way, all you teachers out there, not that many of you are going to die, so don't worry about it. So don't worry about it. Come on.

Kristen Welker: (24:38)

President Trump, let me follow up with you quickly.

Donald Trump: (24:40)

By the way, I will say this. If you go and look at what's happened to New York, it's a ghost town. It's a ghost town. And when you talk about plexiglass, these are restaurants that had dying. These are businesses with no money. Putting up plexiglass is unaffordable, and it's not the answer. I mean, you're going to sit there in a cubicle wrapped around with plastic. These are businesses that are dying. Joe. You can't do 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 vibrant, it's dying. Everyone's leaving New York.

Kristen Welker: (26:15)

Vice President Biden.

Joe Biden: (26:15)

Take a look at what New York has done in terms of turning the curve down, in terms of the number of people dying. And I don't look at this in terms of 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 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 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

Alternative Idea (ClaimRank): Maybe the important claims are those that got fact-checked?

<https://www.rev.com/blog/transcripts/donald-joe-biden-final-presidential-debate-transcript-2020>

US Presidential Debates 2016: PolitiFact

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 [redacted] body's ever seen were a couple of years when he had to [redacted] authorities when he was trying to get a casino license, and they showed he didn't pay any federal income tax.



We rated this claim Mostly False.
<http://www.politifact.com/truth-o-meter/statements/2016/sep/27/hillary-clinton/debate-hillary-clinton-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

US Presidential Debates 2016: NPR

We need strong borders. In the audience tonight we have four mothers of, I mean these are unbelievable people that I've gotten to know over a period of years, whose children have been killed, brutally killed, by people that came into the country illegally. You have thousands of mothers and fathers and relatives all over the country. They are coming in illegally. Drugs are pouring in through the border. We have no country if we have no border. Hillary wants to give amnesty she wants to have open borders. The border secure as you know the border patrol agents sixteen thousand five hundred plus, ICE endorsed me. First time they've ever endorsed a candidate.

A union of immigration and customs enforcement officers [endorsed Trump in September](#), not the federal Immigration and Customs Enforcement bureau.

Phil Ewing

NPR National Security Editor

<https://www.npr.org/2016/10/19/498293478/fact-check-trump-and-clinton-s-final-presidential-debate>

US Presidential Debates 2016: Nine Fact-Checking Sources

Medium	1st	2nd	VP	3rd	Total
ABC News	35	50	29	28	142
Chicago Tribune	30	29	31	38	128
CNN	46	30	37	60	173
FactCheck.org	15	45	47	60	167
NPR	99	92	91	89	371
PolitiFact	74	62	60	57	253
The Guardian	27	39	54	72	192
The New York Times	26	25	46	52	149
The Washington Post	26	19	33	17	95
Total annotations	378	391	428	473	1,670
Annotated sentences	218	235	183	244	880

Agreement Between the Sources

Agreement Level	Number of Sentences	Cumulative Sum
9	1	1
8	6	7
7	5	12
6	19	31
5	26	57
4	40	97
3	100	197
2	191	388
1	492	880
Total number of sentences: 5,415		

Features

1. Sentence-level

TF.IDF bag of words

POS-tags

2. Context-level

Number of words

3. Mixed

Sentiment

Verb tense

Linguistic features

Named Entities

Number of symbols

Named Entities

You even at one time suggested that you would try to negotiate down the national debt of the **United States**.

Thousands of jobs leaving **Michigan**, leaving **Ohio**.

Sidney Blumenthal works for the campaign and close -- very close friend of **Secretary Clinton**.

Linguistic Features

Feature Name	Examples
Bias	capture, create, demand, follow
Negatives	abnormal, bankrupt, cheat, conflicts
Positives	accurate, achievements, affirm
Factives	realize, know, discover, learn
Assertives	think, believe, imagine, guarantee
Hedges	approximately, estimate, essentially
Implicatives	cause, manage, hesitate, neglect
Strong-subj	admire, afraid, agreeably, apologist
Weak-subj	abandon, adaptive, champ, consume

Features

1. Sentence-level Position
2. **Context-level** Size of the segment
3. Mixed Metadata: reaction of the public
Metadata: mention of the opponent

Metadata

Trump: **Hillary Clinton** attacked those same women and attacked them viciously.

Clinton: They're doing it to try to influence the election for **Donald Trump**.

Features

1. Sentence-level LDA
2. Context-level Word2Vec
 Discourse
3. Mixed Contradiction
 Similarity to previously fact-checked claims

Contradiction

Clinton: ... about a potential nuclear competition in Asia, you said, you know, go ahead, enjoy yourselves, folks.

Trump: **I didn't say** nuclear.

Similarity to Fact-Checked Claims

Fact-checked sentence:

"For the last seven-and-a-half years, we've seen
America's place in the **world weakened**."

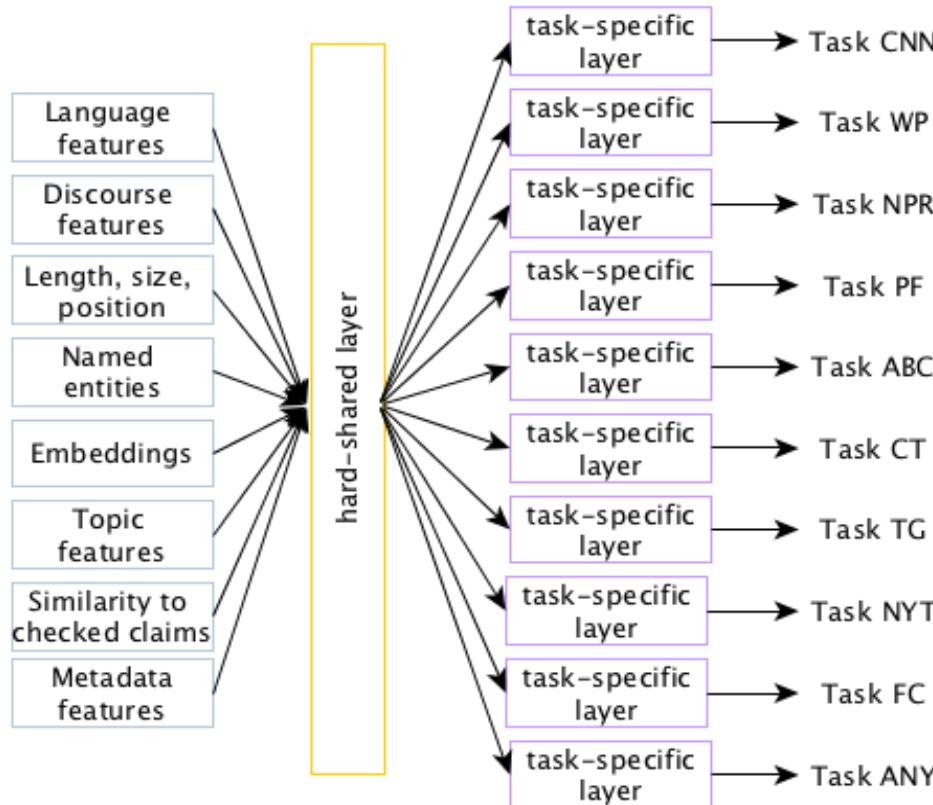
New sentence:

"We've **weakened America's place** in the
world."

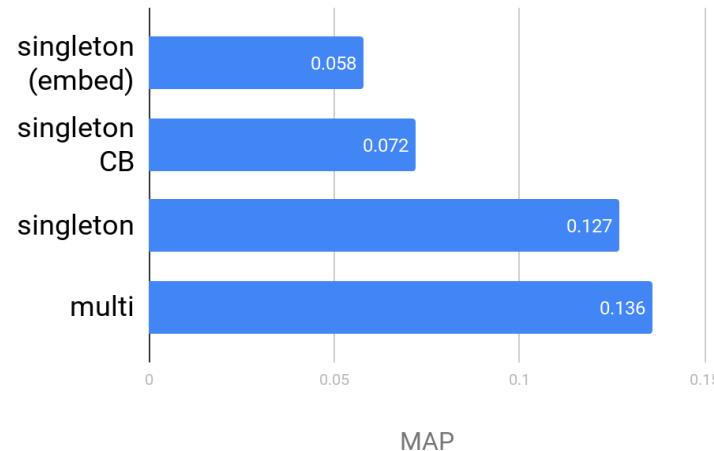
The Role of Context

System	MAP	R-Pr	P@5	P@10	P@20	P@50
All	.427	.432	.800	.725	.713	.600
All, no contextual	.385	.390	.550	.500	.550	.540
Only contextual	.317	.404	.725	.563	.465	.465
<i>CB Platform</i>	.317	.349	.500	.550	.488	.405

Multi-Task Learning



Results averaged over nine fact-checking organizations



Detecting Check-Worthy Claims

<http://claimrank.qcri.org/>

This tool identifies automatically which claims in a given document are most worthy and should be prioritized for fact-checking, e.g., by investigative journalists.



- 0.46 : We have an awful lot of vacant houses in this city.
0.05 : They're in good shape and structurally sound.
0.55 : In the last three years we've taken 3,000 vacant houses, taken them away from the homeowners who neglected them, sold them on the web site BuildingDetroit.org, and moved families back in. Three thousand families occupied.
0.15 : Senator Young says he wants to sell side lots.
0.44 : He's a little late to the party.
0.46 : We've sold 8,200 side lots to next -door neighbors who are maintaining them so they could add swing sets and slides and gardens.
0.28 : And then what we are doing is taking a strategy that works so well in Midtown, and we call it the Strategic Neighborhood Fund.
0.39 : We are starting to go to 10 neighborhoods across the city and investing all kinds of money.
0.52 : We've lined up now \$60 million of private money and this week in West Village we just opened the Coe.
0.26 : They built beautiful new apartments and retail.
0.46 : 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.
0.45 : We're about to kick off a program: we're going to renovate 50 houses around the old Holcomb school.
0.07 : We're going from neighborhood to neighborhood and then for the storefronts.
0.02 : We need to create beautiful areas where businesses want to invest.
0.49 : And I'm very proud to say yesterday Detroit City Council approved my proposal for a \$125 million bond issue.
0.46 : We're going to put \$80 million into it.
0.37 : Twenty three neighborhood quarters your shopping Court is making the beautiful flowers and walkways and the like and recreate those old shopping districts.
0.44 : That's the plan we have for the next four years.

Enter new text Mimic source: General ▾ Sort

Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Márquez, Preslav Nakov

ClaimRank: Detecting Check-Worthy Claims in Arabic and English, NAACL-HLT (Demonstrations) 2018: 26-30

Detecting Check-Worthy Claims

<http://claimrank.qcri.org/>

This tool identifies automatically which claims in a given document are most worthy and should be prioritized for fact-checking, e.g., by investigative journalists.



- 0.53: وبين الصحابا 24 لقوا حتفهم في قصف جوي على منطقة المرج في الغوطة الشرقية، بحسب ما أفاد به الدفاع المدني في الغوطة، كما قتل سبعة مدنيين آخرين بينهم أربعة أطفال في غارات جوية استهدفت مدينة عربين، إضافة لسقوط خمسة قتلى مدنيين -بينهم امرأة و طفل-. في قصف طائرات النظام بلدة مسرايا بغوطة دمشق الشرقية
- 0.5: وقال مراسل الجزيرة في سوريا إن 73 مدنياً -بينهم أطفال ونساء- قتلوا، وأصيب عشرات في قصف من طائرات النظام السوري استهدف اليوم مدن ولداب عربين وميسرايا -منطقة المرج في الغوطة الشرقية المحاصرة في ريف دمشق-
- 0.46: تواصل القصف الجوي والصاروخى المكثف والعنيف من قوات النظام السوري وروسيا على الغوطة الشرقية، وارتفع عدد القتلى خلال 24 ساعة إلى أكثر من 180 مدنياً بينهم أطفال ونساء
- 0.45: وبذلك يرتفع عدد الصحابا في الغوطة الشرقية منذ أمس إلى 183 قتيلاً ومتوفيات الجرحى
- 0.43: وكانت مستشفيات الغوطة -التي تخضع لاتفاق خفض التصعيد- بالಚابين الذين زاد عددهم عن تلاتهCentury، بينهم الكثير من الأطفال، في منطقة تعانى نقصاً حاداً في المواد
- 0.41: والمعدات الطبية جراء حصار محكم تفرضه قوات النظام عليهما منذ 2013
- 0.39: ومنذ أمس الاثنين شنّ سلاح الجو السوري والروسي عشرات الغارات على الأحياء السكنية في مدن الغوطة الشرقية ولدابها مثل سقبا وجسرين وحموربة، فضلاً عن القصف المدفعي، مما أدى إلى توقيف المستشفيات الميدانية في مدينة سقبا عن الخدمة
- 0.3: في حين أعلنت فصائل المعارضة المسلحة مقتل وجرح عدد من عناصر قوات النظام خلال صد هجوماً لهم على مواقعها في الغوطة الشرقية
- 0.26: وقد استخدمت طائرات النظام البراميل المفجرة في قصفها بلداب مسرايا وبيت سوا والزربيقية وحرزما بغوطة دمشق الشرقية
- 0.19: وبينما وصفت المعارضة السورية "الاعتداء الهمجي" على الغوطة بأنه "حرب إبادة"، طالبت الأمم المتحدة بالتوقيف الفوري عن قصف المدنيين، وحضرت من خروج الوضع الإنساني عن السيطرة
- 0.08: وتزامن التصعيد السوري والروسي مع إرسال تعزيزات عسكرية مكثفة تذر بهجوم وشك على معقل العصائل المعارضة الأخير قرب العاصمة
- 0.01: وعلى أثر هذا التصعيد، طلبت محافظة ريف دمشق والحكومة المؤقتة التابعة للمعارضة السورية من هيئة التفاوض تعليق المفاوضات بهدف إنهاء التصعيد في الغوطة الشرقية
- وأشار المراسل إلى أن تصعيد قوات النظام لم يتوقف طوال الليلة الماضية، وقد تواصل صباح اليوم:

Enter new text

Mimic source: PolitiFact

Sort

CLEF-2018 FACT CHECKING LAB

<http://alt.qcri.org/clef2018-factcheck/>

Overview

Dataset: 8,946 claims including 282 check-worthy

The current coverage of the political landscape in both the press and in social media has led to an unprecedented situation. Like never before, a statement in an interview, a press release, a blog note, or a tweet can spread almost instantaneously. This proliferation speed has left little time for double-checking claims against the facts, which has proven critical in politics, e.g., during the 2016 presidential campaign in the USA, which was influenced by fake news in social media and by false claims. Indeed, some politicians were fast to notice that when it comes to shaping public opinion, facts were secondary, and that appealing to emotions and beliefs worked better, especially in social media. It has been even proposed that this was marking the dawn of a post-truth age.

Investigative journalists and volunteers have been working hard trying to get to the root of a claim and to present solid evidence in favor or against it. Manual fact-checking is very time-consuming, and thus automatic methods have been proposed as a way to speed-up the process. For instance, there has been work on checking the factuality/credibility of a claim, of a news article, or of an information source ([Castillo et al., 2011](#); [Mukherjee and Weikum, 2015](#); [Ba et al., 2016](#); [Zubiaga et al., 2016](#); [Ma et al., 2016](#); [Popat et al., 2016](#); [Hardalov et al., 2016](#); [Karadzhov et al., 2017a,b](#); [Nakov et al., 2017](#)). However, less attention has been paid to other steps of the fact-checking pipeline shown in Figure 1, e.g., check worthiness



**CLEF-2018 Fact Checking Lab:
Automatic Identification and Verification
of Claims in Political Debates**

Organizers

- Preslav Nakov, QCRI, HBKU
- Lluís Márquez, Amazon
- Alberto Barrón-Cedeño, QCRI, HBKU
- Wajdi Zaghouani, CHSS, HBKU
- Tamer Elsayed, Qatar University
- Reem Suwaileh, Qatar University
- Pepa Gencheva, Sofia University
- Spas Kyuchukov, Sofia University
- Giovanni Da San Martino, QCRI, HBKU

CLEF-2018 CheckThat! Data

Type	Partition	Sent.	CW
Debates			
 1st Presidential	train	1,403	37
 2nd Presidential	train	1,303	25
 Vice-Presidential	train	1,358	28
 3rd Presidential	test	1,351	77
 9th Democratic	test	1,464	17
Speeches			
 Donald Trump Acceptance	test	375	21
Donald Trump at the World Economic Forum	test	245	11
Donald Trump at a Tax Reform Event	test	412	16
Donald Trump's Address to Congress	test	390	15
Donald Trump's Miami Speech	test	645	35
Total English		8,946	282
Total Arabic		7,254	205

CLEF2019 - CheckThat! Lab

Automatic Identification and Verification of Claims

<https://sites.google.com/view/clef2019-checkthat/home>

Dataset: 23,500 claims including 543 check-worthy

Welcome to CheckThat! Lab at CLEF 2019, the second version of the lab!

Recent Updates

- **Jun 14th, 2019:** Working draft of the lab overview paper is shared [here](#).
- **May 23rd, 2019:** Scorer code for Task 2 released.
- **May 8th, 2019:** Evaluation Cycle 2 of Task 2 sub-task D started.
- **Apr 30th, 2019:** Testing data released to registered participants for Task 1 and Task 2 (sub-tasks A, B, and D).
- **Apr 29th, 2019:** Training data for Task 2 sub-task C is released. Check the [Datasets](#) page.
- **Apr 23rd, 2019:** Submission guidelines and format for Task 2 have been finalized. Check the [Task 2](#) page.
- **Mar 13th, 2019:** Training data for Task 1 is released. Check the [Datasets](#) page.
- **Mar 10th, 2019:** Some training data (CT19-T2-Train-v1.0) for Task 2 (sub-tasks: A, B, and D) is released. Check the [Datasets](#) page.



CLEF-2019 CheckThat! Data

Type	Partition	Sentences	Check-worthy
Debates	Train	10,648	256
	Test	4,584	46
Speeches	Train	2,718	282
	Test	1,883	50
Press Conferences	Train	3,011	36
	Test	612	14
Posts	Train	44	8
Total	Train	16,421	433
	Test	7,079	110

CLEF2020 - CheckThat! Lab

Enabling Automatic Identification and Verification of Claims

in Social Media

Dataset: 64,290 claims including 623 check-worthy



Welcome to CheckThat! Lab at CLEF 2020, the 3rd version of the lab!

[CheckThat! 2019](#)

[CheckThat! 2018](#)



CLEF-2020 CheckThat! Data

Type	Dataset	Transcripts	Sentences	Check-worthy
Debates	Train	18	25,688	254
	Test	7	11,218	56
Speeches	Train	18	7,402	163
	Test	8	7,759	50
Interviews	Train	11	7,044	62
	Test	4	2,220	23
Town-halls	Train	3	2,642	8
	Test	1	317	7
Total	Train	50	42,776	487
	Test	20	21,514	136

Upcoming in 2021

CLEF2021-CheckThat!



CLEF2021 - CheckThat! Lab

Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News

Welcome to CheckThat! Lab at CLEF 2021, the 4th version of the lab!

[CheckThat! 2020](#)

[CheckThat! 2019](#)

[CheckThat! 2018](#)



What is Worth Fact-Checking? *(In Political TV Shows)*

Political TV shows are most discussed on Twitter



Boris Johnson on The Andrew Marr Show, which inspired 1.6 million tweets this year

JEFF OVERS/BBC/GETTY IMAGES

The FullFact Manual Annotation Schema

CLAIM

And the challenge that we have is this productivity challenge that was laid out in the Budget. S1: Can you use sort of ordinary English? I mean, the productivity challenge is hardly going to get everybody fired up, is it?

S3: What it is is this, it takes people in Germany four days to produce what people in the UK take five days to produce.

- 1 Personal experience
- 2 Quantity in the past or present
- 3 Correlation and causation (without a quantity)
- 4 Current laws or rules of operation
- 5 Prediction
- 6 Other type of claim
- 7 Not a claim

DATE: 2017-11-23 NAME: Question Time SENTENCE: What it is this, it takes people in Germany four days to produce what people in the UK take five days to produce. SENTENCE_ID: b09glrz4_201 TIME: 00:16:53.520

Lev Konstantinovskiy, Oliver Price, Mevan Babakar, Arkaitz Zubiaga:

Towards Automated Factchecking: Developing an Annotation Schema and Benchmark for Consistent Automated Claim Detection.
CoRR abs/1809.08193 (2018)

The *FullFact* Manual Annotation Schema

- **Personal experience.** Claims that aren't capable of being checked using publicly-available information, e.g. "I can't save for a deposit."
- **Quantity in the past or present.** Current value of something e.g. "1 in 4 wait longer than 6 weeks to be seen by a doctor." Changing quantity, e.g. "The Coalition Government has created 1,000 jobs for every day it's been in office." Comparison, e.g. "Free schools are outperforming state schools.". Ranking, e.g. "The UK's the largest importer from the Eurozone."
- **Correlation or causation,** Correlation e.g. "GCSEs are a better predictor than AS if a student will get a good degree." Causation, e.g. "Tetanus vaccine causes infertility." Absence of a link, e.g. "Grammar schools don't aid social mobility."
- **Current laws or rules of operation,** Declarative sentences, which generally have the word "must" or legal terms, e.g. "The UK allows a single adult to care for fewer children than other European countries." Procedures of public institutions, e.g. "Local decisions about commissioning services are now taken by organisations that are led by clinicians." Rules and changes, e.g. "EU residents cannot claim Jobseeker's Allowance if they have been in the country for 6 months and have not been able to find work."
- **Prediction,** Hypothetical claims about the future e.g. "Indeed, the IFS says that school funding will have fallen by 5% in real terms by 2019 as a result of government policies."
- **Other type of claim,** Voting records e.g "You voted to leave, didn't you?" Public Opinion e.g "Public satisfaction with the NHS in Wales is lower than it is in England." Support e.g. "The party promised free childcare" Definitions, e.g. "Illegal killing of people is what's known as murder." Any other sentence that you think is a claim.
- **Not a claim,** These are sentences that don't fall into any categories and aren't claims. e.g. "What do you think?.", "Questions to the Prime Minister!"

Category	Subcategory	Counts	Example
Other	Not a claim	54.8%	“Give it all to them, I really don’t mind.”
	Other other	10.4%*	“Molly gives so much of who she is away throughout the film.”
	Support/policy	5.5%*	“He has advocated for a junk food tax.”
	Quote	4.7%*	“The Brexit secretary said he would guarantee free movement of bankers.”
	Trivial claim	1.6%*	“It was a close call.”
	Voting record	0.7%*	“She just lost a parliamentary vote.”
	Public opinion	0.4%*	“A poll showed that most people who voted Brexit were concerned with immigration.”
	Definition	0.0%*	“The unemployed are only those actively looking for work.”
Quantity	Current value	9.9%	“1 in 4 people wait longer than 6 weeks to see a doctor.”
	Changing quantity		
	Comparison		
	Ranking		
Prediction	Hypothetical statements	4.4%	“The IFS says that school funding will have fallen by 5% by 2019.”
	Claims about the future		
Personal experience	Uncheckable	3.0%	“I can’t save for a deposit”
Correlation/causation	Correlation	2.6%	“Tetanus vaccine causes infertility”
	Causation		
	Absence of a link		
Laws/rules of operation	Public institutional procedures	1.9%	“The UK allows a single adult to care for fewer children than other European countries.”
	Rules/rule changes		

What is Worth Fact-Checking? *(In Tweets)*

Eric Trump 
6,630 Tweets



...

Follow

Eric Trump 
@EricTrump

Executive Vice President of The [@Trump](#) Organization. Husband to [@LaraLeaTrump](#).
Large advocate of [@StJude](#) Children's Hospital [@TrumpWinery](#)
[#MakeAmericaGreatAgain](#)

Eric Trump  @EricTrump · Sep 20
Thank you Williamsport, Pennsylvania!!!! 


Eric Trump  @EricTrump · Sep 19
Thank You Milford Pennsylvania: you are incredible and we love you!!! I can't wait to see you again soon!!! Let's [#KeepAmericaGreat](#) [#USAUSAUSA](#) 

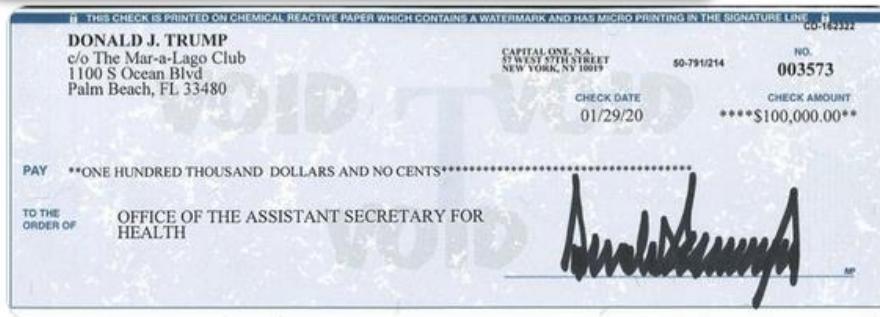

Eric Trump  @EricTrump · Sep 18
Mr. Governor - it was great to be with you today! Our family loves Maine and your incredible Lobstermen and [@realDonaldTrump](#) will never let you down! Let's [#KeepAmericaGreat](#) 


Paul R. LePage  @PaulRLePage1 · Sep 17
Thank you @EricTrump for meeting our Lobstermen at Camp Ellis in Maine today. We know that your father is working hard for these small businesses and our entire lobster industry. His efforts are paying off! [#mepolitics](#) [@realDonaldTrump](#)




Eric Trump
@EricTrump

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

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

D. Trump is donating his salary to combat COVID19?

This is
check-worthy!

What Makes a Tweet Check-Worthy?

- Q0: viral?
- Q1. contains verifiable factual claim?
- Q2. likely to be false?
- Q3. of interest to the general public?
- Q4: harmful to society, person, company, or product(s)?
- Q5. worth fact-checking?

Q1: Is this a Claim?

Verifiable Factual Claim: Does the tweet contain a verifiable factual claim?

A verifiable factual claim is a sentence claiming that something is true, and this can be verified using factual, verifiable information such as statistics, specific examples, or personal testimony.

Factual claims include the following:

- Stating a definition;
- Mentioning quantity in the present or the past;
- Making a verifiable prediction about the future;
- Reference to laws, procedures, and rules of operation;
- References to images or videos (e.g., "*This is a video showing a hospital in Spain.*");
- Statements about correlations or causations. Such correlation and causation needs to be explicit, i.e., sentences like "*This is why the beaches haven't closed in Florida. https://t.co/8x2tcQeg21*" is not a claim because it does not say why explicitly, thus it is not verifiable.

Tweets containing personal opinions and preferences are not factual claims.

Note: if a tweet is composed of multiple sentences or clauses, at least one full sentence or clause needs to be a claim in order for the tweet to contain a factual claim. If a claim exist in a sub-sentence or sub-clause then tweet is not considered to have a factual claim. For example, "*My new favorite thing is Italian mayors and regional presidents LOSING IT at people violating quarantine*" is not a claim, however, it is an opinion. Moreover, if we consider "*Italian mayors and regional presidents LOSING IT at people violating quarantine*" it would be a claim. In addition, when answering this question, annotator should not open the tweet URL.

Labels:

- **YES:** if it contains a verifiable factual claim.
- **NO:** if it does not contain any verifiable factual claim.
- **Don't know or can't not judge:** the content of the tweet does not have enough information to make a judgment. It is recommended to categorize the tweet using this label when the content of the tweet is not understandable at all. For example, it uses a language (i.e., non-English) or references that it is difficult to understand.

Examples:

1. **Tweet:** Please don't take hydroxychloroquine (Plaquenil) plus Azithromycin for #COVID19 UNLESS your doctor prescribes it. Both drugs affect the QT interval of your heart and can lead to arrhythmias and sudden death, especially if you are taking other meds or have a heart condition.
Label: YES
Explanation: There is a claim in the text.

2. **Tweet:** Saw this on Facebook today and it's a must read for all those idiots clearing the shelves #coronavirus #toiletpapercrisis #auspol
Label: NO
Explanation: There is no claim in the text.

Note: This is a YES/NO question. If you answer 'NO' to this question, you do not need to answer the questions (2-5).

Q2: Likely False?

False Information: To what extent does the tweet appear to contain false information?

The stated claim may contain false information. This question labels the tweets with the categories mentioned below. **False Information** appears on social media platforms, blogs, and news-articles to deliberately misinform or deceive the readers.

Labels:

The labels for this question are defined with five point Likert scale. A higher value means that it is more likely to be false.

1. NO, definitely contains no false information
2. NO, probably contains no false information
3. not sure
4. YES, probably contains false information
5. YES, definitely contains false information

To answer this question it is recommended to open the link of the tweet and to look for additional information for the veracity of the claim identified in question 1. For example, if the tweet contains a link to an article from a reputable information source (e.g.,

Reuters, Associated Press, France Press, Aljazeera English, BBC) then the answer could be "... contains no false info".

Examples:

1. **Tweet:** Dominican Republic found the cure for Covid-19 <https://t.co/1CfA162Lq3>

Label: 5. YES, definitely contains false information

Explanation: This is not correct information at the time of this tweet is posted.

2. **Tweet:** This is Dr. Usama Riaz. He spent past weeks screening and treating patients with Corona Virus in Pakistan. He knew there was no PPE. He persisted anyways. Today he lost his own battle with coronavirus but he gave life and hope to so many more. KNOW HIS NAME 🇵🇰❤️ <https://t.co/fISwhLCPmx>

Label: 2: NO: probably contains no false info

Explanation: This seems to be the correct information.

Q3: Interesting to the Public?

Interest to General Public: Will the tweet have an effect on or be of interest to the general public?

Most often people do not make interesting claims, which can be verified by our general knowledge. For example, "Sky is blue" is a claim, however, it is not interesting to the general public. In general, topics such as healthcare, political news and findings, and current events are of higher interest to the general public. Using the five point Likert scale the labels are defined below.

Labels:

1. NO, definitely not of interest
2. NO, probably not of interest
3. not sure
4. YES, probably of interest
5. YES, definitely of interest

Examples:

1. **Tweet:** 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. Its death rate is tiny bcos it's mass testing allows quarantine and bcos it has fewer non reported cases.

Label: YES: probably of interest

Explanation: This information relevant and of high interest for the general population as it reports how a country deals Covid-19.

2. **Tweet:** Fake news peddler Dhruv Rathee had said: ``Corona virus won't spread outside China, we need not worry'' Has this guy ever spoke something sensible? <https://t.co/siBAwIR8Pn>

Label: 2. NO, probably not of interest

Explanation: The information is interesting for the general public as it is an opinion and providing statement of someone else.

Q4: Harmful?

Harmfulness: To what extent is the tweet harmful to the society/person(s)/company(s)/product(s)?

The purpose of this question is to determine if the content of the tweet aims to and can negatively affect the society as a whole, specific person(s), company(s), product(s) or spread rumors about them. The content intends to harm or weaponize the information. A **rumor** involves a form of a statement whose veracity is not quickly or ever confirmed.

Labels: To categorize the tweets we defined the following labels based on the Likert scale. A higher value means a higher degree of harm.

1. NO, definitely not harmful
2. NO, probably not harmful
3. not sure
4. YES, probably harmful
5. YES, definitely harmful

Examples:

1. **Tweet:** How convenient but not the least bit surprising from Democrats! As usual they put politics over American citizens. @SpeakerPelosi withheld #coronavirus bill so DCCC could run ads AGAINST GOP candidates! #tcot

Label: 5. YES, definitely harmful

Explanation: This tweet is weaponized to target Nancy Pelosi and the Democrats in general.

2. **Tweet:** As we saw over the wkend, disinfo is being spread online about a supposed national lockdown and grounding flights. Be skeptical of rumors. Make sure you're getting info from legitimate sources. The @WhiteHouse is holding daily briefings and @cdcgov is providing the latest.

Label: 1. NO, definitely not harmful

Explanation: This tweet is informative and gives advice. It does not attack anyone and is not harmful.

Q5: Check-Worthy?

Need of Verification: Do you think that a professional fact-checker should verify the claim in the tweet?

It is important to verify a factual claim by a professional fact-checker, which can cause harm to the society, specific person(s), company(s), product(s) or government entities. However, not all factual claims are important or worthwhile to be fact-checked by a professional fact-checker as it is a time-consuming procedure. Therefore, the purpose is to categorize the tweet using the labels defined below. While doing so annotator can rely on the answers to the previous questions. For this question, we defined the following labels to categorize the tweets.

Labels:

- **NO, no need to check:** the tweet does not need to be fact-checked, e.g., because it is not interesting, a joke, or does not contain any claim.
- **NO, too trivial to check:** the tweet is worth fact-checking, however, this does not require a professional fact-checker, i.e., a non-expert might be able to fact-check the claim. For example, using reliable sources one can verify the information such as the official website of the WHO, etc. An example of a claim is as follows: "The GDP of the USA grew by 50% last year."
- **YES, not urgent:** the tweet should be fact-checked by a professional fact-checker, however, it is not urgent or critical;
- **YES, very urgent:** the tweet can cause immediate harm to a large number of people, therefore, it should be verified as soon as possible by a professional fact-checker.
- **not sure:** the content of the tweet does not have enough information to make a judgment.

Examples:

1. **Tweet:** Wash your hands like you've been chopping jalapeños and need to change a contact lens" says BC Public Health Officer Dr. Bonnie Henry re. ways to protect against #coronavirus #Covid_19

Label: 2. YES, not urgent

Explanation: Overall it is less important for a professional fact-checker to verify this information. The statement does not harm anyone. The truth value of whether the official said the statement is not important. Also it appears that washing hands is very important to protect oneself from the virus.

2. **Tweet:** ALERT!!!!!! The corona virus can be spread through internationaly printed albums. If you have any albums at home, put on some gloves, put all the albums in a box and put it outside the front door tonight. I'm collecting all the boxes tonight for safety.

Think of your health.

Label: NO, no need to check

Explanation: This is joke and no need to check by a professional fact checker.

CLEF2020 - CheckThat! Lab

Enabling Automatic Identification and Verification of Claims
in Social Media



Welcome to CheckThat! Lab at CLEF 2020, the 3rd version of the lab!

[CheckThat! 2019](#)

[CheckThat! 2018](#)



CLEF-2020 CheckThat! Lab, task 1 (English)

<i>Breaking: Congress prepares to shutter Capitol Hill for coronavirus, opens telework center</i>	✓
<i>China has 24 times more people than Italy...</i>	✗
<i>Everyone coming out of corona as barista</i>	✗
<i>Lord, please protect my family & the Philippines from the corona virus</i>	✗

Dataset	Total	Check-worthy
Train	672	231
Dev	150	59
Test	140	60

CLEF-2020 CheckThat! Lab, task 1 (Arabic)

العقيد رياض الأسعد
@riyadhalasaad

انهيار دفاعات ميليشيات الاحتلال الروسي الايراني المجرم على محور النيرب وحالة فوضى وهروب في صفوف مرتبة الاحتلال بعد تدمير دباباته امام ضربات الثوار ومساندة الجيش التركي والاحتلال الروسي يزج بقواته الجوية بالمعركة محاولا تغيير مجرى المعارك

Translate Tweet

6:35 PM · Feb 20, 2020 · Twitter for Android

71 Retweets and comments 314 Likes

Translation. Defenses of the criminal Russian-Iranian occupation militias on the Nairab axis have collapsed and there is a state of chaos with some occupation mercenaries fleeing after the destruction of their tanks due to rebels' strikes and the support of the Turkish army. The Russian occupation brought air forces into the battle, trying to change the course of the battles.

Multi-task Learning for Q5

English, multiclass

Setting	Acc	W-f1
Single task (Q5 only)	54.8	50.4
Multitask (Q2+Q3+Q4+Q5)	64.6	64.1

Upcoming in 2021

CLEF2021-CheckThat!



CLEF2021 - CheckThat! Lab

Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News

Welcome to CheckThat! Lab at CLEF 2021, the 4th version of the lab!

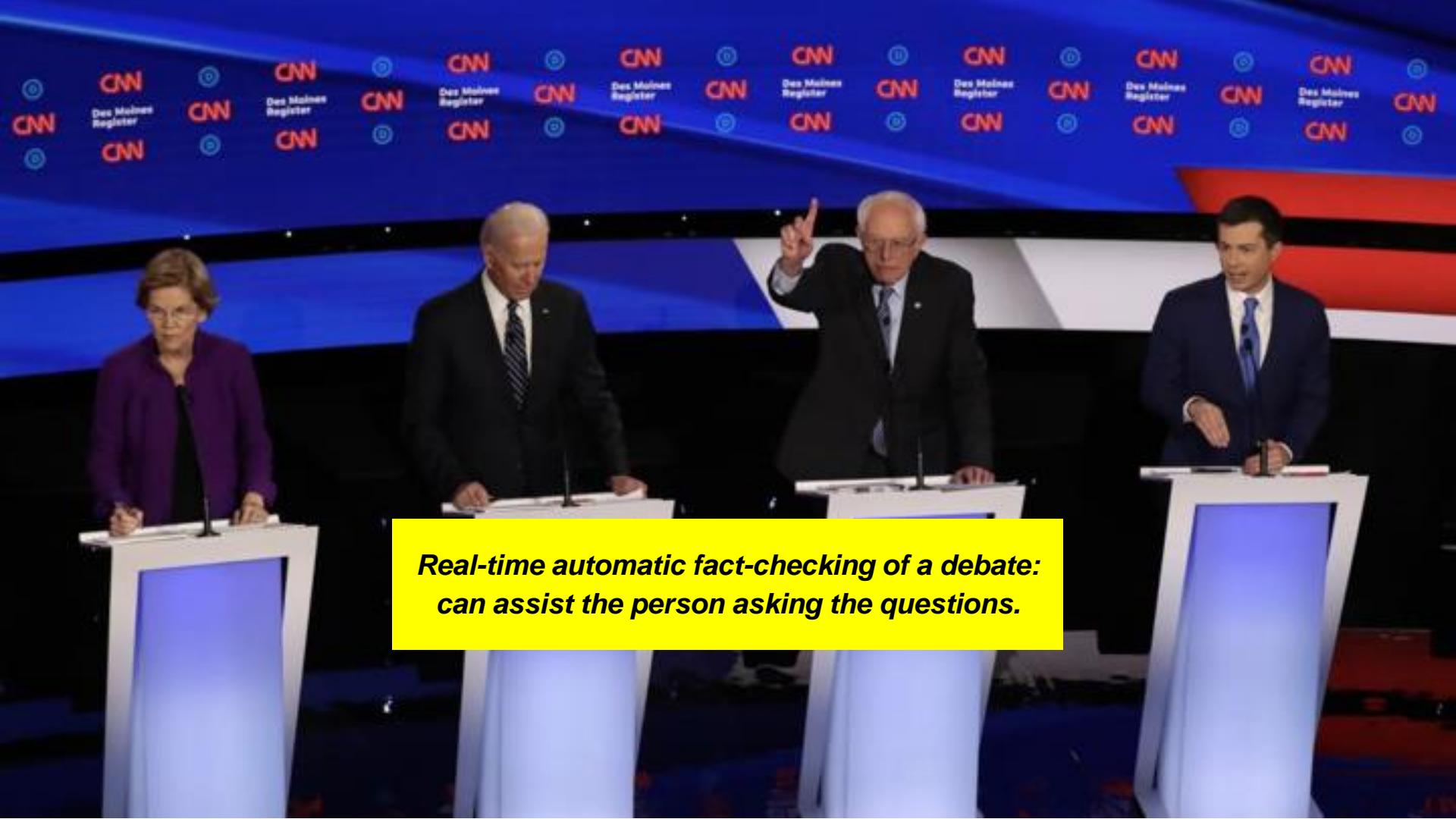
[CheckThat! 2020](#)

[CheckThat! 2019](#)

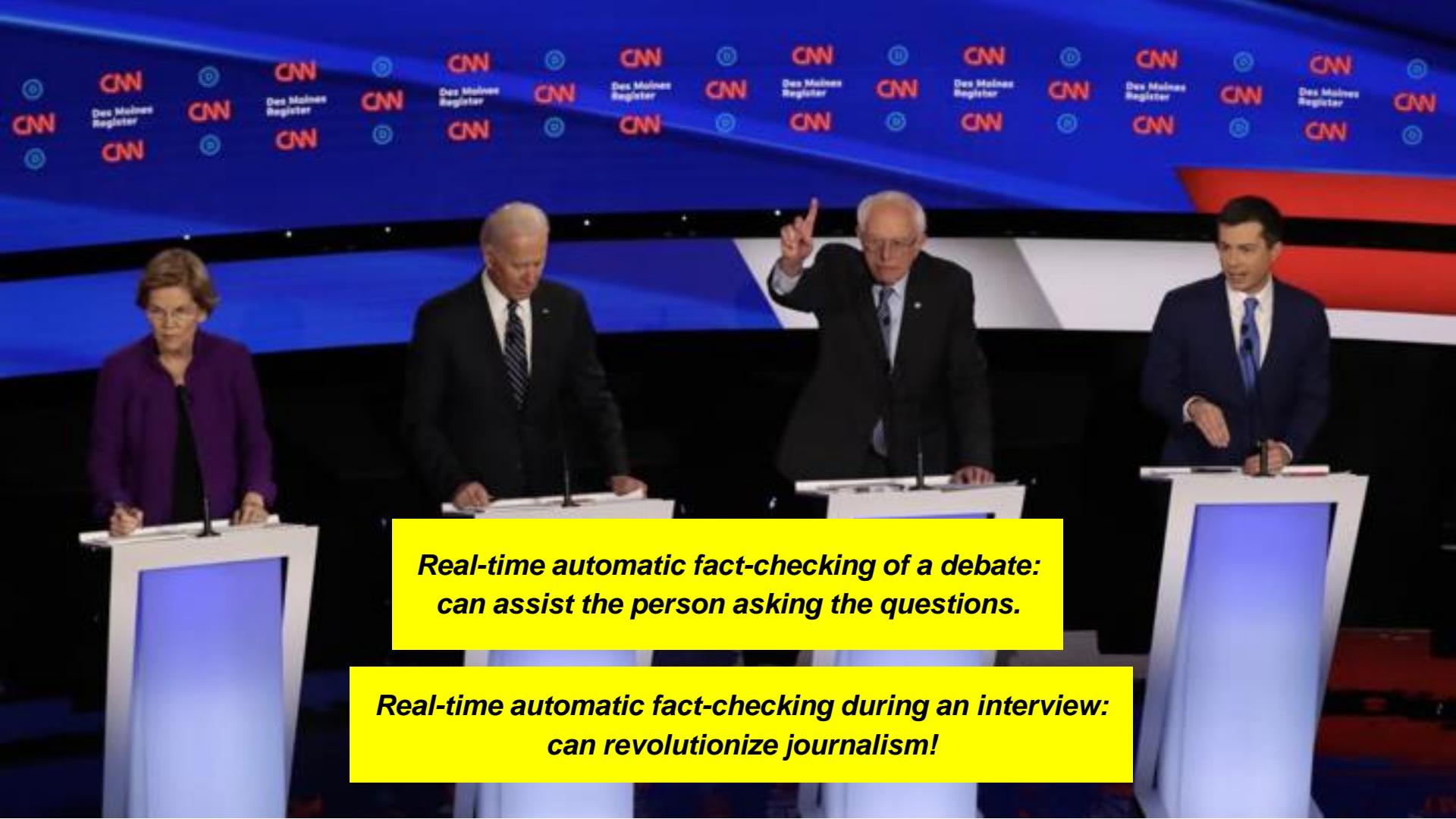
[CheckThat! 2018](#)



Detecting Previously Fact-Checked Claims



*Real-time automatic fact-checking of a debate:
can assist the person asking the questions.*



*Real-time automatic fact-checking of a debate:
can assist the person asking the questions.*

*Real-time automatic fact-checking during an interview:
can revolutionize journalism!*

Automatic Fact-Checking

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



The weird lie Donald Trump has repeated more than 80 times

15 Aug, 2019 5:00am

5 minutes to read



US President Donald Trump. Photo / AP

news.com.au

By: Sam Clench



NEWS

Home | Video | World | UK | Business | Tech | Science | Stories | Entertainment & Arts |

Politics | Parliaments | Brexit

'Strong and stable' - Why politicians keep repeating themselves

By Tom Moseley
Political reporter

27 April 2017

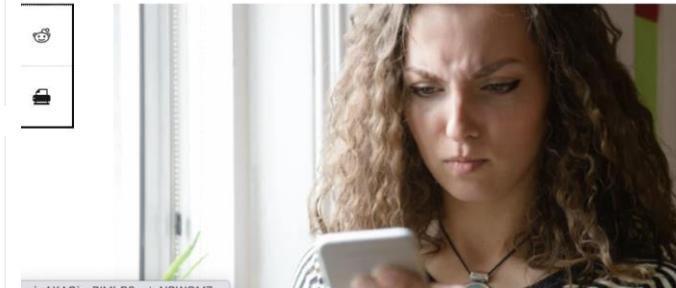


The prime minister sticks relentlessly to her "strong and stable" slogan

I Heard It Before, So It Must Be True

Repeated exposure to implausible statements makes them feel less so

By Susana Martinez-Conde on October 5, 2019  1



READ THIS NEXT

EVOLUTION

Human Speech Evolution Gets Lip Smacking Evidence

8 hours ago — Mark Stratton

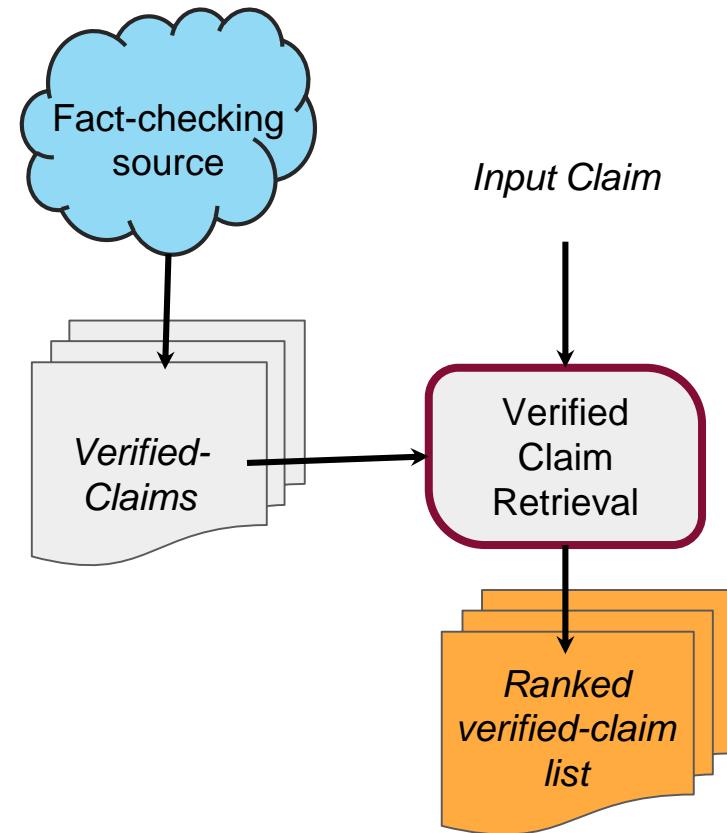
ENVIRONMENT

Carbon Levels Surge Again as Countries Emerge from Lockdown

14 hours ago — Benjamin Storrow and E&E News

Task Definition

Given a check-worthy *input* claim and a set of *verified-claims*, rank those *verified-claims*, so that the claims that can help verify the *input* claim, or a sub-claim in it, are ranked above any claim that is not helpful to verify the *input* claim.



Examples

No	<i>Input claim</i>	Manually Annotated Claim in Politifact
1	<i>Richard Nixon released tax returns when he was under audit</i>	Richard Nixon released tax returns when he was under audit
2	<i>Hillary wants to give amnesty.</i>	Says Hillary Clinton “wants to have open borders.”
3	<i>People with tremendous medical difficulty and medical problems are pouring in, and in many, in many cases, it's contagious.</i>	Says “2,267 caravan invaders have tuberculosis, HIV, chickenpox and other health issues”
4	<i>He actually advocated for the actions we took in Libya and urged that Gadhafi be taken out, after actually doing some business with him one time.</i>	Says Donald Trump is “on record extensively supporting the intervention in Libya.”
5	<i>He actually advocated for the actions we took in Libya and urged that Gadhafi be taken out, after actually doing some business with him one time.</i>	When Moammar Gadhafi was set to visit the United Nations, and no one would let him stay in New York, Trump allowed Gadhafi to set up an elaborate tent at his Westchester County (New York) estate.

This is a Non-Trivial Task!

	Politifact		Snopes	
Similarity Threshold	#	%	#	%
0.75	55	8%	11	1%
0.50	128	17%	75	8%
0.25	201	27%	504	50%

Number of Input–VerClaim pairs with TF.IDF-weighted cosine similarity above threshold

Datasets: Statistics

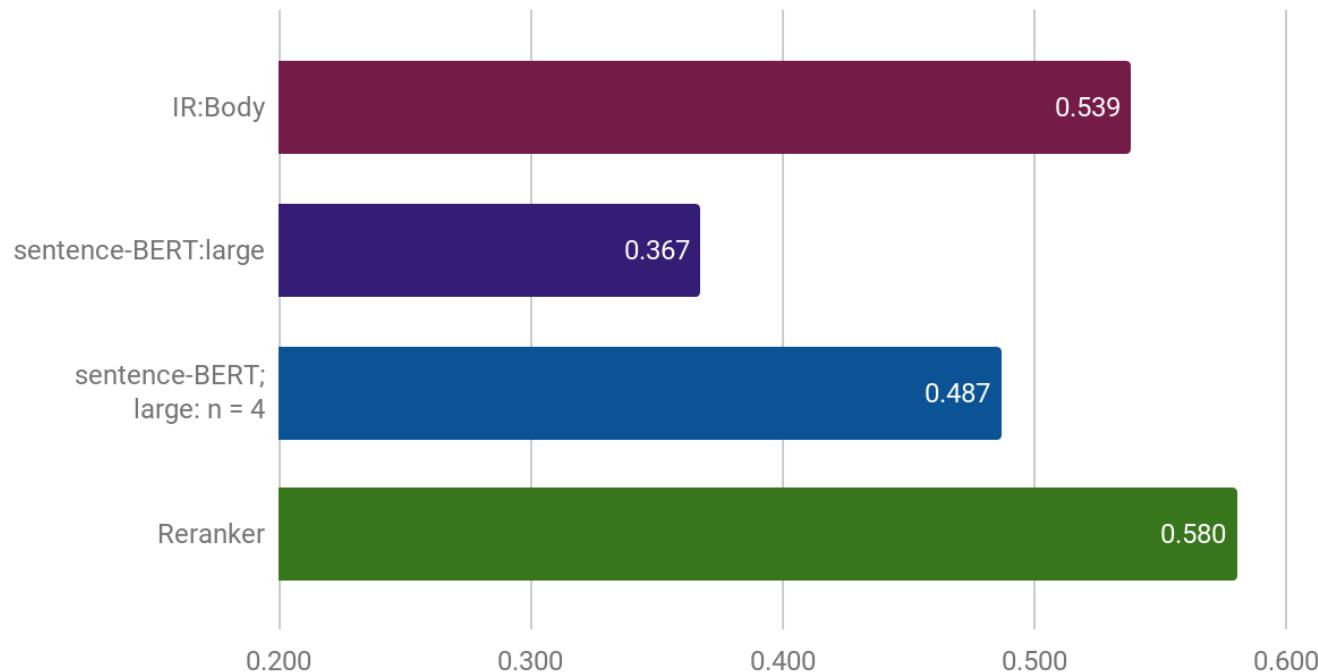
	Politifact	Snopes
<i>Input–VerClaim pairs</i>	768	1,000
– training	614	800
– testing	154	200
Total # of verified claims	16,636	10,396

Experiments:

- IR baseline: use bm25 to match the Input claim against the fact-checking article of the verified claim
- Extract scores
 - BM25: Input → *VerClaim*
 - BM25: Input → *Title of the verified claim fact-checking article*
 - SBERT: cos(Input, *VerClaim*)
 - SBERT: cos(Input, *Title*)
 - (*Politifact* only)
 - BM25: Input → *Body of the verified claim fact-checking article*
 - SBERT: cos(Input, *Body*) for the *n*-best matching sentences from *Body*
- Re-rank the top-*N* list using rankSVM with RBF kernel

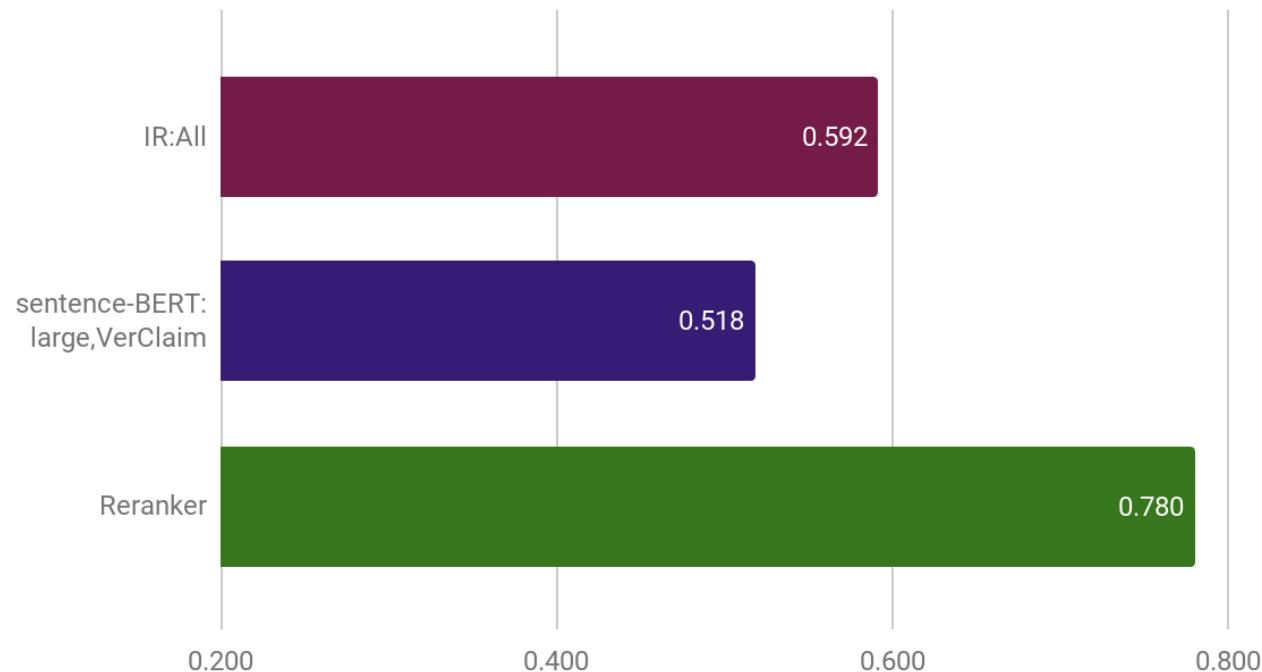
Experiments (Overall): Politifact

Overall Experiments (Politifact) - MAP@3



Experiments (Overall): Snopes

Overall Experiments (Snopes) - MAP@3

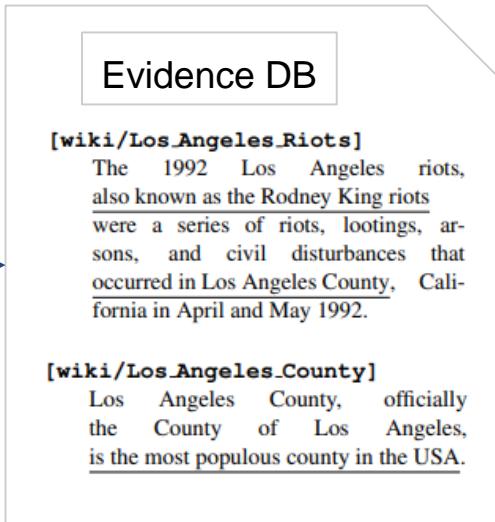


Fact-Checking

Evidence-Based Fact-Checking

Claim: The Rodney King riots took place in the most populous county in the USA.

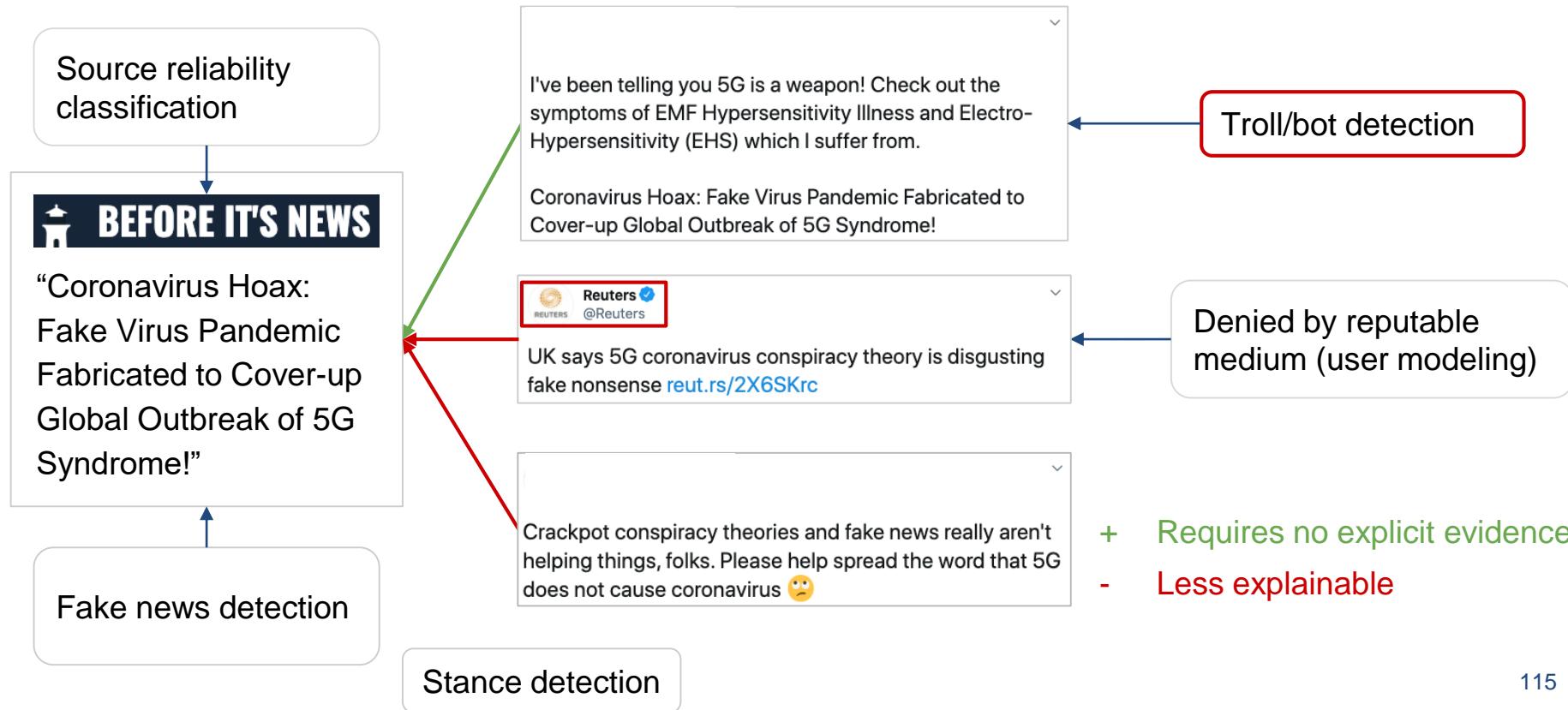
look up →



analyse → **Verdict:** Supported

- + Accurate
- + Highly explainable
- Assumes fact-checkable information
- Requires sufficient evidence

Contextual Fake News Detection



Fact-Checking a Claim Using Wikipedia

Fact-Checking

Explainable

- Fact-checked claims
- **Wikipedia pages**
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- Web
- Social media



The Third Workshop on Fact Extraction and Verification

9th July 2020 - Collocated with ACL 2020

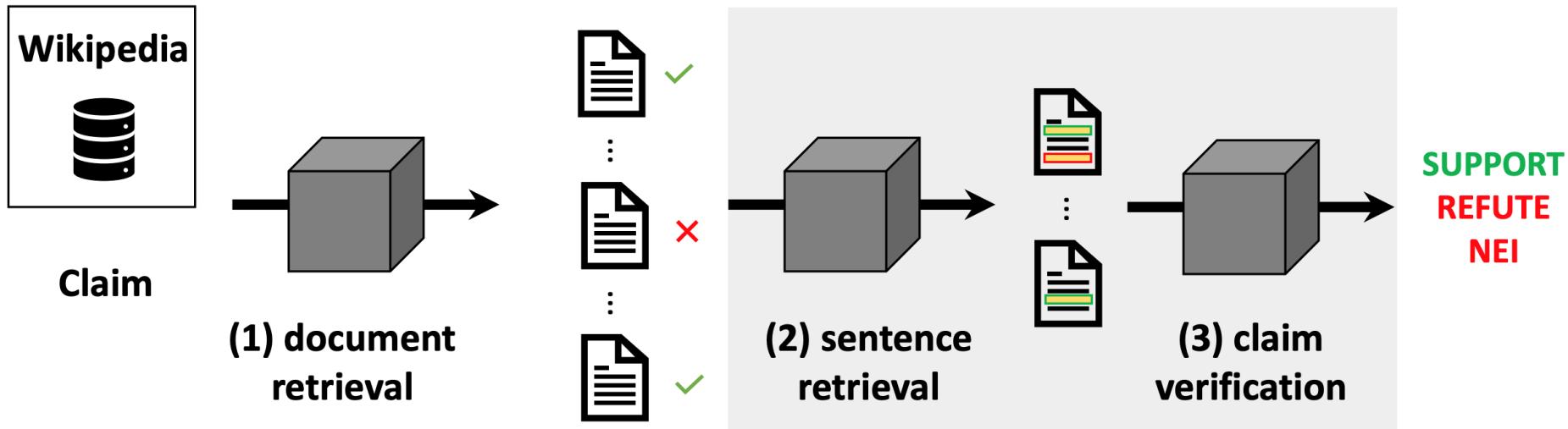
[LEARN MORE](#)

FEVER 1.0 Dataset

Split	SUPPORTED	REFUTED	NEI
Training	80,035	29,775	35,639
Dev	6,666	6,666	6,666
Test	6,666	6,666	6,666

Table 1: Dataset split sizes for SUPPORTED, REFUTED and NOTENOUGHINFO (NEI) classes

FEVER: Multi-Hop Fact-Checking in Wikipedia



FEVER: Multi-Hop Fact-Checking in Wikipedia

Claim: The Rodney King riots took place in the most populous county in the USA.

[wiki/Los_Angeles_Riots]

The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[wiki/Los_Angeles_County]

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

Verdict: Supported

Solving FEVER Using Semantic-Level Graphs

Claim: The Rodney King riots took place in the most populous county in the USA.

Fact knowledge
extracted from
evidence sentences



TRUE

Evidence #1:

The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

3

2

Evidence #2:

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

4

Solving FEVER Using Semantic-Level Graphs

Evidence #1:

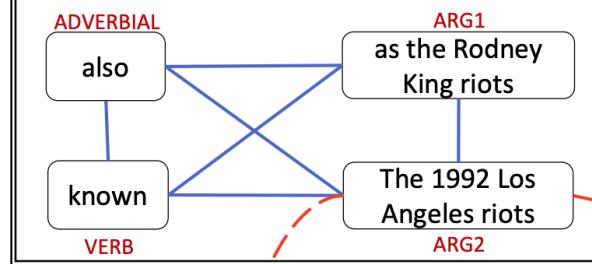
The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

Graph Construction

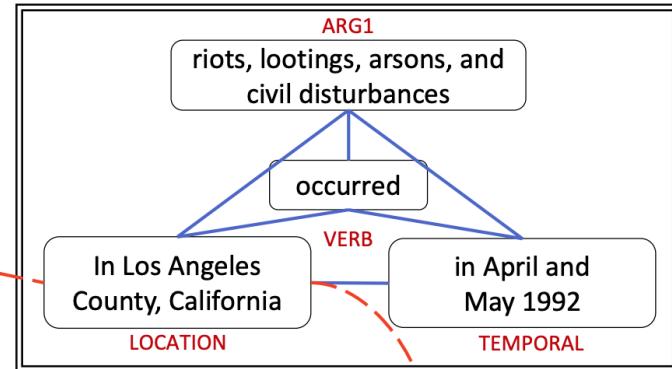
Evidence #2:

Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

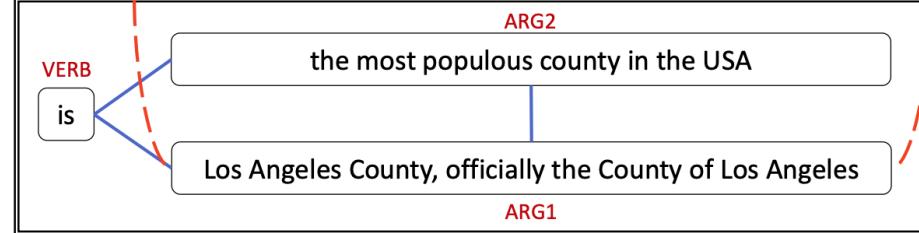
SRL results with verb "known"



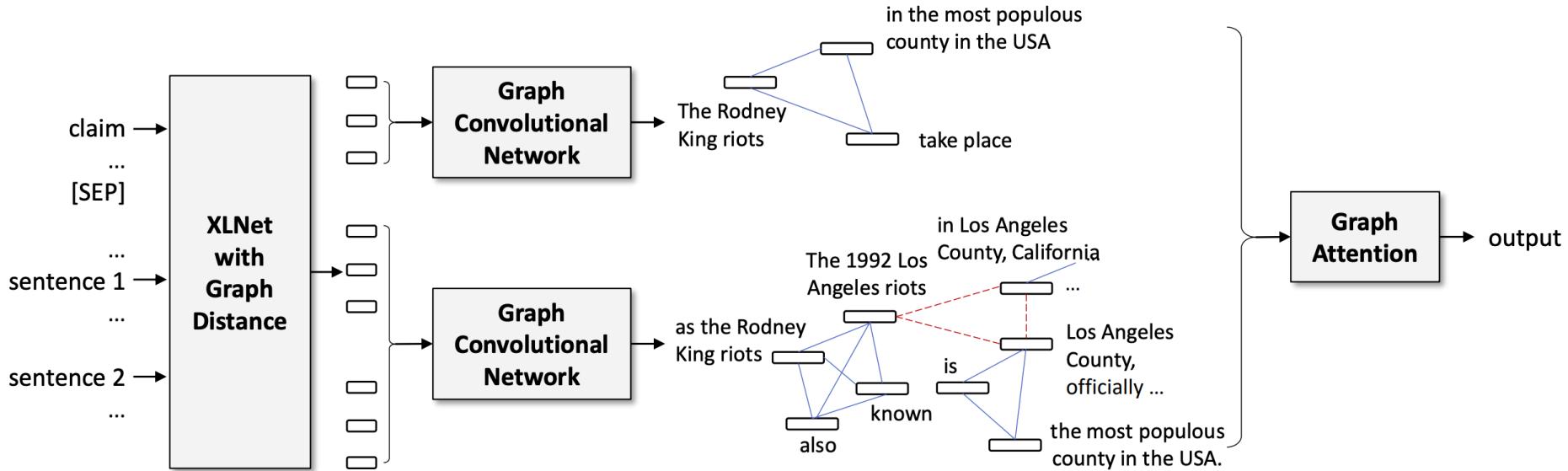
SRL results with verb "occurred"



SRL results with verb "is"



Solving FEVER Using Semantic-Level Graphs



FEVER 2.0: Build It, Break It, Fix It

- Builders
 - solve FEVER 1.0
- Breakers
 - generate adversarial examples for the builders' systems
- Fixers
 - fix the builder's systems using the examples from the breakers

Fact-Checking a Claim Using Tables

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- **Tables**
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- Web
- Social media

Fact-Checking Using Wikipedia Tables

16k Wikipedia tables

<https://tabfact.github.io/>

118k human annotated statements

United States House of Representatives Elections, 1972

District	Incumbent	Party	Result	Candidates
California 3	John E. Moss	democratic	re-elected	John E. Moss (d) 69.9% John Rakus (r) 30.1%
California 5	Phillip Burton	democratic	re-elected	Phillip Burton (d) 81.8% Edlo E. Powell (r) 18.2%
California 8	George Paul Miller	democratic	lost renomination democratic hold	Pete Stark (d) 52.9% Lew M. Warden , Jr. (r) 47.1%
California 14	Jerome R. Waldie	republican	re-elected	Jerome R. Waldie (d) 77.6% Floyd E. Sims (r) 22.4%
California 15	John J. Mcfall	republican	re-elected	John J. Mcfall (d) unopposed

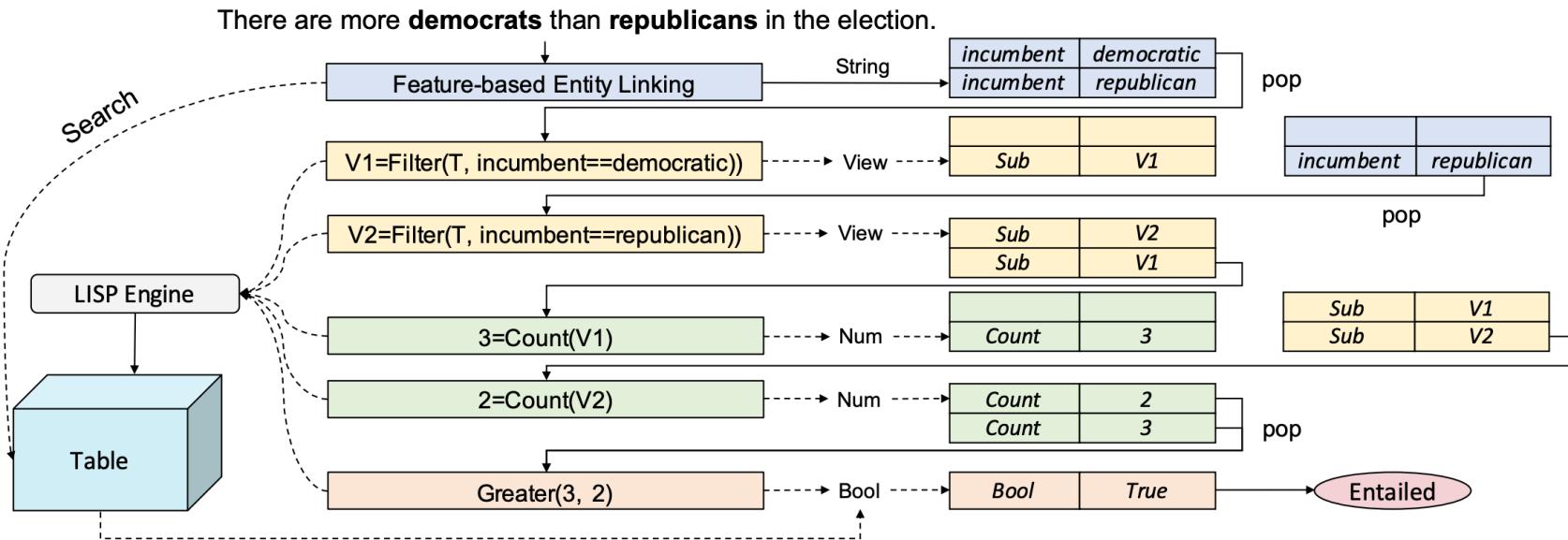
Entailed Statement

1. John E. Moss and Phillip Burton are both re-elected in the house of representative election.
2. John J. Mcfall is unopposed during the re-election.
3. There are three different incumbents from democratic.

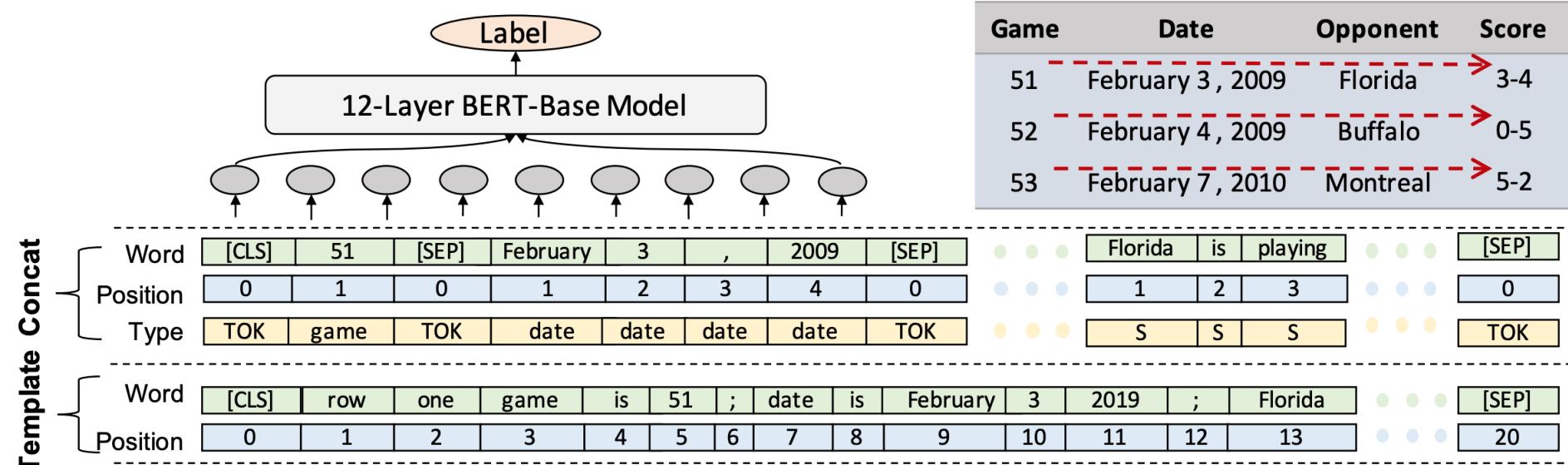
Refuted Statement

1. John E. Moss and George Paul Miller are both re-elected in the house of representative election.
2. John J. Mcfall failed to be re-elected though being unopposed.
3. There are five candidates in total, two of them are democrats and three of them are republicans.

Step 1: Select Relevant Data



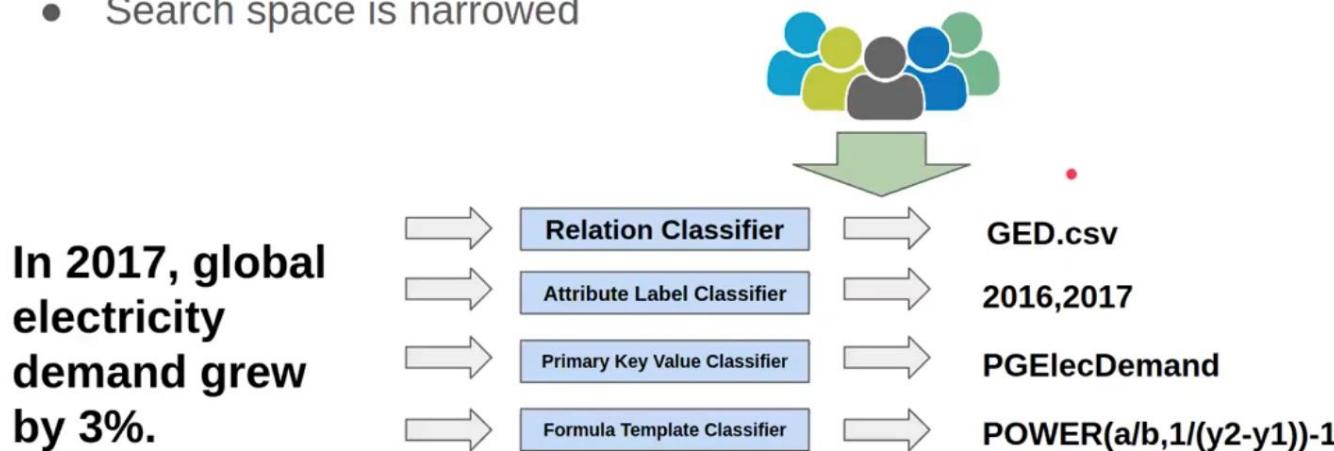
Step 2: Linearize & Apply *Table BERT*



Scrutinizer & CoronaCheck

Claim Translation

- Classifier for each query property
- Classifier accuracy is limited
- We request feedback from the users by asking questions
- Search space is narrowed



Scrutinizer & CoronaCheck

<https://coronacheck.eurecom.fr/en>

CoronaCheck: Computational Fact Checking for Statistical Coronavirus Claims

Verify statistical claims about the coronavirus spread and effects on data from the official sources.

There are 800 cases in China.



Check

The statement above is **false**, click the highlighted text for an explanation:

There are **800** cases in China.

Here are some examples:

The death rate in Italy is much higher than in France.

The total number of confirmed cases for Coronavirus in Brazil remained constant from January to February.

There are 800 cases in China.

Fact-Checking a Claim Using Knowledge Graphs

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- **Knowledge graphs**

Non-explainable

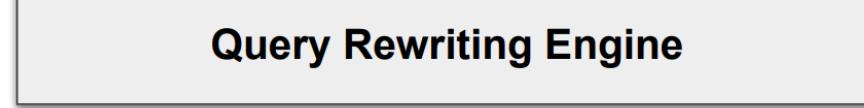
- Claim text
- Perplexity
- Language models
- Web
- Social media

Query: <Sadiq_Khan mayorOf UK> ?



Fact Spotting

Textual Web Sources



Rules

```
citizenOf(X,Y) ← mayorOf(X,Z), capitalOf(Z,Y);  
citizenOf(X,Y) ← bornIn(X,Z), isLocatedIn(Z,Y);
```

Explanations

Text: “Sadiq A. Khan is a British politician of the Labour Party serving as Mayor of London ...”

KG: $\text{capitalOf}(\text{London}, \text{UK})$



Tracing Facts in Knowledge Graphs and Text!

Choose an example...

Knowledge Graphs

Yago

Wikidata

Textual Resources

Wikipedia Trust 0.5

Web (Bing) Trust 0.5

Other Parameters

#Rules 3

#Explanations 5

↑
System Options

Input →

Query

Sadiq_Khan

isCitizenOf

United_Kingdom

Explain!

Rules

isCitizenOf(?x,?y):- hasCapital(?y,?z), mayorOf(?x,?z).
isCitizenOf(?x,?y):- wasBornIn(?x,?z), isLocatedIn(?z,?y).

Explanations

Sort by: Generation Confidence Length Depth Cost

Confidence: 0.583

Facts

[London](#) is located in [United Kingdom](#)

Source: KG

[Sadiq Khan](#) was born in [London](#)

Source: KG

Rules

isCitizenOf(?x, ?y) :- wasBornIn(?x, ?z), isLocatedIn(?z, ?y).

Cost: 13

Depth: 1

#KG accesses: 5

#Text accesses: 2

Confidence: 0.458

Facts

[Sadiq Khan](#) mayor of [London](#)

+ Source: Text

[United Kingdom](#) has capital [London](#)

Source: KG

Rules

isCitizenOf(?x, ?y) :- hasCapital(?y, ?z), mayorOf(?x, ?z).

Cost: 9

Depth: 1

#KG accesses: 2

#Text accesses: 2

Output →

Facts

Sadiq Khan mayor of London

+ Source: Text

Spotted Evidence:

- Khan was elected Mayor of London in the 2016 mayoral election, succeeding Conservative Party Mayor Boris Johnson. (source: [Sadiq Khan](#))
- Khan was elected Mayor of London in May 2016. (source: [Tooting](#))
- It consists of a directly elected executive Mayor of London, currently Sadiq Khan... (source: [Greater London Authority](#))
- Greater London, has a degree of devolved power vested in an elected Mayor of London, currently Sadiq Khan... (source: [Governance of England](#))
- The 2016 London mayoral election was won by Sadiq Khan, a member of the Labour Party. (source: [Deputy Mayor of London](#))

United Kingdom has capital London

Source: KG

Rules

$$\text{isCitizenOf}(\text{?x}, \text{?y}) :- \text{hasCapital}(\text{?y}, \text{?z}), \text{mayorOf}(\text{?x}, \text{?z}).$$

Cost: 9

Depth: 1

#KG accesses: 2

#Text accesses: 2

Fact-Checking a Claim Using Its Text Only

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- **Claim text**
- Perplexity
- Language models
- Web
- Social media

“Liar, Liar Pants on Fire” Dataset

- 12,836 claims from PolitiFact

Dataset Statistics

Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9

Top-3 Speaker Affiliations

Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

“Liar, Liar Pants on Fire” Dataset Example

Statement: *“Under the health care law, everybody will have lower rates, better quality care and better access.”*

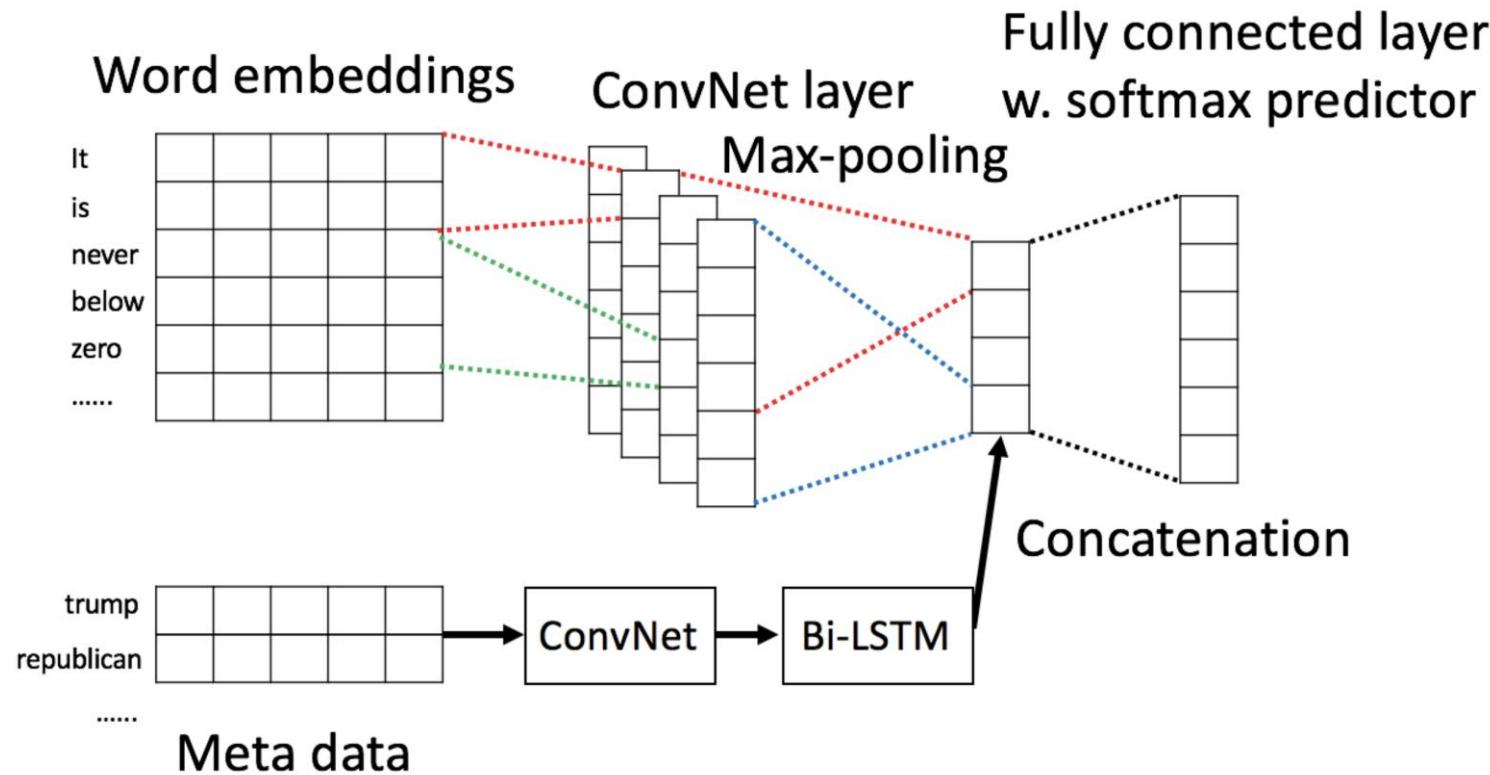
Speaker: Nancy Pelosi

Context: on ‘Meet the Press’

Label: False

Justification: Even the study that Pelosi’s staff cited as the source of that statement suggested that some people would pay more for health insurance. Analysis at the state level found the same thing. The general understanding of the word “everybody” is every person. The predictions don’t back that up. We rule this statement False.

“Liar, Liar Pants on Fire”: CNN-LSTM Model



Fact-Checking a Claim Using Perplexity

Fact-Checking

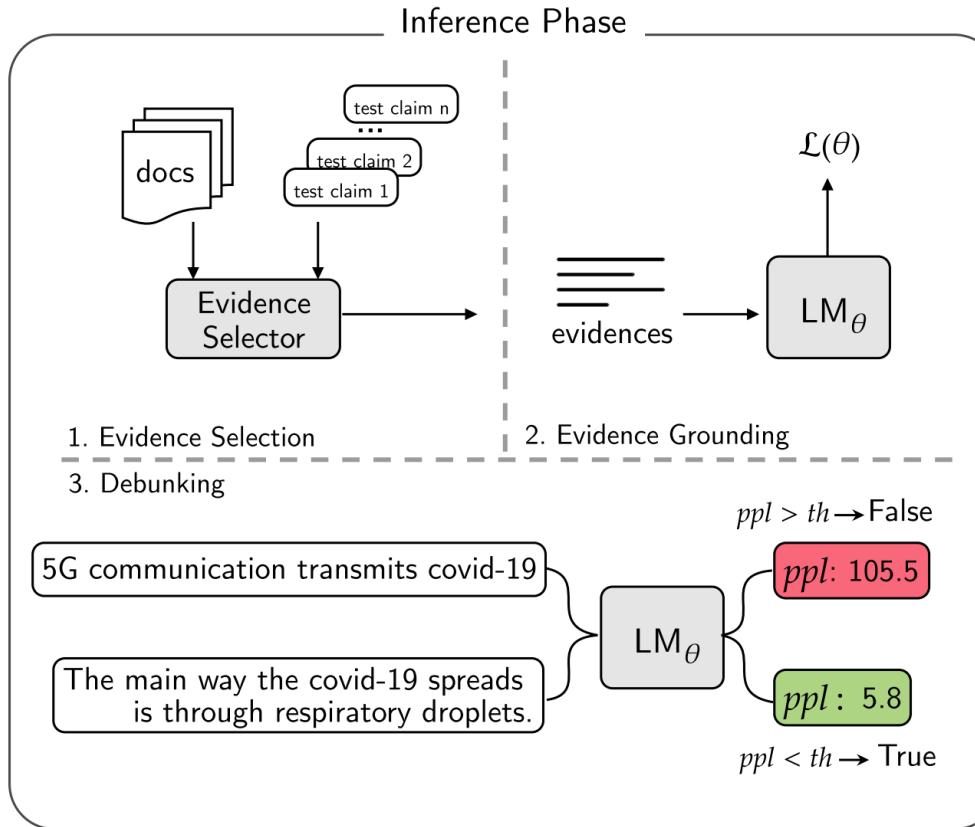
Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- **Perplexity**
- Language models
- Web
- Social media

Fact-Checking Using Perplexity



Fact-Checking Using Perplexity

False Claims

Perplexity

Ordering or buying products shipped from overseas will make a person get COVID-19.	556.2
Sunlight actually can kill the novel COVID-19.	385.0
5G helps COVID-19 spread.	178.2
Home remedies can cure or prevent the COVID-19.	146.2

True Claims

Perplexity

The main way the COVID-19 spreads is through respiratory droplets.	5.8
The most common symptoms of COVID-19 are fever, tiredness, and dry cough.	6.0
The source of SARS-CoV-2, the coronavirus (CoV) causing COVID-19 is unknown.	8.1
Currently, there is no vaccine and no specific antiviral medicine to prevent or treat COVID-19.	8.4

Fact-Checking a Claim Using Language Models

Fact-Checking

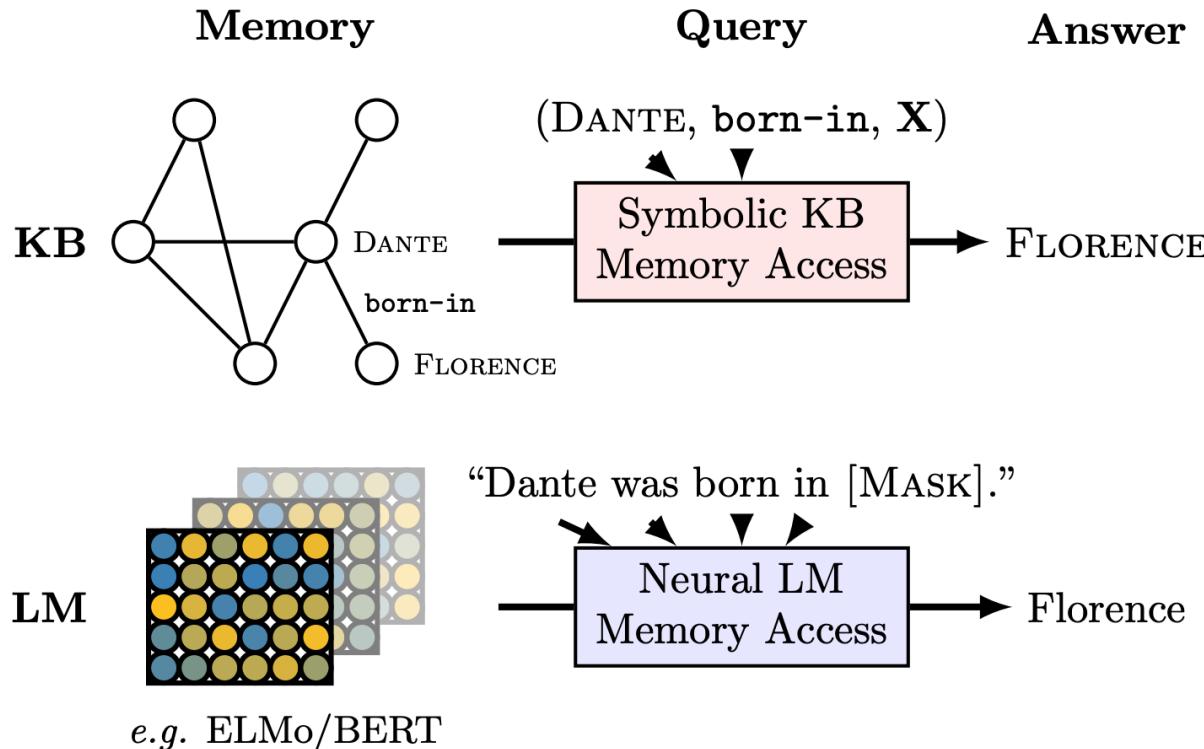
Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

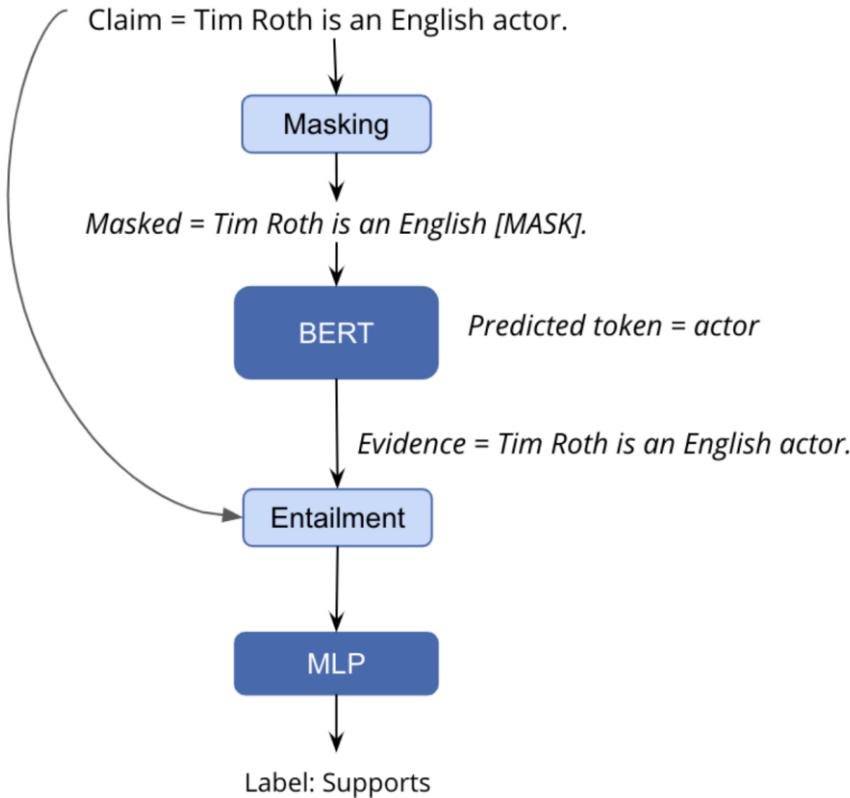
Non-explainable

- Claim text
- Perplexity
- **Language models**
- Web
- Social media

LMs as Knowledge Bases



LMs as Fact-Checkers



Fact-Checking a Claim Using the Web

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- **Web**
- Social media

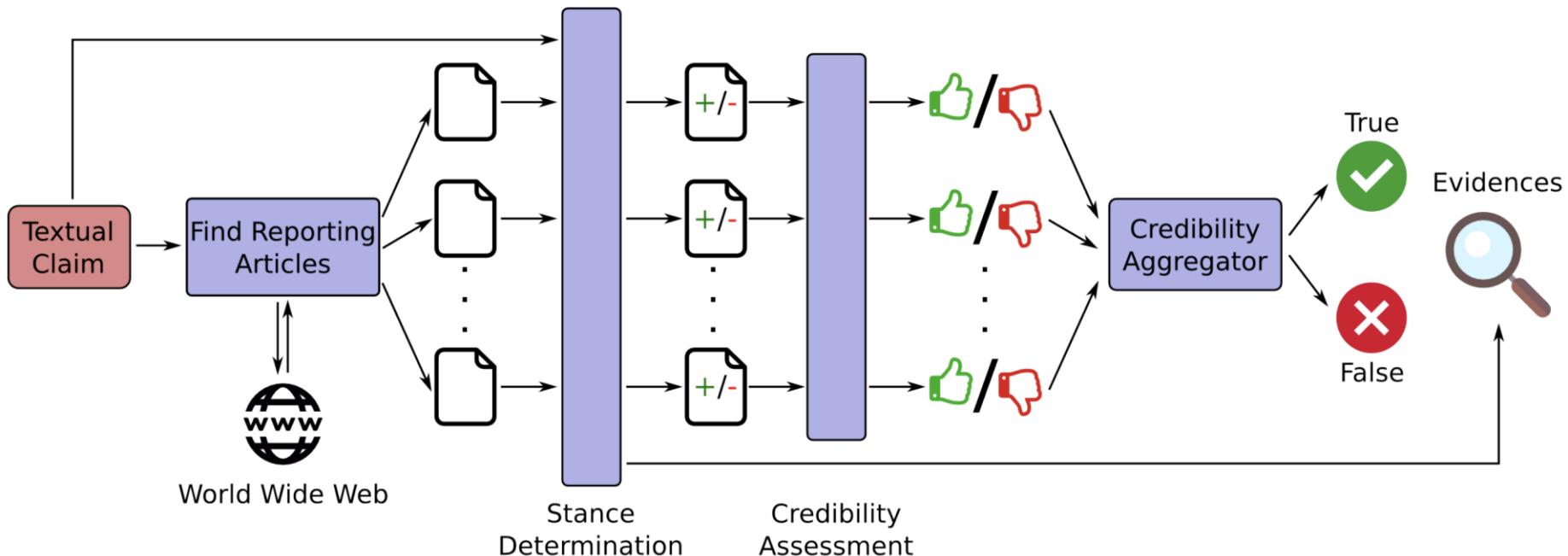
Snopes Dataset



Total Claims	4856
<i>True</i> claims	1277 (26.3%)
<i>False</i> claims	3579 (73.7%)
Web articles	133272
Relevant articles	80421
Relevant web-sources	23260

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media.
WWW (Companion Volume) 2017: 1003-1012

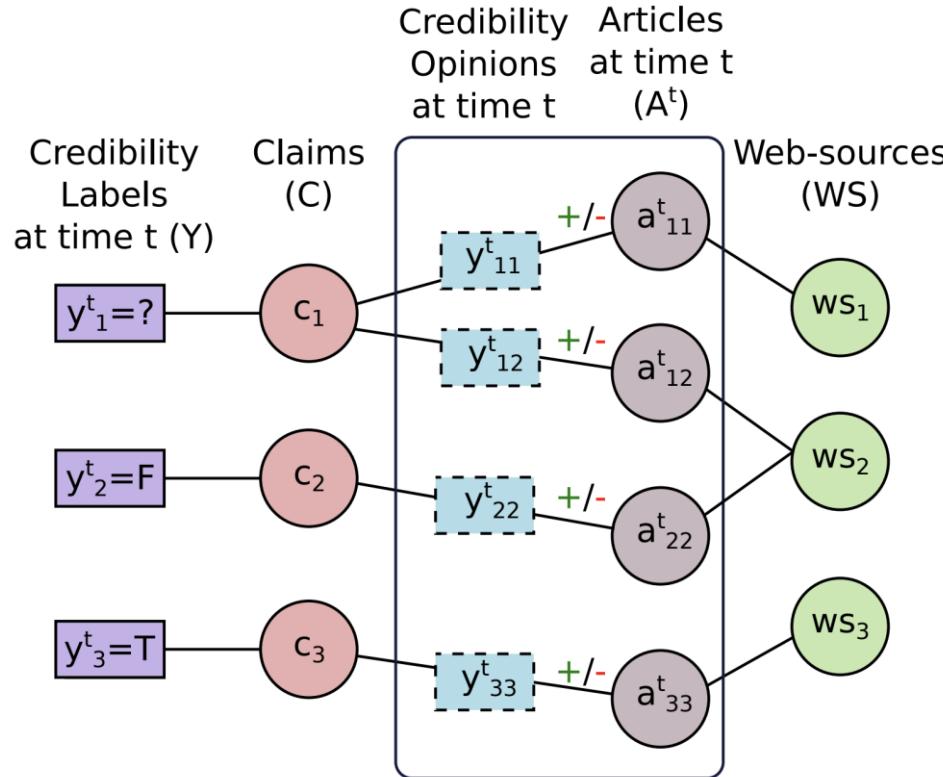
Credibility of Emerging Claims on the Web



Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:

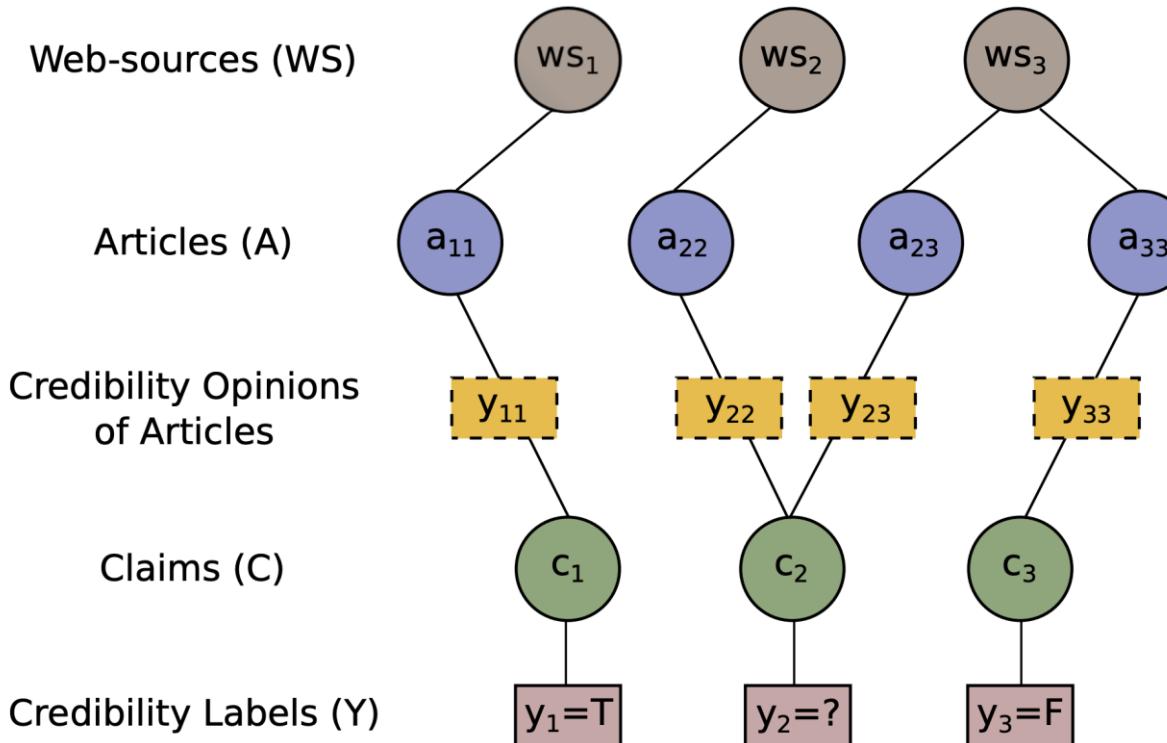
Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media.
WWW (Companion Volume) 2017: 1003-1012

Credibility of Emerging Claims on the Web



Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media.
WWW (Companion Volume) 2017: 1003-1012

Credibility of Emerging Claims on the Web



Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media.
WWW (Companion Volume) 2017: 1003-1012

Fact-Checking for Community Question Answering

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- **Web**
- Social media



SemEval 2019 Task 8: Fact Checking in Community Question Answering Forums

Organized by tsvm - Current server time: Aug. 29, 2019, 12:09 p.m. UTC

Previous

► Current

End

Evaluation

Post-Evaluation

Competition Ends

Jan. 10, 2019, midnight UTC

Feb. 1, 2019, midnight UTC

Never

Learn the Details

Phases

Participate

Results

<http://competitions.codalab.org/competitions/20022>

Overview

Evaluation

Terms and Conditions

SemEval 2019 Task 8 on Fact-Checking in Community Forums

This is the website of the SemEval-2019 task on fact checking in community question answering forums. The task features two subtasks.

Subtask A: decide whether a question asks for a factual information, an opinion/advice or is just socializing. **Subtask B:** decide whether an answer to a factual question is true, false, or does not constitute a proper answer.

To illustrate the problem, we are giving an example from the "Qatar Living" forum:

=====

- *Q:* I have heard its not possible to extend visit visa more than 6 months? Can U please answer me..
Thankzzz...
- *a1:* Maximum period is 9 Months....
- *a2:* 6 months maximum
- *a3:* This has been answered in QL so many times. Please do search for information regarding this. BTW answer is 6 months.

=====

SemEval-2019 Task 8, Subtask A: Is the Question Factual?

Q: *“What is the penalty for jumping on the red light?”*

Factual

Q: *“Can anyone recommend a good Vet in Doha?”*

Opinion

Q: *“What was your first car?”*

Socializing

SemEval-2019 Task 8, Subtask B: Is It True?

False!

Q: “I HAVE HEARD ITS NOT POSSIBLE TO EXTEND VISIT VISA MORE THAN 6 MONTHS? CAN U PLEASE ANSWER ME.. THANKZZZ...”

a₁: “Maximum period is 9 Months....”

True!

a₂: “6 months maximum”

Nonfactual!

a₃: “This has been answered in QL so many times. Please do search for information regarding this. BTW answer is 6 months.”

Modeling Factuality

1. **answer content** (*what is said and how*)
2. **author profile** (*who says it*)
3. **the forum** (*where it is said*)
4. **the entire Web** (*external support*)

Fact-Checking Claims About Images

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- **Web**
- Social media

Fact-Checking Claims About Images: Example



Image shows a Tesla vehicle in space.

Real Image
True Claim

Fact-Checking Claims About Images: Example



Image shows President Vladimir Putin aggressively pulling on U.S. President Barack Obama's tie.

Fake Image
False Claim

Fact-Checking Claims About Images: Example



A photograph shows a palm-sized rabbit.

Real Image
False Claim

Fact-Checking Claims About Images: Dataset

- ▶ Snopes.com is arguably the oldest and the largest fact-checking Web site online.
- ▶ It has a **Fauxtography** section for fact-checking images.
- ▶ We scraped 838 examples: 197 True and 641 False.

image		
claim	A photograph shows Barack Obama dressed as Lucifer for an 'Illuminati ritual.'	Photographs show a "mummified" dinosaur fossil discovered in Alberta, Canada.
label	False	True

Fact-Checking Claims About Images: Dataset

- ▶ **Reuters.com** is the world's largest news provider.
- ▶ *Pictures of the Year*: about 100 photos with a short description published at the end of each year.
- ▶ We scraped **395 examples** from the years 2015–2018.

image		
claim	People observe a minute of silence at the Trocadero in front the Eiffel Tower to pay tribute to the victims of the series of deadly attacks.	Lava erupts from a fissure east of the Leilani Estates subdivision during ongoing eruptions of the Kilauea Volcano.
label	True	True

Fact-Checking Claims About Images

- image:
 - was it manipulated?
- reverse image search
 - does the image appear on good/bad websites?
 - do the pages support the claim?

Fact-Checking Rumors in Social Media

Fact-Checking

Explainable

- Fact-checked claims
- Wikipedia pages
- Tables
- Knowledge graphs

Non-explainable

- Claim text
- Perplexity
- Language models
- Web
- **Social media**

Information Credibility on Twitter

- **Modeling 4 information sources**
 - Message
 - User
 - Topic
 - Propagation

Carlos Castillo, Marcelo Mendoza, Barbara Poblete:
Information credibility on Twitter. WWW 2011: 675-684

Information Credibility on Twitter: Message

Scope	Feature	Description
Msg.	LENGTH CHARACTERS LENGTH WORDS CONTAINS QUESTION MARK CONTAINS EXCLAMATION MARK CONTAINS MULTI QUEST OR EXCL. CONTAINS EMOTICON SMILE CONTAINS EMOTICON FROWN CONTAINS PRONOUN FIRST SECOND THIRD COUNT UPPERCASE LETTERS NUMBER OF URLs CONTAINS POPULAR DOMAIN TOP 100 CONTAINS POPULAR DOMAIN TOP 1000 CONTAINS POPULAR DOMAIN TOP 10000 CONTAINS USER MENTION CONTAINS HASHTAG CONTAINS STOCK SYMBOL IS RETWEET DAY WEEKDAY SENTIMENT POSITIVE WORDS SENTIMENT NEGATIVE WORDS SENTIMENT SCORE	Length of the text of the tweet, in characters ... in number of words Contains a question mark '?' ... an exclamation mark '!' ... multiple question or exclamation marks ... a "smiling" emoticon e.g. :-) ;-) a "frowning" emoticon e.g. :-(;-(< a personal pronoun in 1st, 2nd, or 3rd person. (3 features) Fraction of capital letters in the tweet Number of URLs contained on a tweet Contains a URL whose domain is one of the 100 most popular ones ... one of the 1,000 most popular ones ... one of the 10,000 most popular ones Mentions a user: e.g. @cnnbrk Includes a hashtag: e.g. #followfriday ... a stock symbol: e.g. \$APPL Is a re-tweet: contains 'RT' The day of the week in which this tweet was written The number of positive words in the text ... negative words in the text Sum of ± 0.5 for weak positive/negative words, ± 1.0 for strong ones

Information Credibility on Twitter: *User*

User	REGISTRATION AGE STATUSES COUNT COUNT FOLLOWERS COUNT FRIENDS IS VERIFIED HAS DESCRIPTION HAS URL	The time passed since the author registered his/her account, in days The number of tweets at posting time Number of people following this author at posting time Number of people this author is following at posting time 1.0 iff the author has a 'verified' account ... a non-empty 'bio' at posting time ... a non-empty homepage URL at posting time
------	---	---

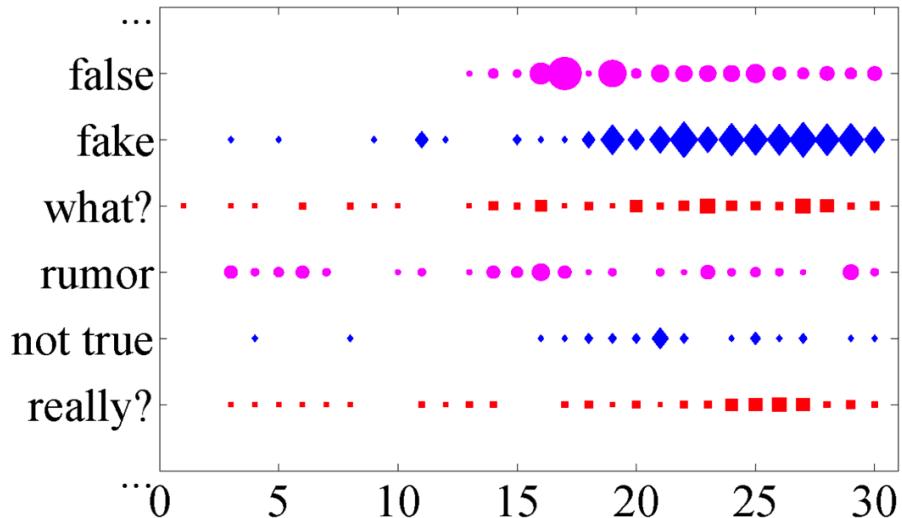
Information Credibility on Twitter: Topic

Topic	COUNT TWEETS AVERAGE LENGTH FRACTION TWEETS QUESTION MARK FRACTION TWEETS EXCLAMATION MARK FRACTION TWEETS MULTI QUEST OR EXCL. FRACTION TWEETS EMOTICON SMILE FROWN CONTAINS PRONOUN FIRST SECOND THIRD FRACTION TWEETS 30PCT UPPERCASE FRACTION TWEETS URL FRACTION TWEETS USER MENTION FRACTION TWEETS HASHTAG FRACTION TWEETS STOCK SYMBOL FRACTION RETWEETS AVERAGE SENTIMENT SCORE FRACTION SENTIMENT POSITIVE FRACTION SENTIMENT NEGATIVE FRACTION POPULAR DOMAIN TOP 100 FRACTION POPULAR DOMAIN TOP 1000 FRACTION POPULAR DOMAIN TOP 10000 COUNT DISTINCT EXPANDED URLs SHARE MOST FREQUENT EXPANDED URL COUNT DISTINCT SEEMINGLY SHORTENED URLs COUNT DISTINCT HASHTAGS SHARE MOST FREQUENT HASHTAG COUNT DISTINCT USERS MENTIONED SHARE MOST FREQUENT USER MENTIONED COUNT DISTINCT AUTHORS SHARE MOST FREQUENT AUTHOR AUTHOR AVERAGE REGISTRATION AGE AUTHOR AVERAGE STATUSES COUNT AUTHOR AVERAGE COUNT FOLLOWERS AUTHOR AVERAGE COUNT FRIENDS AUTHOR FRACTION IS VERIFIED AUTHOR FRACTION HAS DESCRIPTION AUTHOR FRACTION HAS URL	Number of tweets Average length of a tweet The fraction of tweets containing a question mark '?' ... an exclamation mark '!' ... multiple question or exclamation marks ... emoticons smiling or frowning (2 features) ... a personal pronoun in 1st, 2nd, or 3rd person. (3 features) ... more than 30\% of characters in uppercase The fraction of tweets containing a URL ... user mentions ... hashtags ... stock symbols The fraction of tweets that are re-tweets The average sentiment score of tweets The fraction of tweets with a positive score ... with a negative score The fraction of tweets with a URL in one of the top-100 domains ... in one of the top-1,000 domains ... in one of the top-10,000 domains The number of distinct URLs found after expanding short URLs The fraction of occurrences of the most frequent expanded URL The number of distinct short URLs The number of distinct hashtags The fraction of occurrences of the most frequent hashtag The number of distinct users mentioned in the tweets The fraction of user mentions of the most frequently mentioned user The number of distinct authors of tweets The fraction of tweets authored by the most frequent author The average of AUTHOR REGISTRATION AGE The average of AUTHOR STATUSES COUNT ... of AUTHOR COUNT FOLLOWERS ... of AUTHOR COUNT FRIENDS The fraction of tweets from verified authors ... from authors with a description ... from authors with a homepage URL
-------	---	---

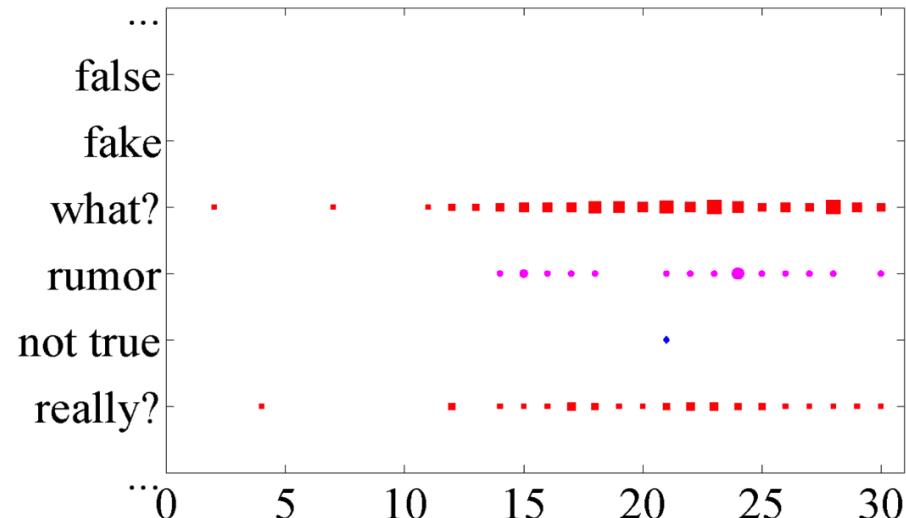
Information Credibility on Twitter: *Propagation*

Prop.	PROPAGATION INITIAL TWEETS PROPAGATION MAX SUBTREE PROPAGATION MAX AVG DEGREE PROPAGATION MAX AVG DEPTH PROPAGATION MAX LEVEL	The degree of the root in a propagation tree The total number of tweets in the largest sub-tree of the root, plus one The maximum and average degree of a node that is not the root (2 feat.) The depth of a propagation tree (0=empty tree, 1=only initial tweets, 2=only re-tweets of the root) and its per-node average (2 features) The max. size of a level in the propagation tree (except children of root)
-------	---	--

Rumor Detection in Social Media: Timing Matters



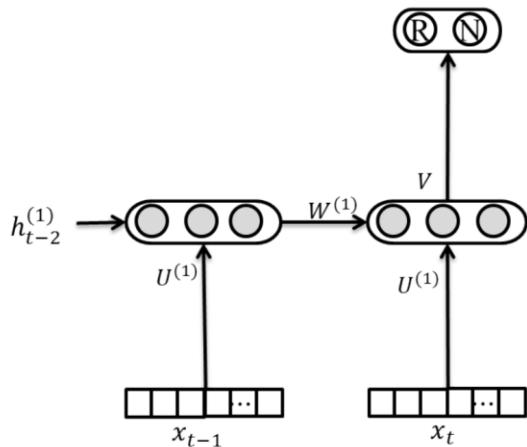
(a) In rumors



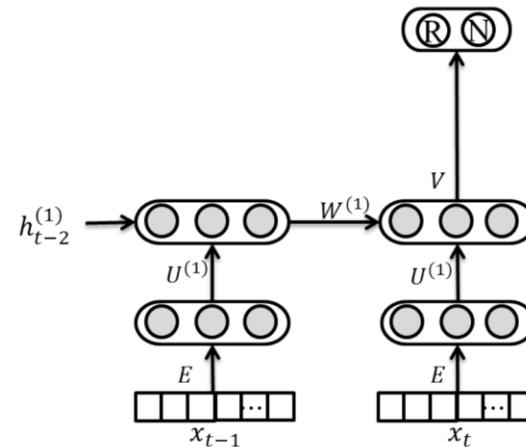
(b) In non-rumors

Rumor Detection in Social Media: Timing Matters

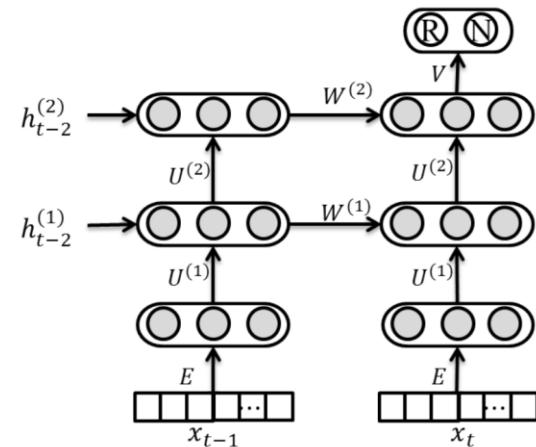
Output
Hidden-2
Hidden-1
Embedding
Input (vocab)



(a) Basic tanh-RNN



(b) 1-layer LSTM/GRU + embedding



(c) 2-layer GRU + embedding

SemEval-2017 Task 8

RumourEval: Determining rumour veracity and support for rumours

<http://alt.qcri.org/semeval2017/task8/>

RumourEval: Determining rumour veracity and support for rumours

Media is full of false claims. Even Oxford Dictionaries named "post-truth" as the word of the year. This makes it more important than ever to build systems that can identify the truth of a story, and the kind of discourse there is around it.

RumourEval invites people and groups to create systems that can do just that. We provide data for training and testing a system, and then in January 2017, will release some evaluation data. You will build a system ahead of that, and then run the system on the evaluation data, and send us the results. We will share the results, and all systems participating will be invited to present their work and analysis at a major conference in summer 2017 as part of the large SemEval workshop.

This task aims to identify and handle rumours and reactions to them, in text. We present an annotation scheme, a large dataset covering multiple topics -- each having their own families of claims and replies -- and concrete subtasks.



RumourEval 2019

<https://competitions.codalab.org/competitions/19938>

Organized by ggorrell - Current server time: Aug. 31, 2019, 6 a.m. UTC

Previous

Evaluation

Jan. 10, 2019, midnight UTC

► Current

Post-Evaluation

Feb. 1, 2019, midnight UTC

End

Competition Ends

Never

Learn the Details

Phases

Participate

Results

Overview

Evaluation

Terms and Conditions

Welcome to RumourEval 2019!

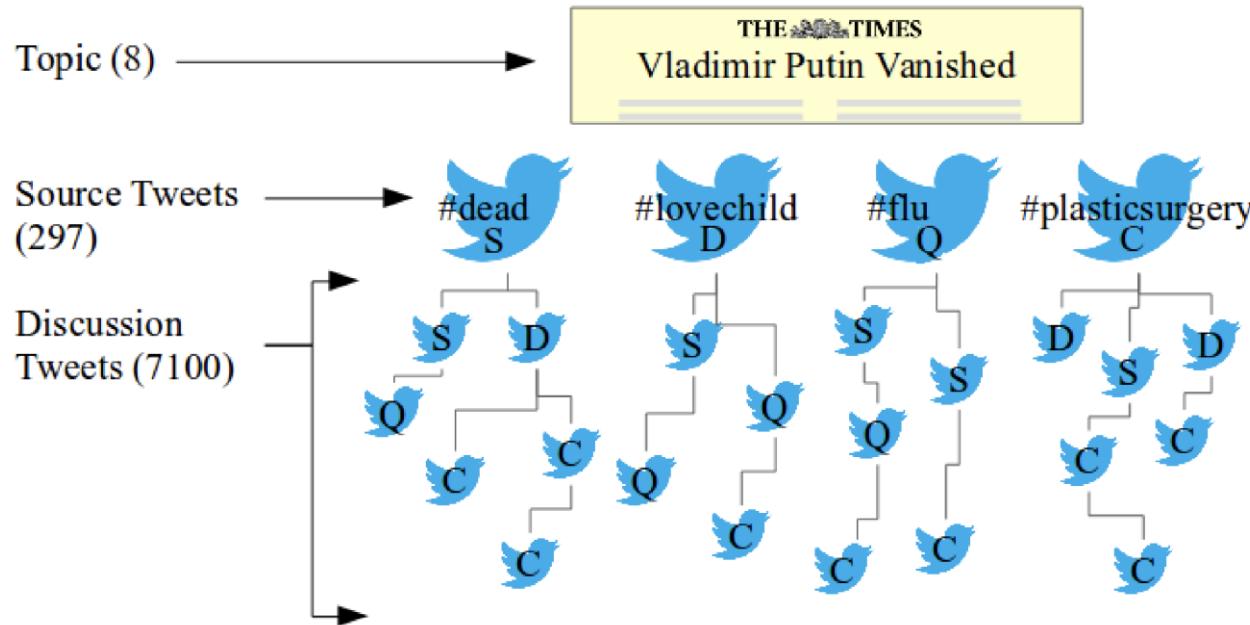
The core mission is to automatically determine the veracity of rumours. The task falls into two parts; task A, in which responses to a rumourous post are classified according to stance, and task B, in which the statements themselves are classified for veracity. Each is described in more detail below.

Latest news! Trial data is now available in the practise phase, consisting of the English language Twitter training/dev data and a toy Reddit set to illustrate the format. Full training/dev data forthcoming.

Please do join the [Google group](#) for the task, where you will find answers to your questions.

181

RumourEval (SemEval-2019 Task 7)



RumourEval (SemEval-2019 Task 7)

[*depth=0*] **u1:** These are not timid colours; soldiers back guarding Tomb of Unknown Soldier after today's shooting #StandforCanada –PICTURE– [support]

[*depth=1*] **u2:** @u1 Apparently a hoax. Best to take Tweet down. [deny]

[*depth=1*] **u3:** @u1 This photo was taken this morning, before the shooting. [deny]

[*depth=1*] **u4:** @u1 I don't believe there are soldiers guarding this area right now. [deny]

[*depth=2*] **u5:** @u4 wondered as well. I've reached out to someone who would know just to confirm that. Hopefully get response soon. [comment]

[*depth=3*] **u4:** @u5 ok, thanks. [comment]

Fact-Checking: Using Multi-Source Data



"I was totally against the war in Iraq."

— Donald Trump on Wednesday, September 7th, 2016 in
the NBC Commander-In-Chief Forum



Analysis

Planned Parenthood's false stat: 'Thousands' of women died every year before Roe

Planned Parenthood's president claims thousands of women died from unsafe abortions before Roe v. Wade. That was out of date decades ago.



Glenn Kessler · Abortion · 5 days ago



Viral image

-



FactCheck.org

A meme claims to show a Time magazine cover from
1977 warning of the "Coming Ice Age."

A tagging system for fact-checks

A new way to identify fact-check articles for
search engines and apps.

ClaimReview

Thing > CreativeWork > Review > ClaimReview

<https://schema.org/ClaimReview>

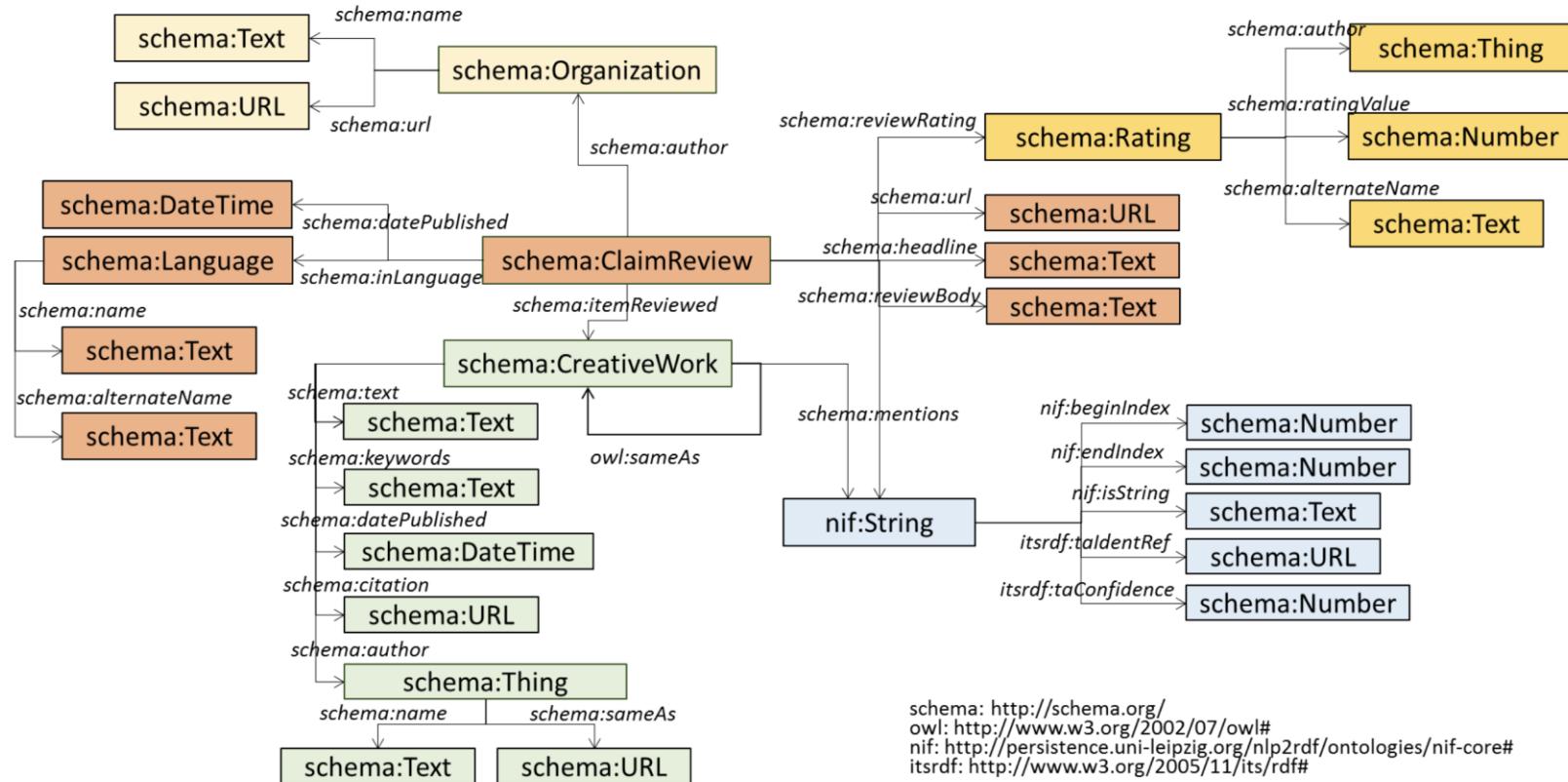
A fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed).

[more...]

Property	Expected Type	Description
Properties from ClaimReview		
claimReviewed	Text	A short summary of the specific claims reviewed in a ClaimReview.
Properties from Review		
itemReviewed	Thing	The item that is being reviewed/rated.
reviewAspect	Text	This Review or Rating is relevant to this part or facet of the itemReviewed.
reviewBody	Text	The actual body of the review.
reviewRating	Rating	The rating given in this review. Note that reviews can themselves be rated. The reviewRating applies to rating given by the review. The aggregateRating property applies to the review itself, as a creative work.
Properties from CreativeWork		
about	Thing	The subject matter of the content. Inverse property: subjectOf .
abstract	Text	An abstract is a short description that summarizes a CreativeWork.

ClaimsKG

<https://data.gesis.org/claimskg/site/>



schema: <http://schema.org/>
 owl: <http://www.w3.org/2002/07/owl#>
 nif: <http://persistence.uni-leipzig.org/nlp2rdf/ontologies/nif-core#>
 itsrdf: <http://www.w3.org/2005/11/its/rdf#>

ClaimsKG

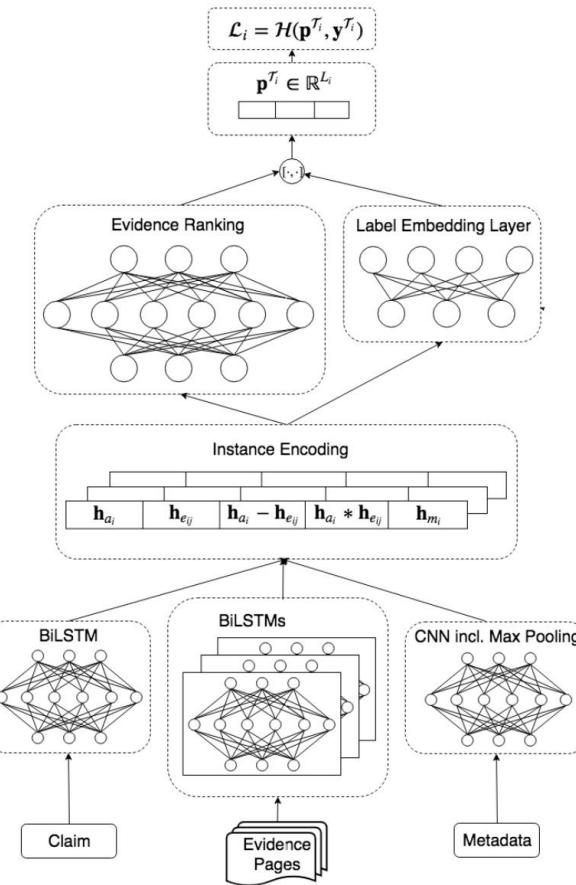
<https://data.gesis.org/claimskg/site/>

Statistics

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

MultiMC

Feature	Value
ClaimID	farg-00004
Claim	Mexico and Canada assemble cars with foreign parts and send them to the U.S. with no tax.
Label	distorts
Claim URL	https://www.factcheck.org/2018/10/factchecking-trump-on-trade/
Reason	None
Category	the-factcheck-wire
Speaker	Donald Trump
Checker	Eugene Kiely
Tags	North American Free Trade Agreement
Claim Entities	United States, Canada, Mexico
Article Title	FactChecking Trump on Trade
Publish Date	October 3, 2018
Claim Date	Monday, October 1, 2018



MultIMC

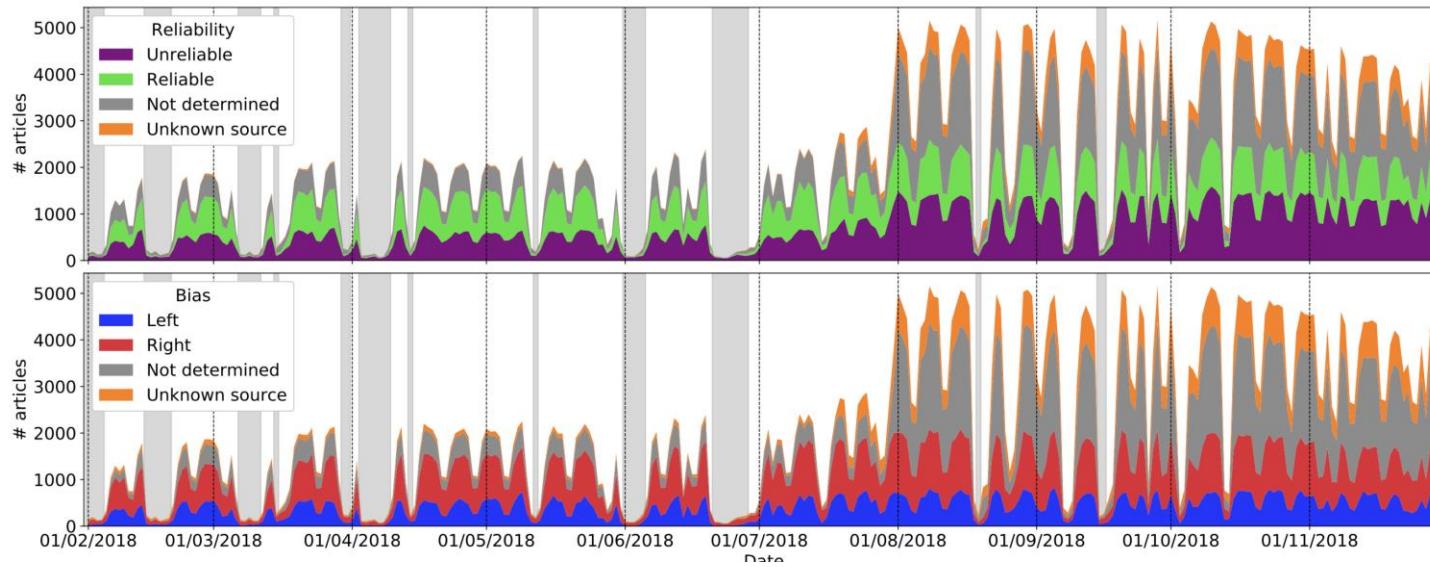
Domain	# Insts	# Labels	Labels
abbc	436	3	in-between, in-the-red, in-the-green
afck	433	7	correct, incorrect, mostly-correct, unproven, misleading, understated, exaggerated
bove	295	2	none, rating: false
chct	355	4	verdict: true, verdict: false, verdict: unsubstantiated, none
clk	38	3	incorrect, unsupported, misleading
faan	111	3	factscan score: false, factscan score: true, factscan score: misleading
faly	71	5	true, none, partly true, unverified, false
fani	20	3	conclusion: accurate, conclusion: false, conclusion: unclear
farg	485	11	false, none, distorts the facts, misleading, spins the facts, no evidence, not the whole story, unsupported, cherry picks, exaggerates, out of context 0, 1, 2, 3, 4, 10
goop	2943	6	
hoer	1310	7	facebook scams, true messages, bogus warning, statirical reports, fake news, unsubstantiated messages, misleading recommendations
huca	34	3	a lot of baloney, a little baloney, some baloney
mpws	47	3	accurate, false, misleading
obry	59	4	mostly_true, verified, unobservable, mostly_false
para	222	7	mostly false, mostly true, half-true, false, true, pants on fire!, half flip
peck	65	3	false, true, partially true
pomt	15390	9	half-true, false, mostly true, mostly false, true, pants on fire!, full flop, half flip, no flip
pose	1361	6	promise kept, promise broken, compromise, in the works, not yet rated, stalled
ranz	21	2	fact, fiction
snes	6455	12	false, true, mixture, unproven, mostly false, mostly true, miscaptioned, legend, outdated, misattributed, scam, correct attribution
thet	79	6	none, mostly false, mostly true, half true, false, true
thal	74	2	none, we rate this claim false
tron	3423	27	fiction!, truth!, unproven!, truth! & fiction!, mostly fiction!, none, disputed!, truth! & misleading!, authorship confirmed!, mostly truth!, incorrect attribution!, scam!, investigation pending!, confirmed authorship!, commentary!, previously truth! now resolved!, outdated!, truth! & outdated!, virus!, fiction! & satire!, truth! & unproven!, misleading!, grass roots movement!, opinion!, correct attribution!, truth! & disputed!, inaccurate attribution!
vees	504	4	none, fake, misleading, false
vogo	653	8	none, determination: false, determination: true, determination: mostly true, determination: misleading, determination: barely true, determination: huckster propaganda, determination: false, determination: a stretch
wast	201	7	4 pinnochos, 3 pinnochos, 2 pinnochos, false, not the whole story, needs context, none

Dataset	# Claims	Labels	metadata	Claim Sources
I: Veracity prediction w/o evidence				
Wang (2017)	12,836	6	Yes	Politifact
Pérez-Rosas et al. (2018)	980	2	No	News Websites
II: Veracity				
Bachenko et al. (2008)	275	2	No	Criminal Reports
Mihalcea and Strapparava (2009)	600	2	No	Crowd Authors
Mitra and Gilbert (2015)†	1,049	5	No	Twitter
Ciampaglia et al. (2015)†	10,000	2	No	Google, Wikipedia
Popat et al. (2016)	5,013	2	Yes	Wikipedia, Snopes
Shu et al. (2018)†	23,921	2	Yes	Politifact, gossipcop.com
Datacommons Fact Check ²	10,564	2-6	Yes	Fact Checking Websites
III: Veracity (evidence encouraged, but not provided)				
Barrn-Cedeo et al. (2018)	150	3	No	factcheck.org, Snopes
IV: Veracity + stance				
Vlachos and Riedel (2014)	106	5	Yes	Politifact, Channel 4 News
Zubiaga et al. (2016)	330	3	Yes	Twitter
Derczynski et al. (2017)	325	3	Yes	Twitter
Baly et al. (2018)	422	2	No	ara.reuters.com, verify-sy.com
Thorne et al. (2018)†	185,445	3	No	Wikipedia
V: Veracity + evidence relevancy				
MultiFC	36,534	2-40	Yes	Fact Checking Websites

Fake News Detection: Datasets

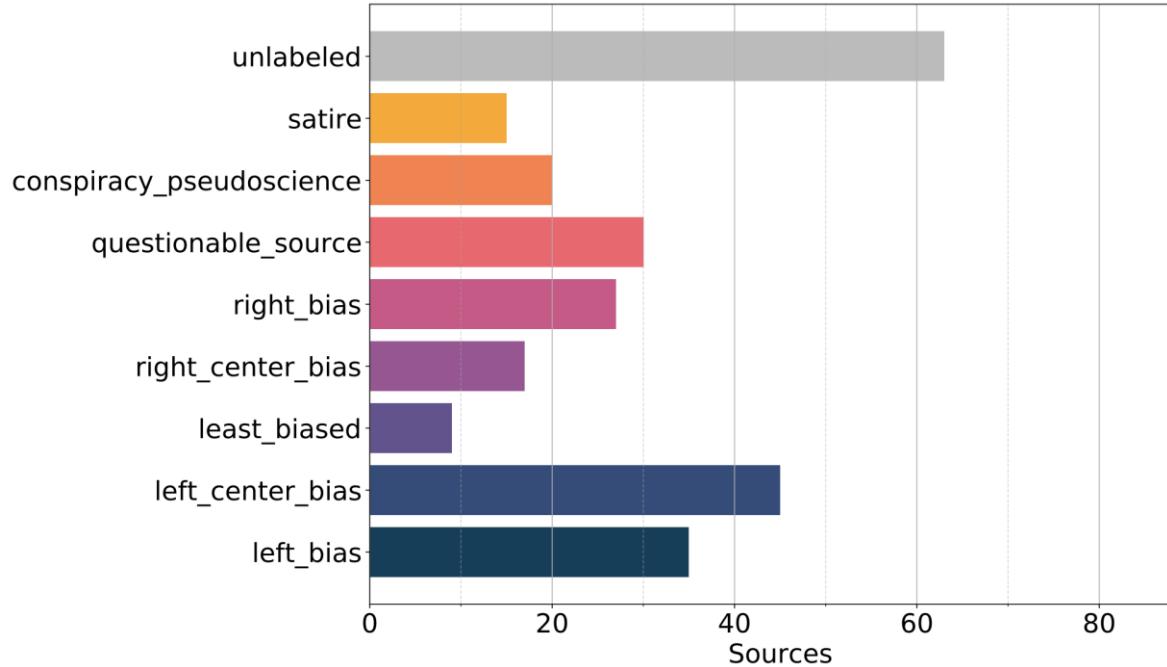
NELA-GT-2018 Fake News Dataset

- **713k articles; 194 news and media outlets**
- **distant supervision for annotations**

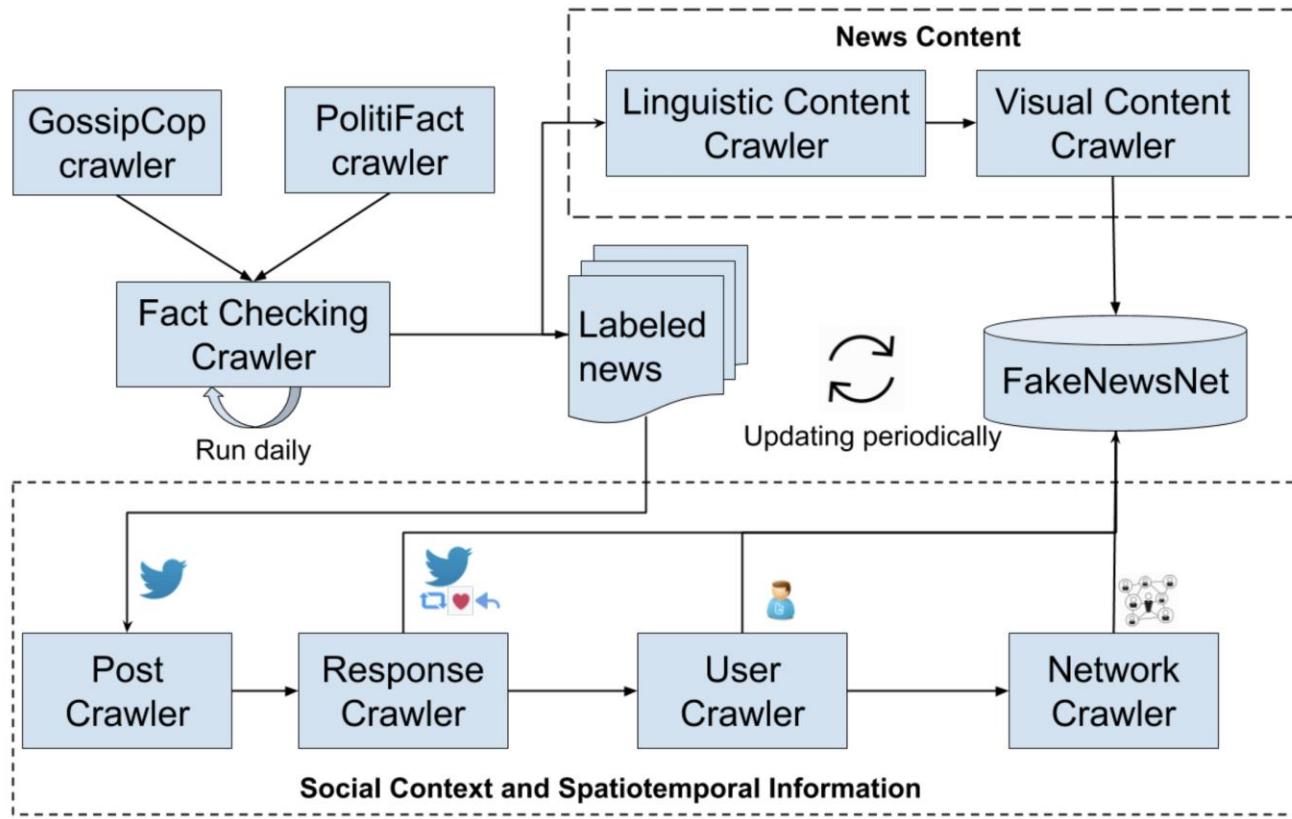


NELA-GT-2019 Fake News Dataset

- **1.12M articles; 260 news and media outlets**
- **distant supervision for annotations**



FakeNewsNet



FakeNewsNet

	Category	Features	PolitiFact		GossipCop	
			Fake	Real	Fake	Real
News Content	Linguistic	# News articles	432	624	5,323	16,817
		# News articles with text	420	528	4,947	16,694
	Visual	# News articles with images	336	447	1,650	16,767
Social Context	User	# Users posting tweets	95,553	249,887	265,155	80,137
		# Users involved in likes	113,473	401,363	348,852	145,078
		# Users involved in retweets	106,195	346,459	239,483	118,894
		# Users involved in replies	40,585	18,6675	106,325	50,799
	Post	# Tweets posting news	164,892	399,237	519,581	876,967
	Response	# Tweets with replies	11,975	41,852	39,717	11,912
		# Tweets with likes	31692	93,839	96,906	41,889
		# Tweets with retweets	23,489	67,035	56,552	24,955
	Network	# Followers	405,509,460	1,012,218,640	630,231,413	293,001,487
		# Followees	449,463,557	1,071,492,603	619,207,586	308,428,225
		Average # followers	1299.98	982.67	1020.99	933.64
		Average # followees	1440.89	1040.21	1003.14	982.80
Spatiotemporal Information	Spatial	# User profiles with locations	217,379	719,331	429,547	220,264
		# Tweets with locations	3,337	12,692	12,286	2,451
	Temporal	# Timestamps for news pieces	296	167	3,558	9,119
		# Timestamps for response	171,301	669,641	381,600	200,531

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, Huan Liu:

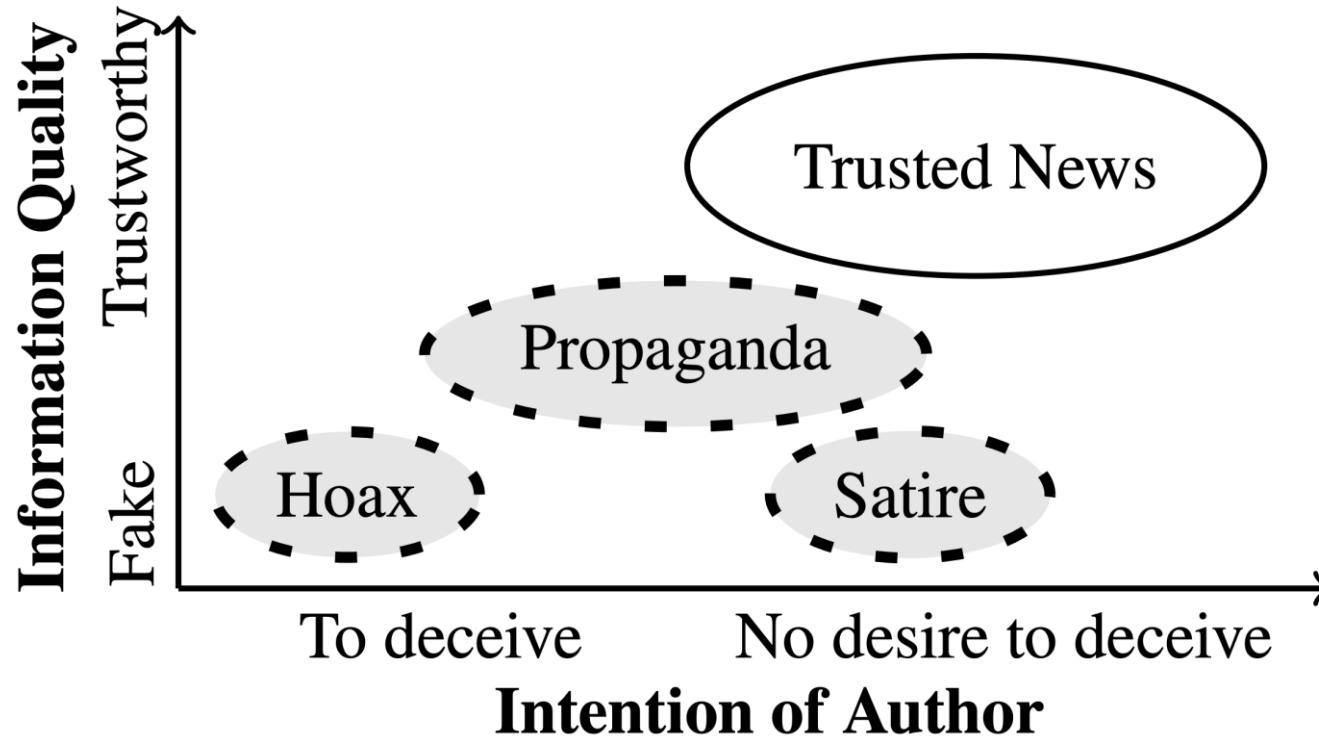
FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media.
CoRR abs/1809.01286 (2018)

FakeNewsNet

Dataset	Features	News Content		Social Context				Spatiotemporal Information	
		Linguistic	Visual	User	Post	Response	Network	Spatial	Temporal
BuzzFeedNews		✓							
LIAR		✓							
BS Detector		✓							
CRED BANK		✓		✓	✓			✓	✓
BuzzFace		✓			✓	✓			✓
FacebookHoax		✓		✓	✓	✓			
FakeNewsNet		✓	✓	✓	✓	✓	✓	✓	✓

Fake News Detection: Some Approaches

The Language of Fake News



The Language of Fake News

LEXICON MARKERS	RATIO	SOURCE	EXAMPLE TEXT	MAX
Swear (LIWC)	7.00	Borowitz Report	... Ms. Rand, who has been damned to eternal torment ...	S
2nd pers (You)	6.73	DC Gazette	You would instinctively justify and rationalize your ...	P
Modal Adverb	2.63	American News	... investigation of Hillary Clinton was inevitably linked ...	S
Action Adverb	2.18	Activist News	... if one foolishly assumes the US State Department ...	S
1st pers singular (I)	2.06	Activist Post	I think its against the law of the land to finance riots ...	S
Manner Adverb	1.87	Natural News	... consequences of deliberately engineering extinction.	S
Sexual (LIWC)	1.80	The Onion	... added that his daughter better not be pregnant .	S
See (LIWC)	1.52	Clickhole	New Yorkers ... can bask in the beautiful image ...	H
Negation(LIWC)	1.51	American News	There is nothing that outrages liberals more than ...	H
Strong subjective	1.51	Clickhole	He has one of the most brilliant minds in basketball.	H
Hedge (Hyland, 2015)	1.19	DC Gazette	As the Communist Party USA website claims ...	H
Superlatives	1.17	Activist News	Fresh water is the single most important natural resource	P
Weak subjective	1.13	American News	... he made that very clear in his response to her.	P
Number (LIWC)	0.43	Xinhua News	... 7 million foreign tourists coming to the country in 2010	S
Hear (LIWC)	0.50	AFP	The prime minister also spoke about the commission ...	S
Money (LIWC)	0.57	NYTimes	He has proposed to lift the state sales tax on groceries	P
Assertive	0.84	NYTimes	Hofstra has guaranteed scholarships to the current players.	P
Comparatives	0.86	Assoc. Press	... from fossil fuels to greener sources of energy	P

Fake vs. Real News vs. Satire

This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News

Benjamin D. Horne and Sibel Adali

Rensselaer Polytechnic Institute
110 8th Street, Troy, New York, USA
{horneb, adalis}@rpi.edu

Abstract

The problem of fake news has gained a lot of attention as it is claimed to have had a significant impact on 2016 US Presidential Elections. Fake news is not a new problem and its spread in social networks is well-studied. Often an underlying assumption in fake news discussion is that it is written to look like real news, fooling the reader who does not check for reliability of the sources or the arguments in its content. Through a unique study of three data sets and features that capture the style and the language of articles, we show that this assumption is not true. Fake news in most cases is more similar to satire than to real news, leading us to conclude that persuasion in fake news is achieved through heuristics rather

homophily (Bakshy, Messing, and Adamic 2015). Furthermore, misleading or wrong information have a higher potential to become viral (Bessi et al. 2015) and lead to negative discussions (Zollo et al. 2015b). Even more troubling, results show resistance by individuals to information challenging established beliefs and that attempts to debunk conspiracy theories are largely ineffective (Zollo et al. 2015a).

As an informed public is crucial to any operating democracy, incorrect information is especially dangerous in political news. As many widely discredited claims have become viral and distributed widely in social networks during the 2016 elections (Silverman 2016), the topic of “fake news”

Benjamin D. Horne, Sibel Adali:

This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News.
CoRR abs/1703.09398 (2017)

dEFEND: Explainable Fake News Detection

Fake News

Iranian Official Drops Bombshell: Obama Secretly Gave Citizenship to 2500 Iranians as Part of Nuke Deal

By [REDACTED] - July 2, 2018

148 Comments

A senior Iranian cleric and member of parliament has just dropped a bombshell.

He is claiming that the Obama administration, as part of negotiating during the Iran Deal, granted U.S. citizenship to 2500 Iranians including family members of government officials.

...

There have been so many things hidden from the public about the Iran Deal if this was one more thing given up in bribe, it wouldn't be hard to believe.

Comments

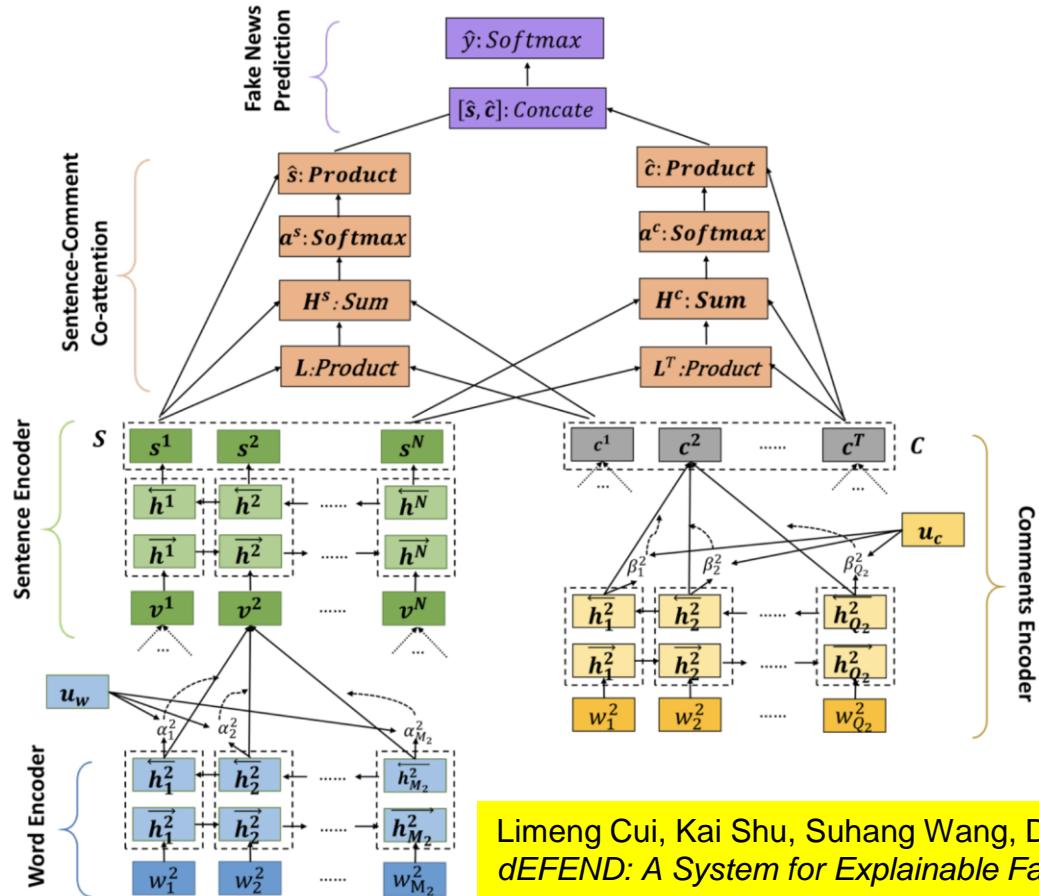
If you had done your research, you would know that the president does not have the power to give citizenship. This would have to be done as an act of congress... (0.0160)

Isn't graft and payoffs normally a offense even for a ex-president? (0.0086)

Wow! What's frightening is where will it end? We could be seeing some serious issues here. (0.0051)

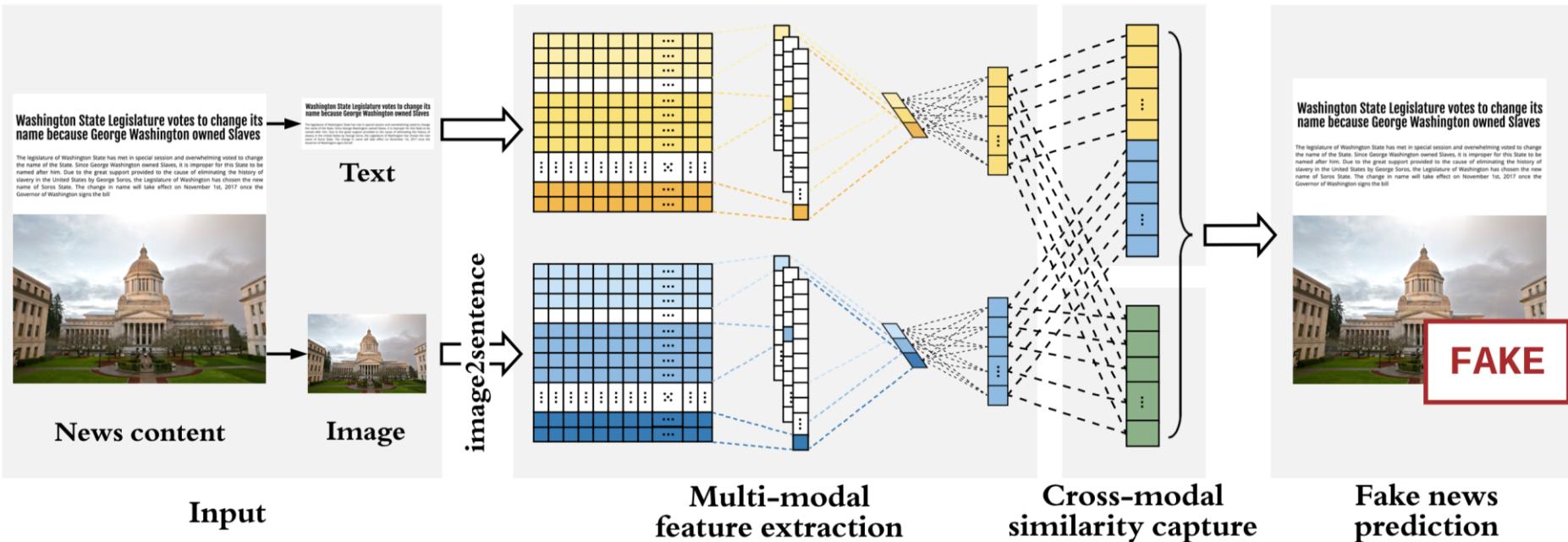
Walkaway from their (0.0080)

dEFEND: Explainable Fake News Detection



Limeng Cui, Kai Shu, Suhang Wang, Dongwon Lee, Huan Liu:
dEFEND: A System for Explainable Fake News Detection. CIKM 2019: 2961-2964

SAFE: Similarity-Aware Multi-modal Fake News Detection



SAFE: Similarity-Aware Multi-modal Fake News Detection

Washington State Legislature votes to change its name because George Washington owned Slaves



(a) $s = 0.024$

MORGUE EMPLOYEE CREMATED BY MISTAKE WHILE TAKING A NAP

Beaumont, Texas | An employee of the Jefferson County morgue died this morning after being accidentally cremated by one of his coworkers.



(b) $s = 0.044$

Angelina Jolie & Jared Leto Dating After

Brad Pitt Divorce – Report



(c) $s = 0.001$

Fig. 5. Fake News

"Face the Nation" transcripts, August 26, 2012: Rubio, Priebus, Barbour, Blackburn



(a) $s = 0.966$

98 Degrees' 2017 Macy's Parade Performance Will Take You Right Back To The '90s



(b) $s = 0.975$

Chrissy Teigen and John Legend Have First Date Night Since Welcoming Son Miles: Pic!



(c) $s = 0.983$

Xinyi Zhou, Jindi Wu, Reza Zafarani:

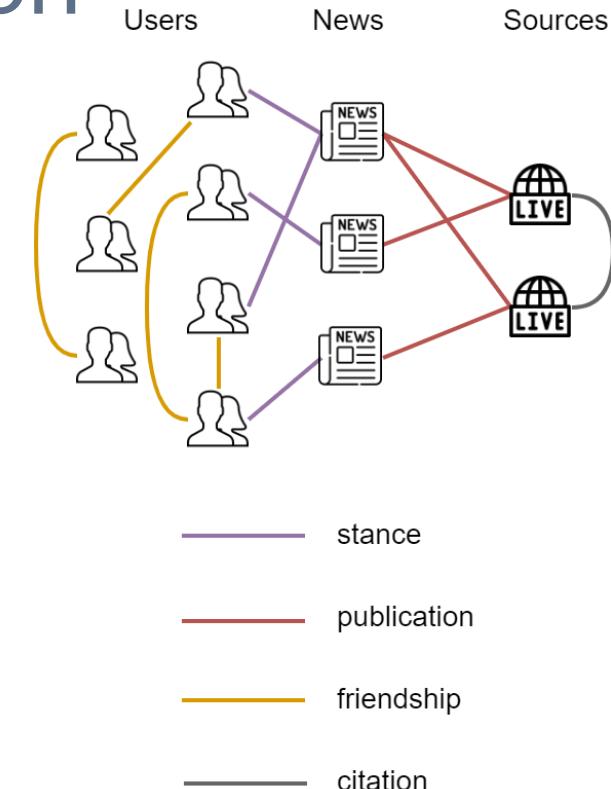
SAFE: Similarity-Aware Multi-modal Fake News Detection. PAKDD (2) 2020: 354-367

FANG: Factual News Graph

Social entities: **users, news, sources**

Social interactions: **stance, publication, friendship, citation**

	User	News	Source
User	Friendship	Stance	
News	Stance		Publication
Source		Publication	Citation



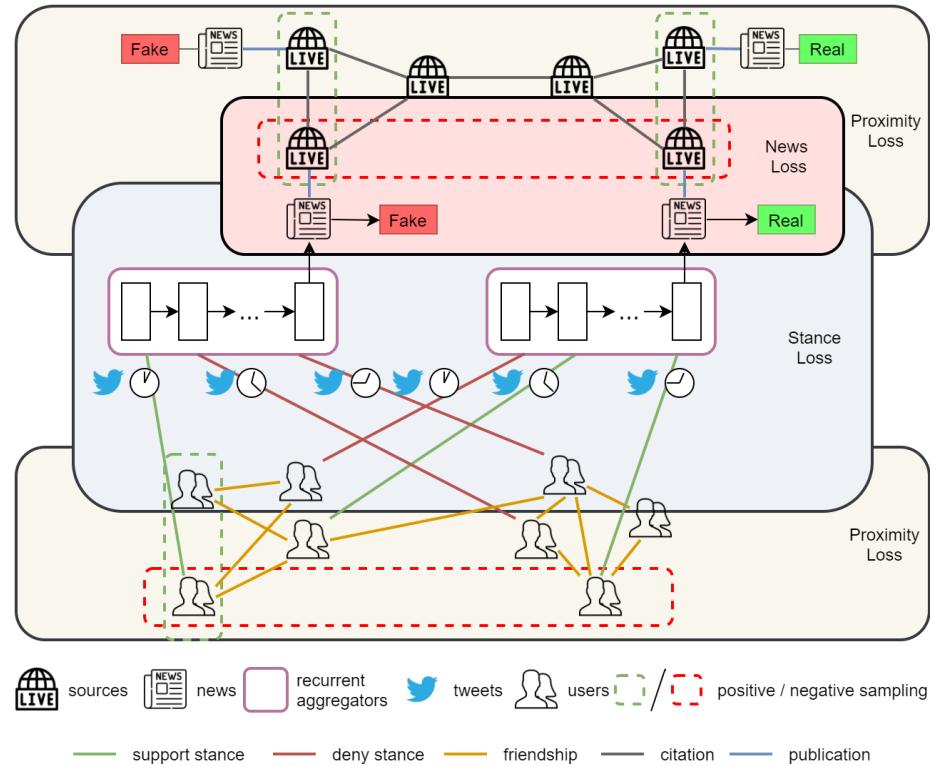
FANG: Factual News Graph

Representations are jointly optimized on

- Proximity loss (unsupervised)
- Stance loss (self-supervised)
- Fake news detection loss (supervised)

Linear combination of the 3 losses

$$\mathcal{L}_{total} = \mathcal{L}_{prox.} + \mathcal{L}_{stance} + \mathcal{L}_{news}$$

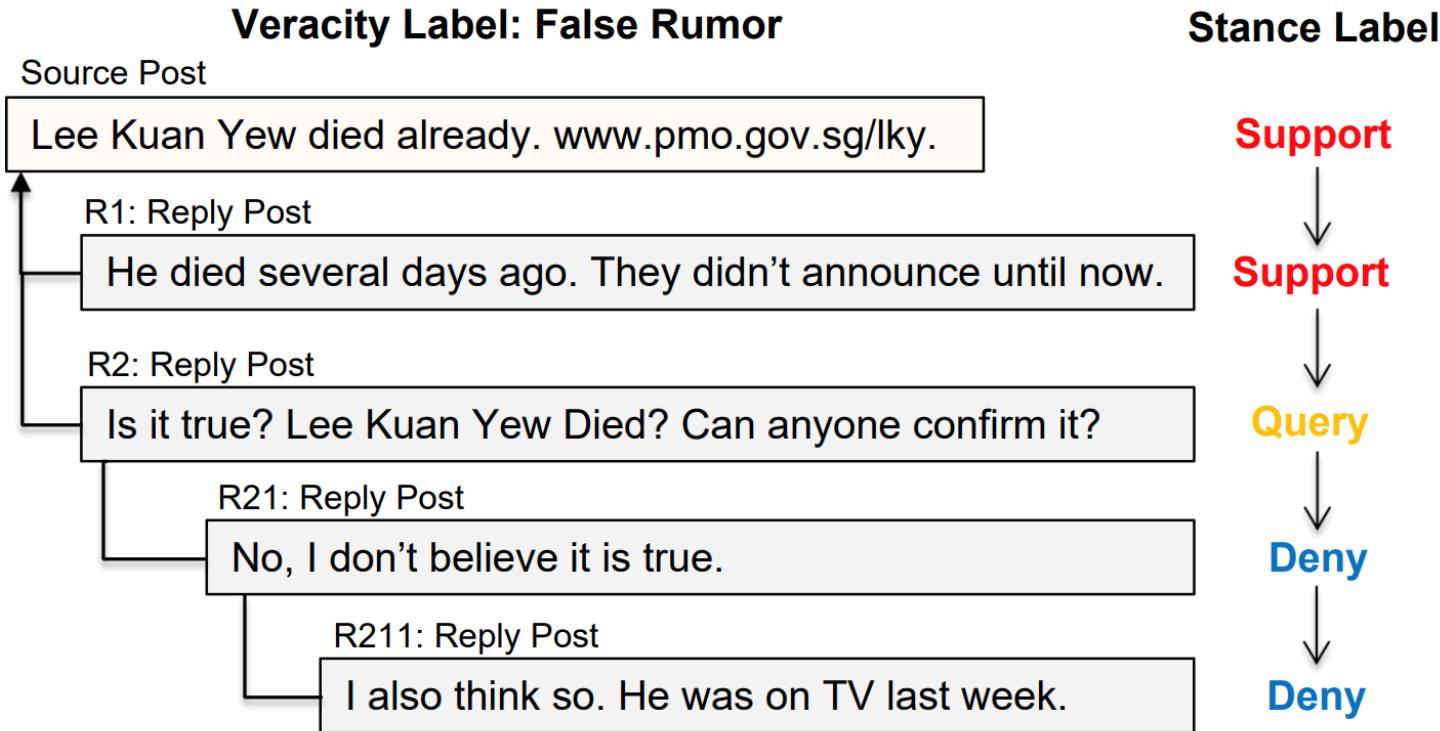


Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, Min-Yen Kan:

FANG: Leveraging Social Context for Fake News Detection Using Graph Representation. CIKM 2020: 1165-1174

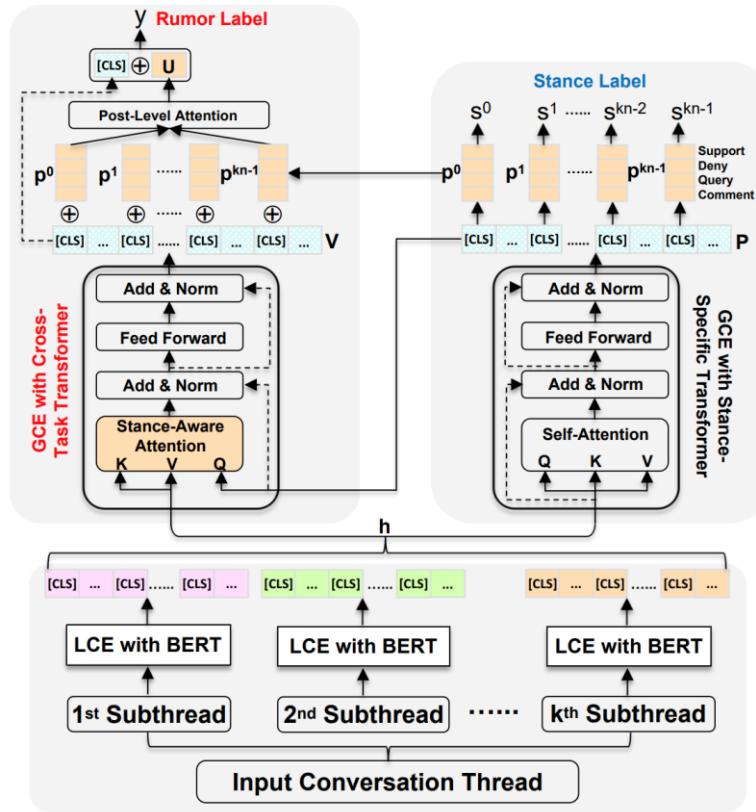
Stance With Respect to a Claim

Stance Detection for Rumour Verification



209

Stance Detection for Rumour Verification



210

Stance Detection for Rumour Verification

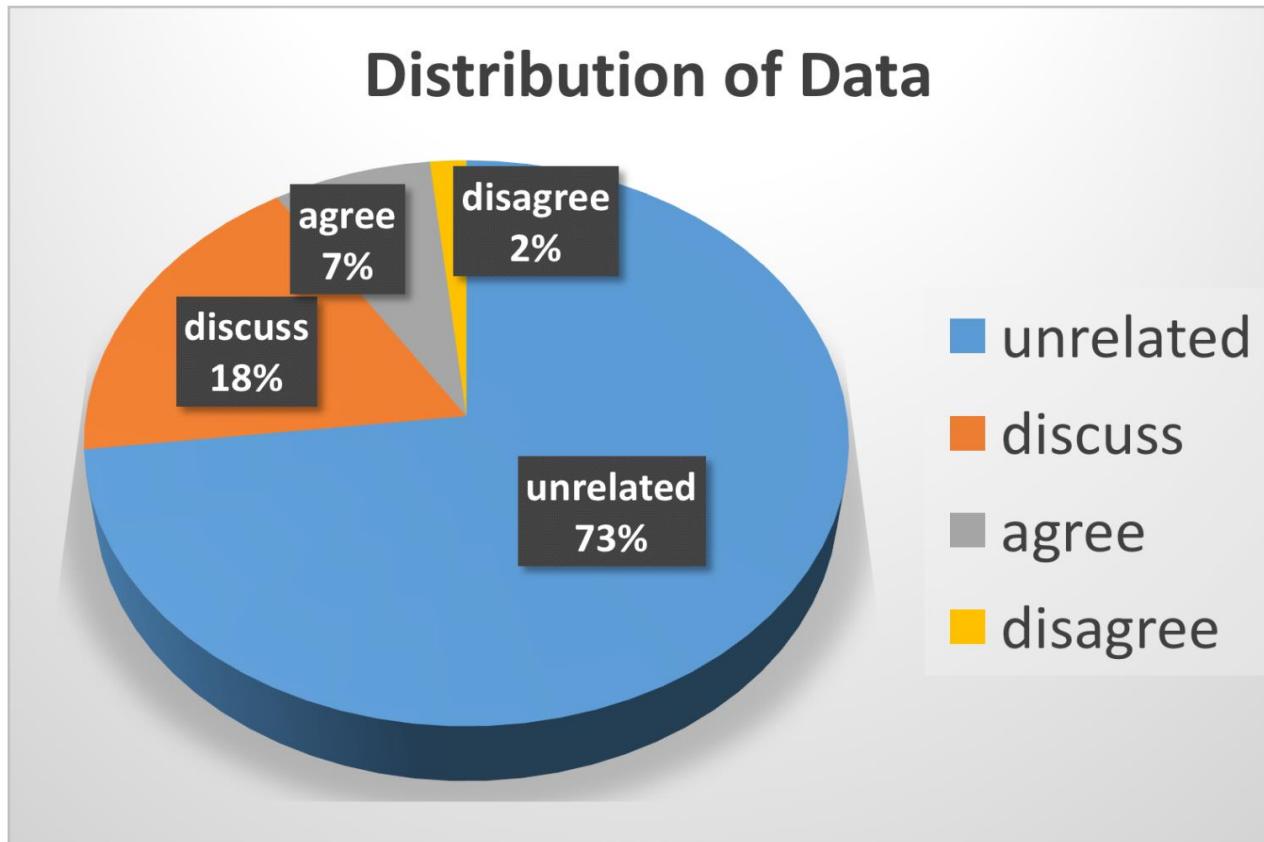
Dataset	#Threads	#Tweets	Stance Labels				Rumor Veracity Labels		
			#Support	#Deny	#Query	#Comment	#True	#False	#Unverified
SemEval-17	325	5,568	1,004	415	464	3,685	145	74	106
PHEME	2,402	105,354		-			1,067	638	697

Fake News Challenge

Headline: Hundreds of Palestinians flee floods in Gaza as Israel opens dams	
Agree (AGR)	GAZA CITY (Ma'an) – Hundreds of Palestinians were evacuated from their homes Sunday morning after Israeli authorities opened a number of dams near the border, flooding the Gaza Valley in the wake of a recent severe winter storm. The Gaza Ministry of Interior said in a statement that civil defense services and teams from the Ministry of Public Works had evacuated more than 80 families from both sides of the Gaza Valley (Wadi Gaza) after their homes flooded as water levels reached more than three meters [...]
Discuss (DSC)	Palestinian officials say hundreds of Gazans were forced to evacuate after Israel opened the gates of several dams on the border with the Gaza Strip, and flooded at least 80 households. Israel has denied the claim as "entirely false". [...]
Disagree (DSG)	Israel has rejected allegations by government officials in the Gaza strip that authorities were responsible for released storm waters flooding parts of the besieged area. "The claim is entirely false, and southern Israel does not have any dams," said a statement from the Coordinator of Government Activities in the Territories (COGAT). "Due to the recent rain, streams were flooded throughout the region with no connection to actions taken by the State of Israel." At least 80 Palestinian families have been evacuated after water levels in the Gaza Valley (Wadi Gaza) rose to almost three meters. [...]
Unrelated (UNR)	Apple is continuing to experience 'Hairgate' problems but they may just be a publicity stunt [...]

Fake News Challenge Dataset

49,972 examples

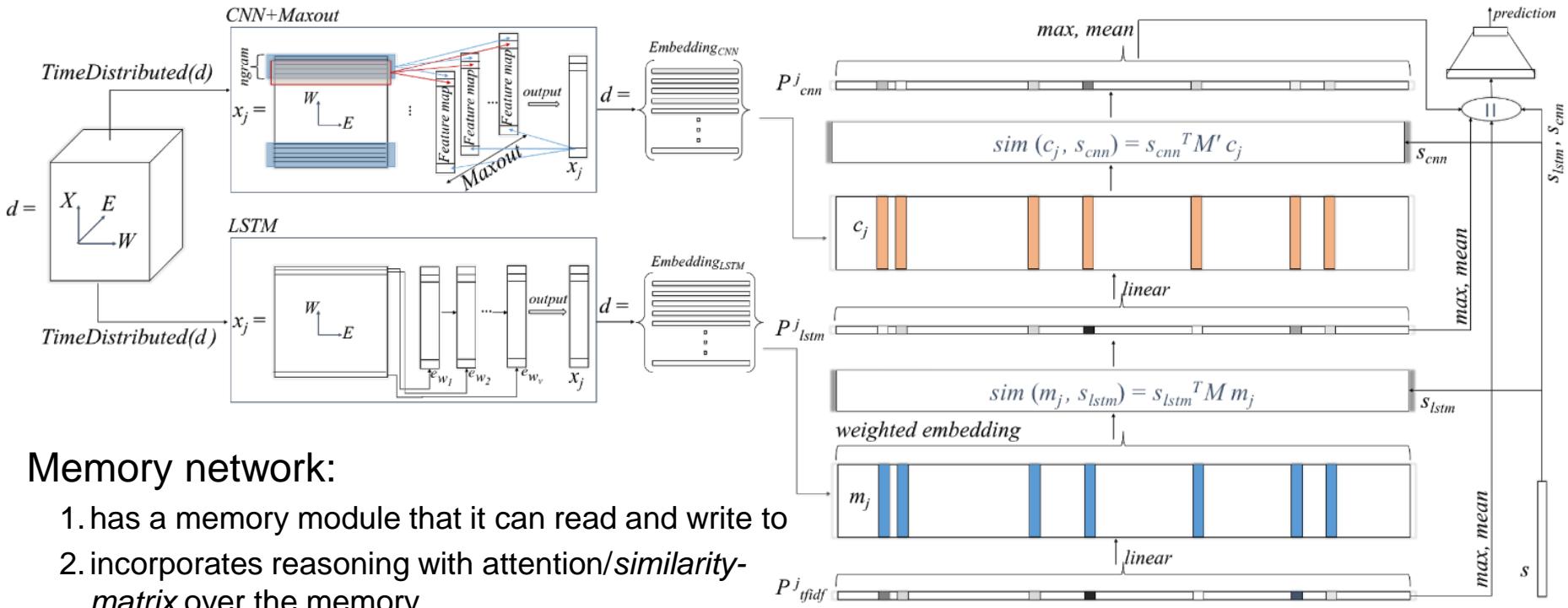


Fake News Challenge: Analysis

Systems	FNC-FNC					
	FNC	F_1m	AGR	DSG	DSC	UNR
Majority vote	.394	.210	0.0	0.0	0.0	.839
TalosComb	.820	.582	.539	.035	.760	.994
TalosTree	.830	.570	.520	.003	.762	.994
TalosCNN	.502	.308	.258	.092	0.0	.882
Athene	.820	.604	.487	.151	.780	.996
UCLMR	.817	.583	.479	.114	.747	.989
featMLP	.825	.607	.530	.151	.766	.982
stackLSTM	.821	.609	.501	.180	.757	.995
Upper bound	.859	.754	.588	.667	.765	.997

Very bad performance
on the *Disagree* class,
which is too small

End-to-End Memory Network (MemNN)



Memory network:

1. has a memory module that it can read and write to
2. incorporates reasoning with attention/similarity-matrix over the memory.

Stance Detection Datasets (English)

Datasets	# samples				
	Train	Dev	Test	Total	classes
arc (Hanselowski et al., 2018; (Habernal et al., 2018)	12,382	1,851	3,559	17,792	unrelated (75%), disagree (10%), agree (9%), discuss (6%)
argmin (Stab et al., 2018)	6,845	1,568	2,726	11,139	argument_against (56%), argument_for (44%)
fnc1 (Pomerleau and Rao, 2017)	42,476	7,496	25,413	75,385	unrelated (73%), discuss (18%), agree (7%), disagree (2%)
iac1 (Walker et al., 2012b)	4,227	454	924	5,605	pro (56%), anti (34%), other (10%)
ibmcs (Bar-Haim et al., 2017)	935	104	1,355	2,394	pro (55%), con (45%)
perspectrum (Chen et al., 2019)	6,978	2,071	2,773	11,822	support (52%), undermine (48%)
scd (Hasan and Ng, 2013)	3,251	624	964	4,839	for (60%), against (40%)
semeval2016t6 (Mohammad et al., 2016)	2,497	417	1,249	4,163	against (51%), favor (25%), none (24%)
semeval2019t7 (Gorrell et al., 2019)	5,217	1,485	1,827	8,529	comment (72%), support (14%), query (7%), deny (7%)
snopes (Hanselowski et al., 2019)	14,416	1,868	3,154	19,438	support (74%), refute (26%)

Emergent Dataset: Stance + Factuality

- 300 rumours
- 2,595 news articles

Claim: Robert Plant ripped up an \$800 million contract offer to reunite Led Zeppelin

Source: mirror.co.uk (shares: 39,140)

Headline: Led Zeppelin's Robert Plant turns down £500MILLION to reform supergroup

Stance: *for*

Source: usnews.com (shares: 850)

Headline: No, Robert Plant Didn't Rip Up an \$800 Million Contract

Stance: *against*

Source: forbes.com (shares: 3,360)

Headline: Robert Plant Reportedly Tears Up \$800 Million Led Zeppelin Reunion Contract

Stance: *observing*

Veracity: *False*

STANCE + FACTUALITY + RATIONALE (FOR ARABIC)

CORPUS

Domain: Syria/Middle-East

Claim Sources:

- True claims: *Reuters* [3]
- False claims: *VERIFY* [4]

انسحب تنظيم الدولة الإسلامية داعش من مدينة الرقة بشكل كامل
ISIS has completely pulled out from the City of Raqqa

<http://elaph.com/Web/News/2017/12/1180724.html> ✓

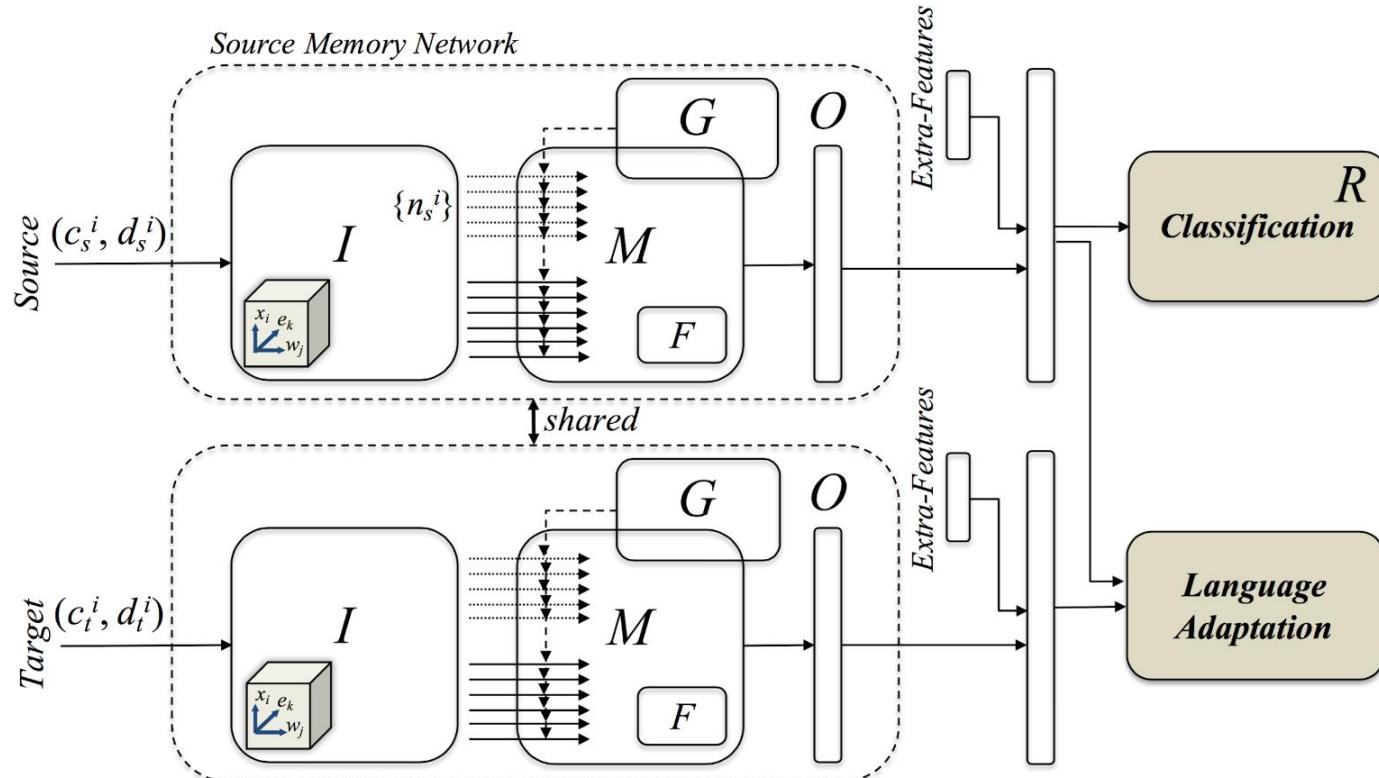
- الرقة: في الخامس من نوفمبر 2016 شنت قوات سوريا الديمقراطية التي تضم قوات كردية بشكل أساسي مع قوات عربية، هجوما على مدينة الرقة، أierz معقل لتنظيم الدولة الإسلامية في سوريا، ودخلت المدينة في يونيو. وأعلنت في 17 أكتوبر 2017 السيطرة الكاملة على المدينة و«انتهاء العمليات العسكرية».

<http://www.alhayat.com/article/890385> ✗

ومني تنظيم «داعش» بخسائر ميدانية متلاحقة في سوريا، حيث يوشك حالياً على خسارة مدينة الرقة (شمال) معقله في سوريا. ويتصدى التنظيم لهجومين منفصلين في محافظة دير الزور (شرق)، الأول تقوده القوات النظامية بدعم روسي في مدينة دير الزور وريفها الغربي، والثاني تنفذه «قوات سوريا الديمقراطية» بدعم من التحالف الدولي بقيادة أمريكية في الريف الشرقي.

Claims	Retrieved Documents	Stance (document-to-claim)			
		Agree	Disagree	Discuss	Unrelated
False: 219	1,239	103	82	159	895
True: 203	1,803	371	5	250	1,177
Total: 402	3,042	474	87	409	2,072

Cross-Lingual Memory Network



**Stance
With Respect to a Topic**

SemEval-2016 Task 6: Tweet Stance Detection

A foetus has rights too! Make your voice heard.

(Target: *women have the right to abortion* → AGAINST)

All Mitt Romney cares about is making money for the rich.

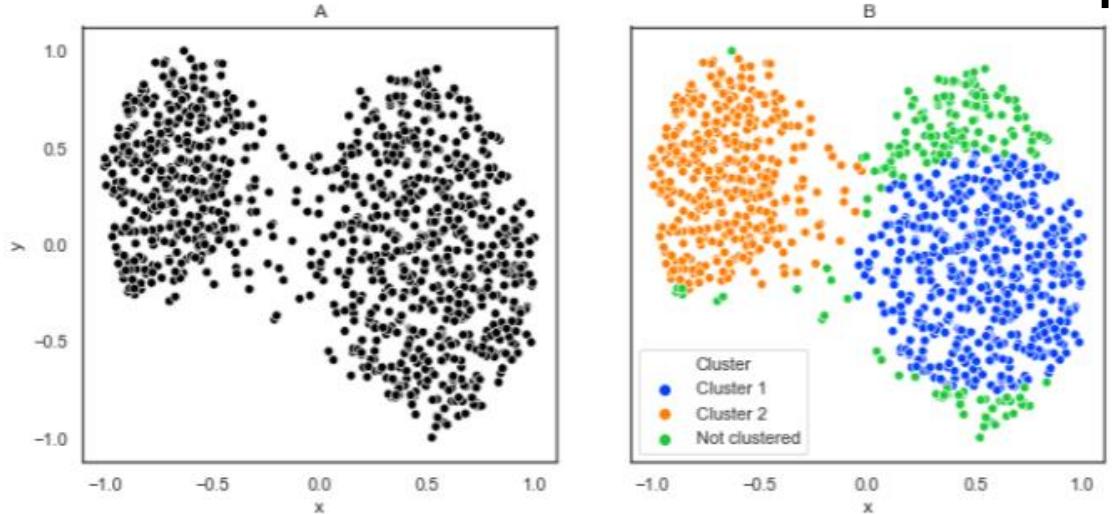
(Target: *Barack Obama* → INFAVOR)

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, Colin Cherry:
SemEval-2016 Task 6: Detecting Stance in Tweets. SemEval@NAACL-HLT 2016: 31-41

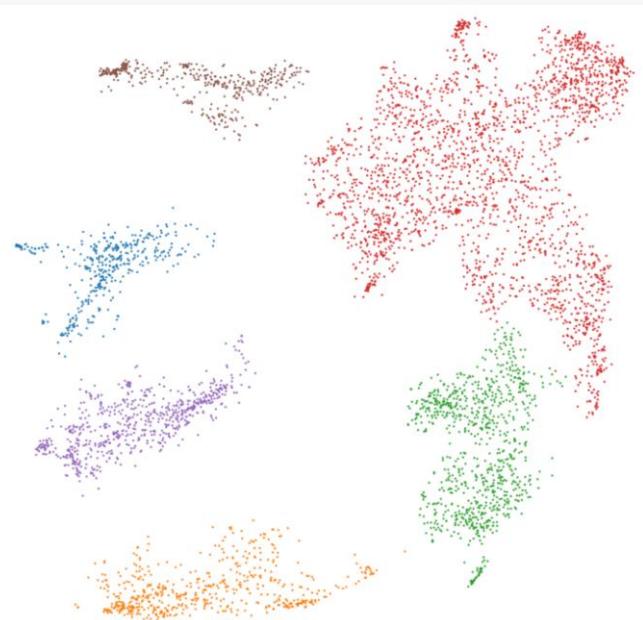
Saif M. Mohammad, Parinaz Sobhani, Svetlana Kiritchenko:
Stance and Sentiment in Tweets. ACM Trans. Internet Techn. 17(3): 26:1-26:23 (2017)

User Stance Detection

Retweets: UMAP & mean shift Pro/Anti Trump



Embeddings: UMAP & HDBSCAN Turkish voters



Kareem Darwish, Peter Stefanov, Michaël J. Aupetit, Preslav Nakov:
Unsupervised User Stance Detection on Twitter. ICWSM 2020: 141-152

Media Stance Detection

$$V(I) = 2 \frac{\frac{tf(I, C_0)}{total(C_0)}}{\frac{tf(I, C_0)}{total(C_0)} + \frac{tf(I, C_1)}{total(C_1)}} - 1$$

$$Cat(V) = \begin{cases} --, & \text{if } s \in [-1, -0.6) \\ -, & \text{if } s \in [-0.6, -0.2) \\ 0, & \text{if } s \in [-0.2, 0.2) \\ +, & \text{if } s \in [0.2, 0.6) \\ ++, & \text{if } s \in [0.6, 1] \end{cases}$$

Medium	Average	climate change	gun control	ilhan	immigration	midterm	police & racism	Kavanaugh	vaccine
thehill.com	+	0	++	+	+	+	+	++	++
theguardian.com	++	++	++	++	++	++	++	++	++
washingtonpost.com	++	++	++	++	++	++	++	++	++
breitbart.com	--	--	--	--	--	--	--	--	--
foxnews.com	--	--	--	--	--	--	--	--	--
nytimes.com	++	+	++	+	+	+	++	++	++

Political Bias Detection

Article-Level Bias Detection



As Lockdowns Ease, a Focus on Safety Risks and the Need for Normalcy

As the U.S. death toll from the COVID-19 coronavirus pandemic surpasses 100,000 in the coming days, Americans are moving to stay safe and avoid spreading the virus while lockdowns are...

From the Center

From Camping To Dining Out: Here's How Experts Rate The Risks Of 14 Summer Activities

NPR Online News L L C R R

From the Right

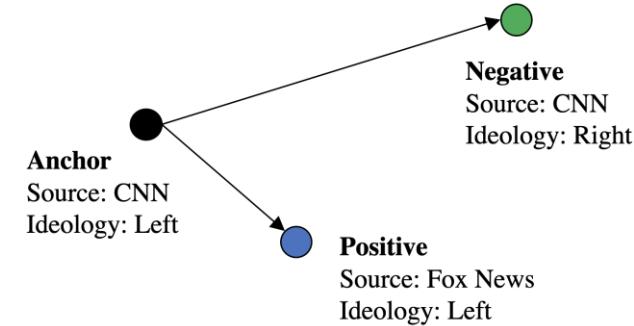
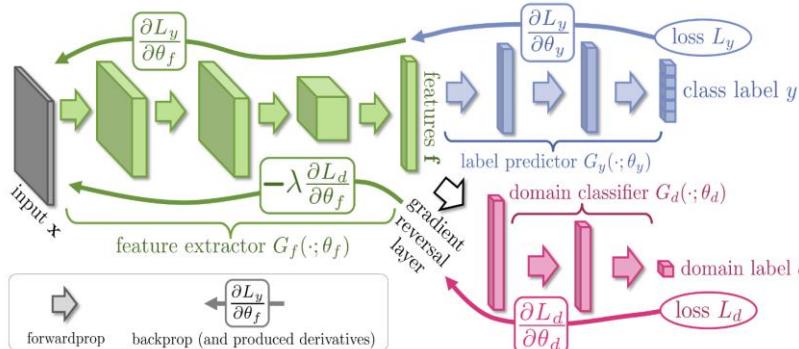
The covid-19 lockdown has served its purpose. It's time to end it.

Marc A. Thiessen L L C R R

From the Left

How to weigh the risk of going out in the coronavirus pandemic, in one chart

Vox L L C R R



YouTube Channel Bias Detection

YouTube Politics Media Annotate About

Search...

Daily Mail

Daily Mail (MailOnline) is one of the world's largest English speaking news sites. Feed your daily addiction with trending and viral videos from all our top channels: US news, UK news & World news, Celebrity & Showbiz, Sports, Femail, Science & Tech, Health, Money, Travel, and many more.

Cristhian Rivera appears in court for Molly Tibbetts' death x

Cristhian Rivera, the 24-year-old suspected illegal immigrant charged with murdering Mollie Tibbetts, has arrived in court in Iowa to make his first appearance since taking police to her body. Original Article:

<http://www.dailymail.co.uk/news/article-6087793/Illegal-immigrant-24-charged-Mollie-Tibbetts-murder-makes-court-appearance.html> Original Video: <http://www.dailymail.co.uk/video/news/video-1749634/Video-Man>

cristhian rivera court molly tibbetts death abduction college

disappearance iowa missing girl missing jogger missing person

mollie mollie tibbet mollie tibbetts mollie tibbetts boyfriend

mollie tibbetts missing mollie tibbetts missing iowa

Analyzed videos



Media Bias Detection

Medium	factuality	bias	Average	climate change	gun control	ilhan	immigration	midterm	police & racism	Kavanaugh	vaccine
thehill.com	H	L-C	+	0	++	+	+	+	+	++	++
theguardian.com	H	L-C	++	++	++	++	++	++	++	++	++
washingtonpost.com	H	L-C	++	++	++	++	++	++	++	++	++
breitbart.com	VL	Far R	--	--	--	--	--	--	--	--	--
foxnews.com	M	R	--	--	--	--	--	--	--	--	--
nytimes.com	H	L-C	++	+	++	+	+	+	++	++	++
cnn.com	M	L	+	+	++	+	++	+	+	++	+
apple.news			+	0	0	+	0	0	+	+	++
dailycaller.com	M	R	--	--	--	--	--	--	--	--	--
rawstory.com	M	L	++	++	++	++	++	++	++	++	++
huffingtonpost.com	H	L	++	++	++	++	++	+	++	++	++
truepundit.com	L	--	--	--	--	--	--	--	--	--	--

Peter Stefanov, Kareem Darwish, Atanas Atanasov, Preslav Nakov:
Predicting the Topical Stance and Political Leaning of Media using Tweets. ACL 2020: 527-537

Source Reliability (and Bias) Estimation

What was Written

- Published Articles 
- Twitter Profile 
- YouTube Videos 

NEWS MEDIUM

Who Read it

- Twitter Followers' Bios 
- Facebook Followers' Demographics 
- YouTube Metadata 

Undercover Nurse Exposes NYC's COVID 'Epicenter' – Revealing Widespread Medical Malpractice

Matt Stoller Explains COVID's Billionaire and Big Government Power-play

What are the real economic implications of the coronavirus lockdown? Will the economy, or society ever regain its former footing?

This interview discusses the macro financial aspects of the COVID crisis with Matt Stoller, author of *Goliath: The 100-Year War Between Monopoly Power and Democracy*, where explains how the coronavirus lockdowns have allowed corporate monopolies to flourish – and he also forecasts the coming disappearance of most small businesses, the problems that emerge when monopolies control health care, the emergence of a Wall Street-oriented "planned economy", the failure of progressives to stop the theft of trillions, and much more.

FoxNews

MSNBC

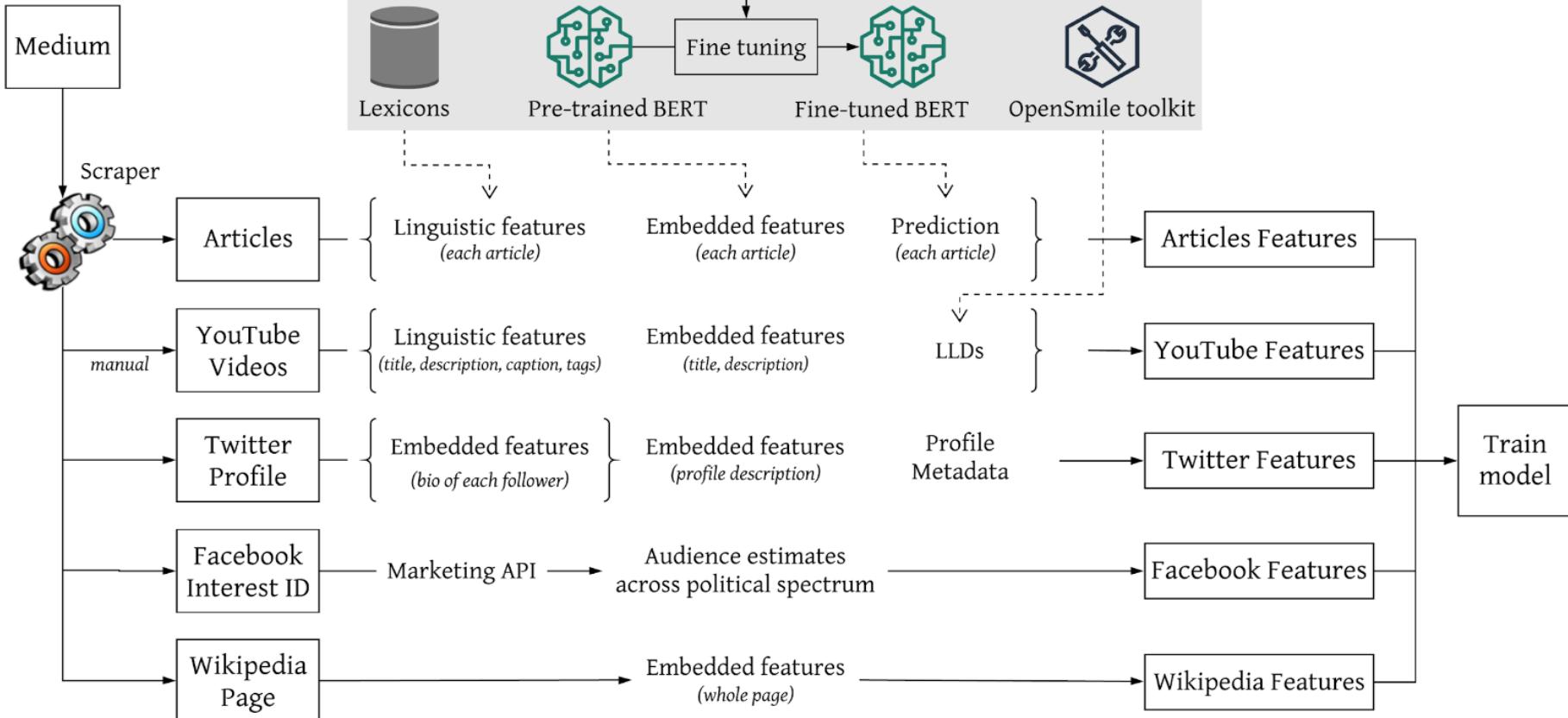
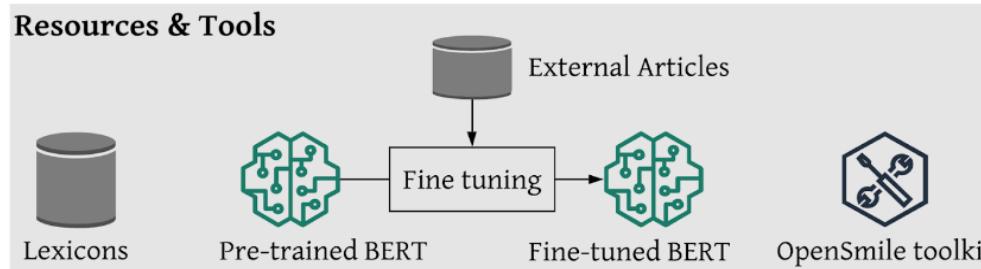
InfoWars

Believer in Jesus #1. **Trump Supporter** all the way. Freedom supporter. #usarmyvet #WWIWGA
Anywhere  Joined June 2009
650 Following 342 Followers
#prolife#trump2020
sticks and stones may break my bones but there will always be something to offend a feminist #allilivesmatter #georgefloydisacriminal
Ireland  Joined November 2016
1,287 Following 52 Followers
Human. Anti-Fascist. Lover of Lasagna. Constitutional. #ripgo #dissident #BlackLivesMatter **#voteouttrump2020 #FoxNewsIsRacist**
#EnoughsEnough
United States of America  Joined September 2019
545 Following 127 Followers
PRO-LIFE BUILD THE WALL Not one bit ashamed of my beliefs. Been standing for SSB & POA for 64 years---Tacoma baseball & Husky FB since 1962-South Tacoma #MAGA
Tacoma, WA 98408  Joined July 2011
2,814 Following 714 Followers

InfoWars
From Wikipedia, the free encyclopedia
"Info Wars" redirects here. For the 2004 film, see *Info Wars (film)*.
InfoWars is a far-right American **conspiracy theory** and **fake news website** owned by Alex Jones.^[14] It was Talk shows and other content for the site are created primarily in studios at an undisclosed location in an internet cafe in Austin, Texas, and broadcast via satellite to approximately 100 countries worldwide. The website receives over 10 million visits, making its reach greater than some mainstream news websites such as *The Economist* and *Newsweek*.
The site has regularly published fake stories which have been linked to harassment of victims.^[a] In February 2018, Jones was sued for \$100 million for his role in the harassment of a woman who worked for him and was harassed by Jones.^[25] InfoWars, and in particular Jones, advocate numerous conspiracy theories particularly about the 9/11 attacks and Sandy Hook shooting. InfoWars has issued retractions various times as a result of legal challenges from victims. The website has been removed from several platforms, including Facebook, Twitter, YouTube, iTunes,^{[26][27][28][29]} and Roku.^[30]
InfoWars earns revenue from the sale of products pitched by Jones during the show, including dietary supplements and a media outlet.^[31]

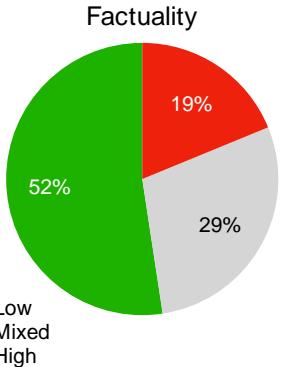
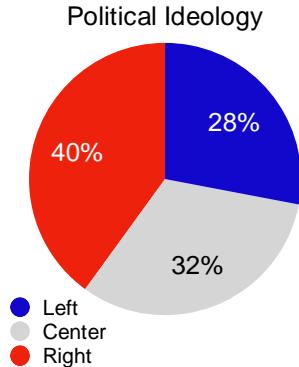
What's Written About it?

- Wikipedia page content 

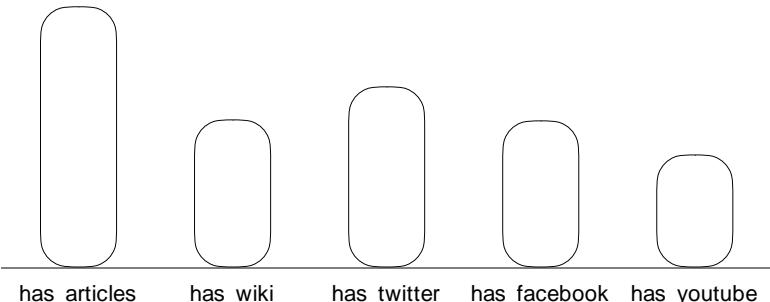


Dataset

<https://mediabiasfactcheck.com/>



864 annotated media sources



Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James R. Glass, Preslav Nakov:
What Was Written vs. Who Read It: News Media Profiling Using Text Analysis and Social Media Context. ACL 2020: 3364-3374

MEDIA BIAS/FACT CHECK

The Most Comprehensive Media Bias Resource

New York Times



Factual Reporting	
Very High	
HIGH	
Mostly Factual	
Mixed	
Low	
Very Low	

History

The *New York Times* (sometimes abbreviated to NYT) is an American daily newspaper, founded and continuously published in New York City since September 18, 1851 by The New York Times Company.

The *New York Times* was originally founded by U.S. journalist and politician [Henry Jarvis Raymond](#) and former banker [George Jones](#). Henry Jarvis Raymond is the founding editor and according to a [NYT article](#) Raymond was the political power behind the *New York Times*. "We will be Conservative, in all cases where we think Conservatism is good for the public; — and we will be Radical in everything which may seem to us to require radical treatment and radical reform."

The NYT was initially published by Raymond and Jones & Company. Raymond was also prominent in forming the [Republican Party](#) during the 1850s. George Jones who was a banker and politician solicited funds to launch the newspaper. Another early investor to the paper was the first president of [Wells Fargo & Company](#), and the founder of the United States Express Company, and director of the American Express Co., [Edwin D. Morgan](#). The *New York Times* was purchased by Adolph Ochs in 1896. Ochs had one daughter [Iphigenia Bertha Ochs](#) and she married to [Arthur Hays Subitzer](#), who became publisher of the NYT after Ochs died. [Dean Baquet](#) is the current Executive Editor.

[Read our profile on United States government and media.](#)

Funded by / Ownership

The *New York Times*, is controlled by the Ochs-Subitzer family through Class B shares. Since 1967, the company has been listed on the New York Stock Exchange under the symbol NYT. Class B shares are those that are [held privately](#). The owner and publisher of the *New York Times* is The New York Times Company and the Chairman is [Arthur Gregg "A.G." Sulzberger](#), succeeding his father Arthur Ochs Sulzberger Jr. He is the sixth member of the Ochs-Sulzberger family to serve as publisher since the newspaper was purchased by Adolph Ochs in 1896.

[Mark Thompson](#) became president and chief executive officer of The New York Times Company in 2012 and the current board of directors can be found [here](#).

According to Forbes, during the financial crisis of 2008, Mexican telecom mogul [Carlos Slim Helú](#) loaned The *New York Times* Company \$1 billion to help it weather the market collapse, and as of 2017, [Gerry Smith](#) is the chairman of The *New York Times*. According to a [Bloomberg article](#) President Trump accused Ochs of using his ownership stake in the Times to influence the paper's negative coverage of him. The Times and a representative for Ochs deny this claim. The *New York Times* is funded through advertising and subscription fees.

Analysis / Bias

The *New York Times'* coverage includes News (World News, National News, Business News) Opinion Pieces, Editorials, Arts, Movies, Theater, Travel, NYC Guide, Food, Home & Garden, and Fashion & Style.

A [Politico article](#) states Donald Trump has frequently criticized the *New York Times* on his [Twitter account](#). Trump labels the NYT as "fake news", "naive", "dumb", and "the failing New York Times". However, [CNBC report](#) the company has shown strong growth by adding 157,000 new digital subscriptions in the last quarter of 2017.

The NYT looks at the issues from a progressive perspective and is regarded as "liberal". According to a [Pew Research Center's media polarization report](#) "the ideological Placement of Each Source's Audience" places the audience for the *New York Times* as "consistently liberal". Further, since 1960 The *New York Times* has only endorsed Democratic Presidential Candidates.

In review, the NYT utilizes emotionally loaded language in their headlines such as "[Trump Again Faultily Blames Democrats for His Separation Tactic](#)" and "[A Financial's Profit-Minded Mission to Open a Chamber of Detention for Immigrants](#)". However, they use credible sources such as [Fox News](#), [Breitbart](#), [The Hill](#), [Politico](#), [The Washington Post](#). Story selection is typically balanced, however wording tends to lean left in most cases.

A [factual search](#) shows the *New York Times* has made false claims in reporting, but always makes corrections to their stories as soon as new information is available. Further, failed fact checks occurred on Op-Ed pages and not straight news reporting.

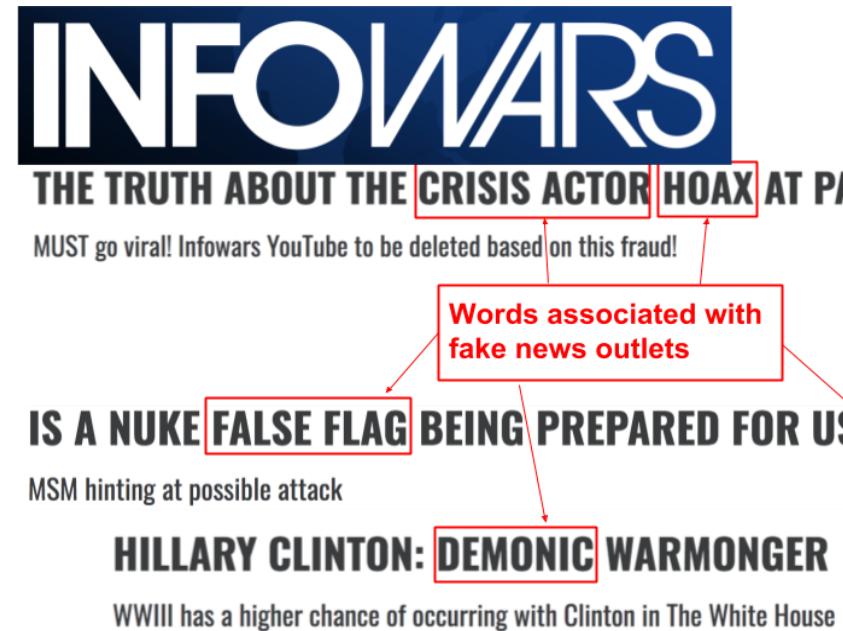
- We have a lot of research associated with high B.M.I.s. But correlation doesn't prove causation, and this is a very important body of research showing that weight stigma and weight cycling can explain more if any of the associations we see between higher weights and poor health outcomes." — [MUSTLY FALSE](#)

Overall, we rate the *New York Times* Left-Center based on word and story selection that moderately favors the left, but highly factual and considered one of the most reliable sources for news information due to proper sourcing and well respected journalists/editors. The failed fact checks that occurred were on Op-Ed's and not straight news reporting. (5/18/2016) Update (M. Hulme 5/18/2019)

Source: <https://www.nytimes.com/>

Article: Title & Body

- Connection between title vs. body
- Linguistic structure: function words, pronouns
- Length & complexity
- Sentiment
- Bias
- Subjectivity
- Topic
- Morality



Wikipedia

- Has Wikipedia page?
- Embedding for the text in
 - Infobox
 - Summary
 - Content
 - Categories
 - Table of Contents

National Review

From Wikipedia, the free encyclopedia

This article is about the American magazine. For other uses, see National Review (disambiguation).

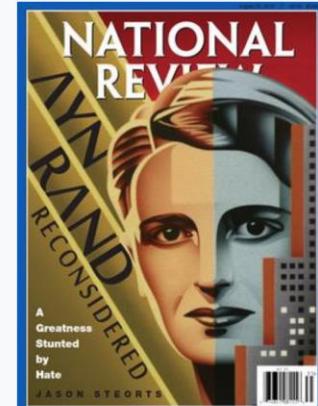
National Review (NR) is an American semi-monthly conservative editorial magazine focusing on news and commentary pieces on political, social, and cultural affairs. The magazine was founded by the author William F. Buckley Jr. in 1955.^[3] It is currently edited by Rich Lowry.

Since its founding, the magazine has played a significant role in the development of conservatism in the United States, helping to define its boundaries^[3] and promoting fusionism while establishing itself as a leading voice on the American right.^{[3][4][5]}

The online version, **National Review Online**, is edited by Charles C. W. Cooke and includes free content and articles separate from the print edition.^[6]

Contents <small>[hide]</small>
<ul style="list-style-type: none">1 History<ul style="list-style-type: none">1.1 Background1.2 Early years1.3 Contributors1.4 Mission to conservatives1.5 Defining the boundaries of conservatism1.6 After Goldwater2 Political views and content<ul style="list-style-type: none">2.1 Donald Trump

National Review



National Review cover for August 30, 2010

Editor	Rich Lowry
Categories	Editorial magazine, conservatism
Frequency	Biweekly

Twitter

- **Counts:** friends, statuses, favorites
- **Has Twitter account?**
- **Has location information?**
- **Verified?**
- **Years in existence**
- **Text description**



The image shows a Twitter profile for the user '@loki_jcrusader'. The profile picture features an eagle with its wings spread, set against a background of the American flag's stars and stripes. The bio reads: 'An Army Combat Vet ('99 to present). NCO. Crusader. Republican. Take back our country. #MAGA America First!'. It includes location 'USA' and joined date 'Joined February 2012'. The stats at the top right are: Tweets 645, Following 404, Followers 333, Likes 843. A 'Follow' button is visible. Below the profile, there are three tabs: 'Tweets', 'Tweets & replies', and 'Media'. The first tweet shown is from Dan Bongino (@dbongino) on July 22, 2022, which has been retweeted by the user. The tweet content is: 'Dear Media Hacks, Now that the FISA application has been released, & you've been completely humiliated for promoting the Russian collusion hoax, may I suggest that you all collectively submit letters of resignation? Your reputations are forever tarred. Thanks for considering.' The tweet has 1.5K likes, 17K retweets, 44K replies, and 0 quotes.

URL

- **Length**
- **Use of special characters**
 - digits
 - dashes
 - underscores
- **Use of https**
- **Is it hosted on a blog platform?**
- **Can the URL be chopped into meaningful words?**
- **What is the top-level domain?**

<http://www.angrypatriotmovement.com/>

<http://alternativemediasyndicate.com>

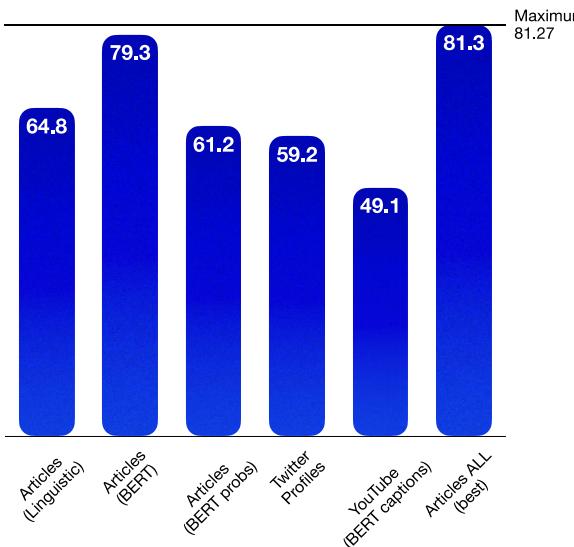
<http://100percentfedup.com/>

<http://abcnews.com.co>

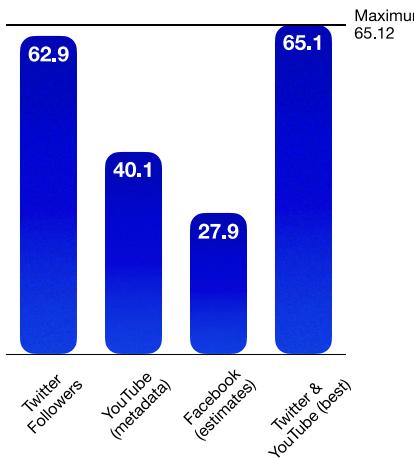
Results

Political Ideology Prediction

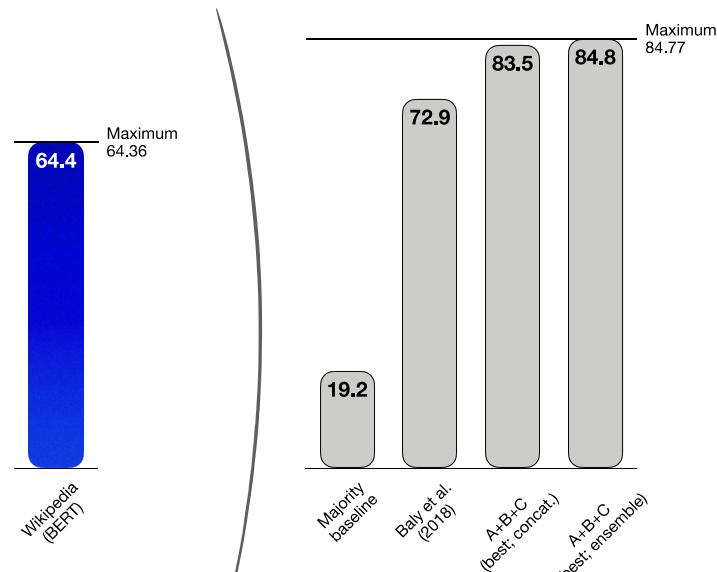
Results are in macro-F1
(accuracy shows similar trend)



A. What was written



B. Who read it



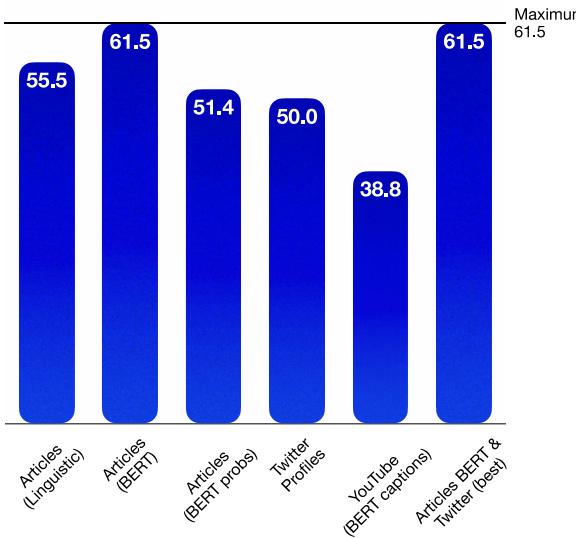
C. What's written about it

OVERALL

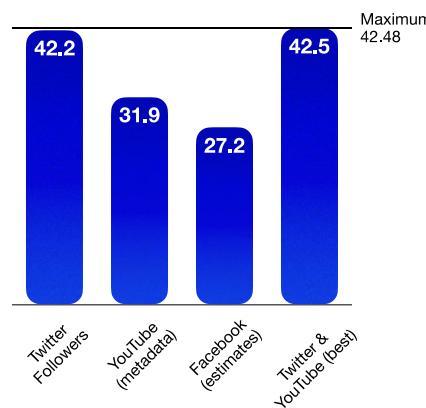
Results

Factuality Prediction

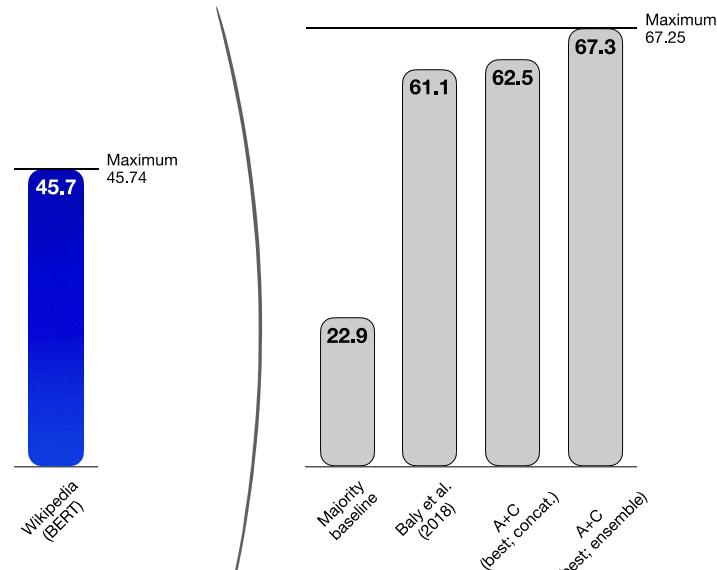
Results are in macro-F1
(accuracy shows similar trend)



A. What was written

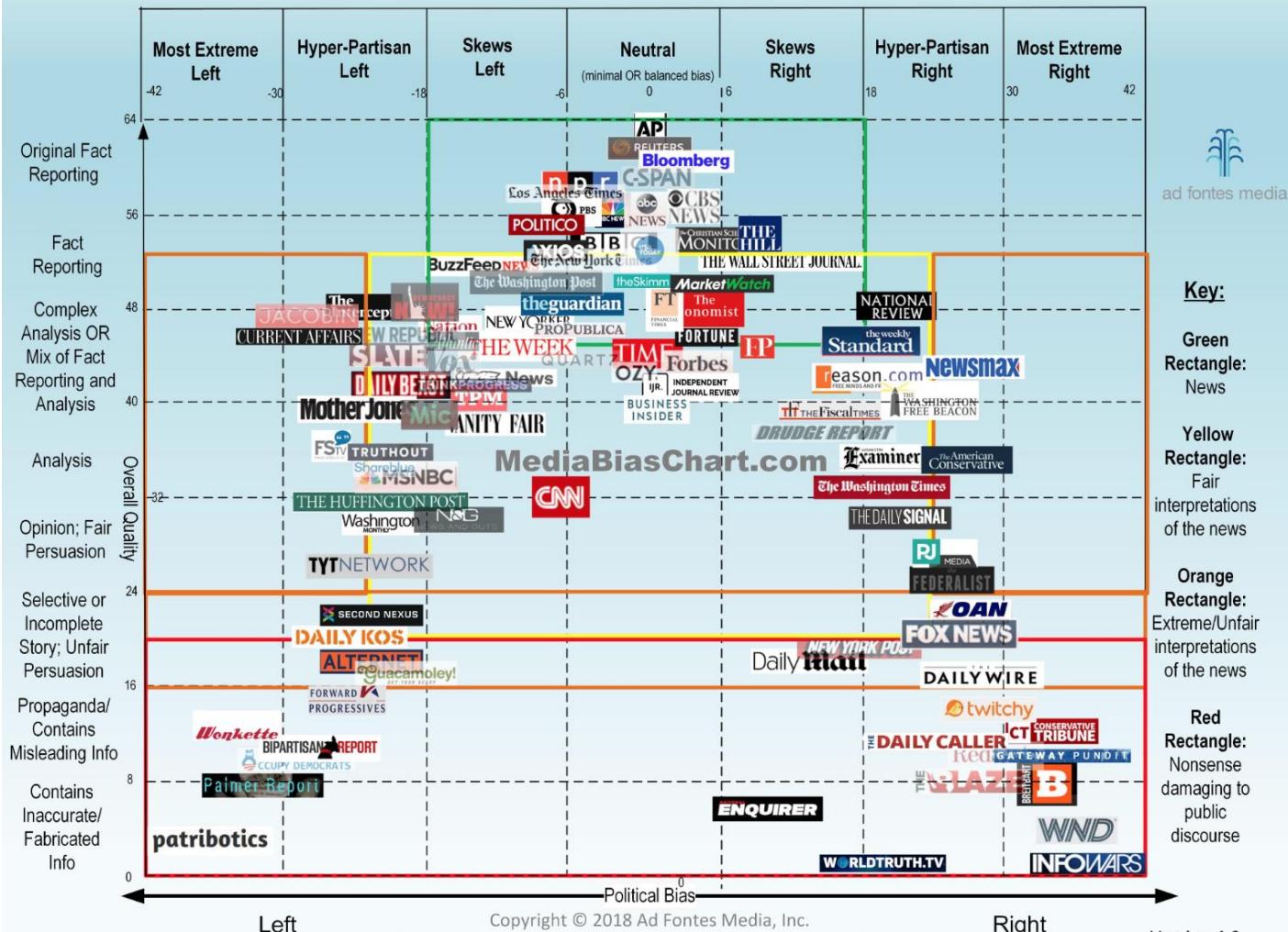


B. Who read it



C. What's written about it

OVERALL

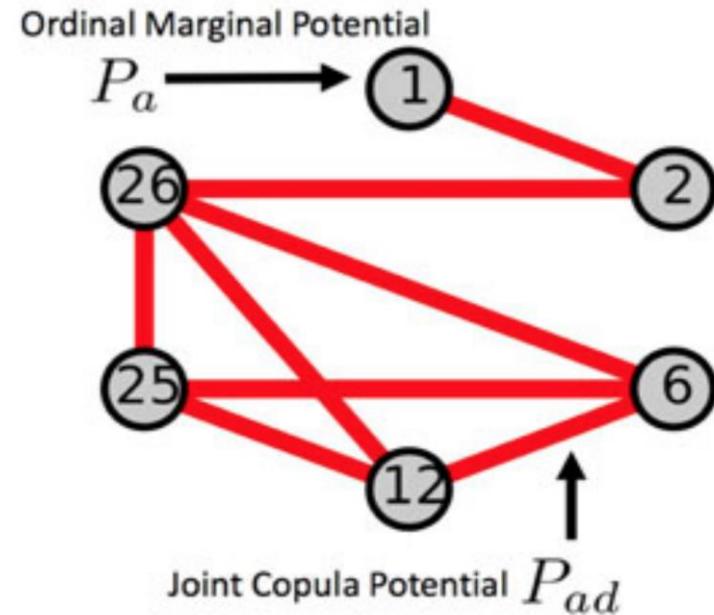


Copyright © 2018 Ad Fontes Media, Inc.
 Methodology, licenses and usage info available at mediabiaschart.com
 Created by Vanessa Otero
 twitter: @vloter0

Version 4.0
 August 2018

Multitask Ordinal Regression for Factuality & Bias

- Multi-task copula ordinal regression
- Learn jointly
 - Factuality
 - Left-center-right bias
 - 7-point
 - 5-point
 - 3-point
 - Centrality
 - Hyper-partisanship



Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James R. Glass, Preslav Nakov:

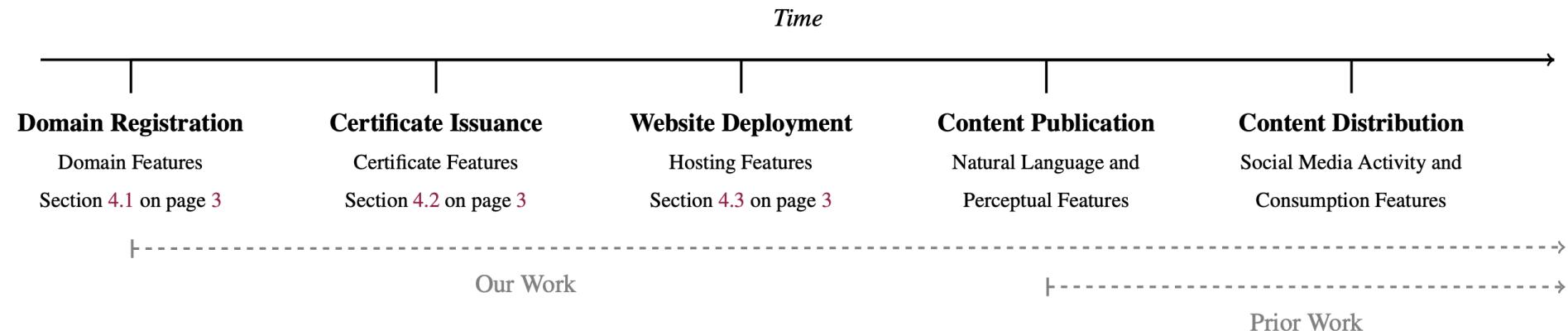
Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media.
NAACL-HLT (1) 2019: 2109-2116

Source Reliability

Using Stance wrt Fact-Checked Claims

$$tw(s) = \frac{\#articles_support_true + \#articles_refute_false}{\#total_articles}$$

Source Reliability Using Infrastructure



Source Reliability Using Infrastructure

Name	Category	Description	Rank	Data Type
News Keyword(s) in Domain	Domain	The domain name contains one or more keywords that imply it serves news (e.g., “herald,” “tribune,” or “chronicle”).	1	Boolean
Domain Name Length	Domain	The number of characters in the domain name.	3	Numeric
“News” in Domain	Domain	The domain name contains the specific keyword “news.”	8	Boolean
WHOIS Privacy	Domain	The domain registrant is using a WHOIS proxy service or registrar privacy option.	9	Boolean
Registrar Name	Domain	The organization with whom the domain was registered.	11	Categorical
Nameserver SLD	Domain	The second-level domain of the nameserver.	14	Categorical
Nameserver AS	Domain	The autonomous system of the nameserver’s IP address.	16	Categorical
Registrant Organization	Domain	The organization of the registrant.	17	Categorical
Registrant Country	Domain	The country of the registrant.	19	Categorical
Time Since Domain Registration	Domain	The time elapsed since the domain was originally registered.	21	Numeric
Domain Lifespan	Domain	The time period between the domain’s initial registration and expiration dates.	22	Numeric
Time to Domain Expiration	Domain	The time until the domain’s registration expires.	23	Numeric
Time Since Domain Update	Domain	The time since the domain’s configuration was updated.	25	Numeric
Nameserver Country	Domain	The country where the nameserver is located, using IP geolocation.	27	Categorical
Novelty TLD	Domain	The TLD is novelty (e.g., .news, .xyz, or .club).	29	Boolean
Digit in Domain	Domain	The domain name contains numeric characters.	30	Boolean
Hyphen in Domain	Domain	The domain name contains a hyphen.	31	Boolean
Domain Resolves	Domain	The domain name resolves to an IP address.	32	Boolean

Source Reliability Using Infrastructure

SAN Count	Certificate	The number of domains in the Subject Alternate Name extension field.	2	Numeric
SAN Contains Wildcard	Certificate	The Subject Alternate Name extension field contains a wildcard entry for a domain.	7	Boolean
Expired Certificate	Certificate	The certificate is expired.	10	Boolean
Certificate Available	Certificate	A certificate is configured at the domain (i.e., a certificate is provided during a TLS handshake on the HTTPS port).	12	Boolean
Self-signed Certificate	Certificate	The certificate is signed by the domain owner, not a CA.	13	Boolean
Domain-validated Certificate	Certificate	The domain owner has obtained a certificate from a CA through domain validation rather than organization validation or extended validation.	18	Boolean
Certificate Issuer Name	Certificate	The organization or individual who issued the certificate.	24	Categorical
Certificate Issuer Country	Certificate	The country where the certificate was issued.	26	Categorical
Certificate Lifetime	Certificate	The certificate's period of validity.	28	Numeric
WordPress Plugins	Hosting	WordPress plugins used by the website.	4	Categorical
Website AS	Hosting	The autonomous system of the website's IP address.	5	Categorical
WordPress CMS	Hosting	The website uses WordPress as its content management system.	6	Boolean
WordPress Theme	Hosting	The WordPress theme used by the website.	15	Categorical
Website Country	Hosting	The country where the website is located, using IP geolocation.	20	Categorical
Website Available	Hosting	A website is hosted at the domain (i.e., content is returned in response to an HTTP request for the base URL, following redirects).	33	Boolean

NELA-GT-2018: Source Reliability Data

Section	Description	(NewsGuard points)	Coloring
NewsGuard	1. Does not repeatedly publish false content	(22.0)	
	2. Gathers and presents information responsibly	(18.0)	
	3. Regularly corrects or clarifies errors	(12.5)	
	4. Handles the difference between news and opinion responsibly	(12.5)	
	5. Avoids deceptive headlines	(10.0)	
	6. Website discloses ownership and financing	(7.5)	
	7. Clearly labels advertising	(7.5)	
	8. Reveals who's in charge, including any possible conflicts of interest	(5.0)	
	9. Provides information about content creators	(5.0)	
	10. Aggregated score computed from 1-9		-
	11. Column 10 thresholded at 60 points		
Pew Research Center	12. Trust from consistently-liberals		
	13. Trust from mostly-liberals		
	14. Trust from mixed groups		
	15. Trust from mostly-conservatives		
	16. Trust from consistently-conservatives		
	17. Aggregated trust from 12-16		

NELA-GT-2018: Source Reliability Data

Wikipedia	18. Existence of source on Wikipedia's list of fake news sources	█
Open Sources	19. Marked reliable	█
	20. Marked blog	█
	21. Marked clickbait	█
	22. Marked rumor	█
	23. Marked fake	█
	24. Marked unreliable	█
	25. Marked biased	█
	26. Marked conspiracy	█
	27. Marked hate speech	█
	28. Marked junk science	█
	29. Marked political	█
	30. Marked satire	█
	31. Marked state news	█
Media Bias / Fact Check	32. Factual reporting from 5 (good) down to 1 (bad)	█ █ █ █ █
	33. Special label; conspiracy, pseudoscience or questionable source (purple), and satire (orange)	█ █
	34. Political leaning / bias from left to right.	█ █ █ █ █

NELA-GT-2018: Source Reliability Data

Allsides	35. Political leaning / bias	
BuzzFeed	36. Political leaning / bias, but only left and right	
PolitiFact	37. Has brought story labelled as "pants on Fire!"	
	38. Has brought story labelled as false	
	39. Has brought story labelled as mostly false	
	40. Has brought story labelled as half-true	
	41. Has brought story labelled as mostly true	
	42. Has brought story labelled as true	
Alexa Ranking	The Alexa ranking of the source.	Numerical
# Articles	The number of articles collected from the source.	Numerical
First Observed	The date of first articles collected from the source.	dd-mm-yyyy

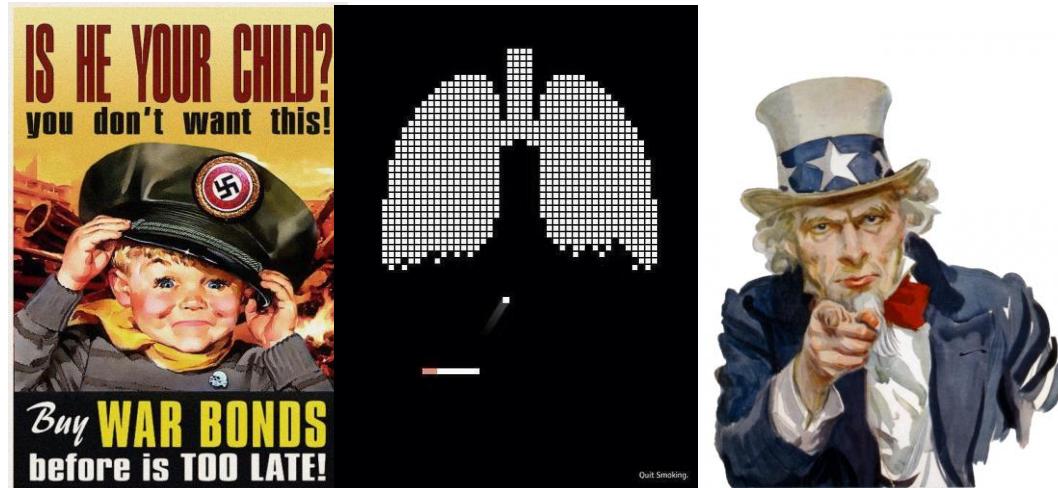
Online Propaganda Detection



Propaganda: A Definition

- “Communications that **deliberately** misrepresent symbols, appealing to emotions and prejudices and bypassing rational thought, to **influence** its audience **towards a specific goal**”*

- Not all influence has ill intent
- Avoid detection
- Pursue an agenda



*definition re-elaborated from *Institute for Propaganda Analysis (Ed.). (1938). How to Detect Propaganda. In Propaganda Analysis. Volume I of the Publications of the Institute for Propaganda Analysis (pp. 210–218).*

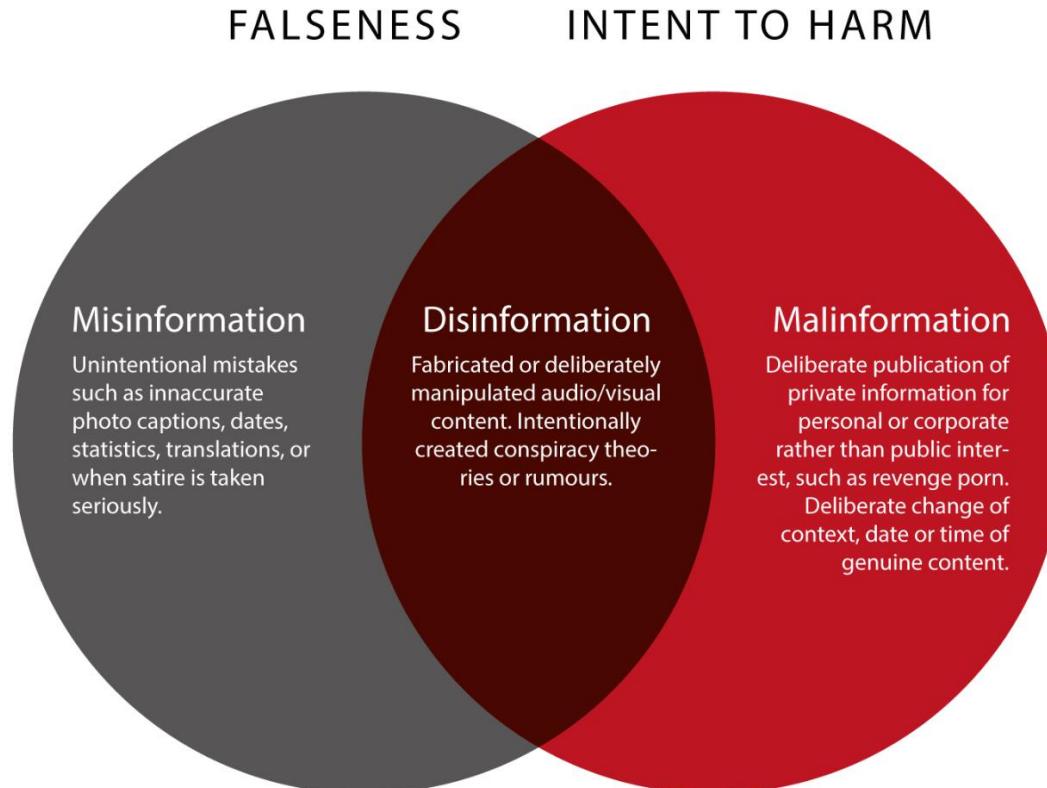
Propaganda: A Historical Perspective

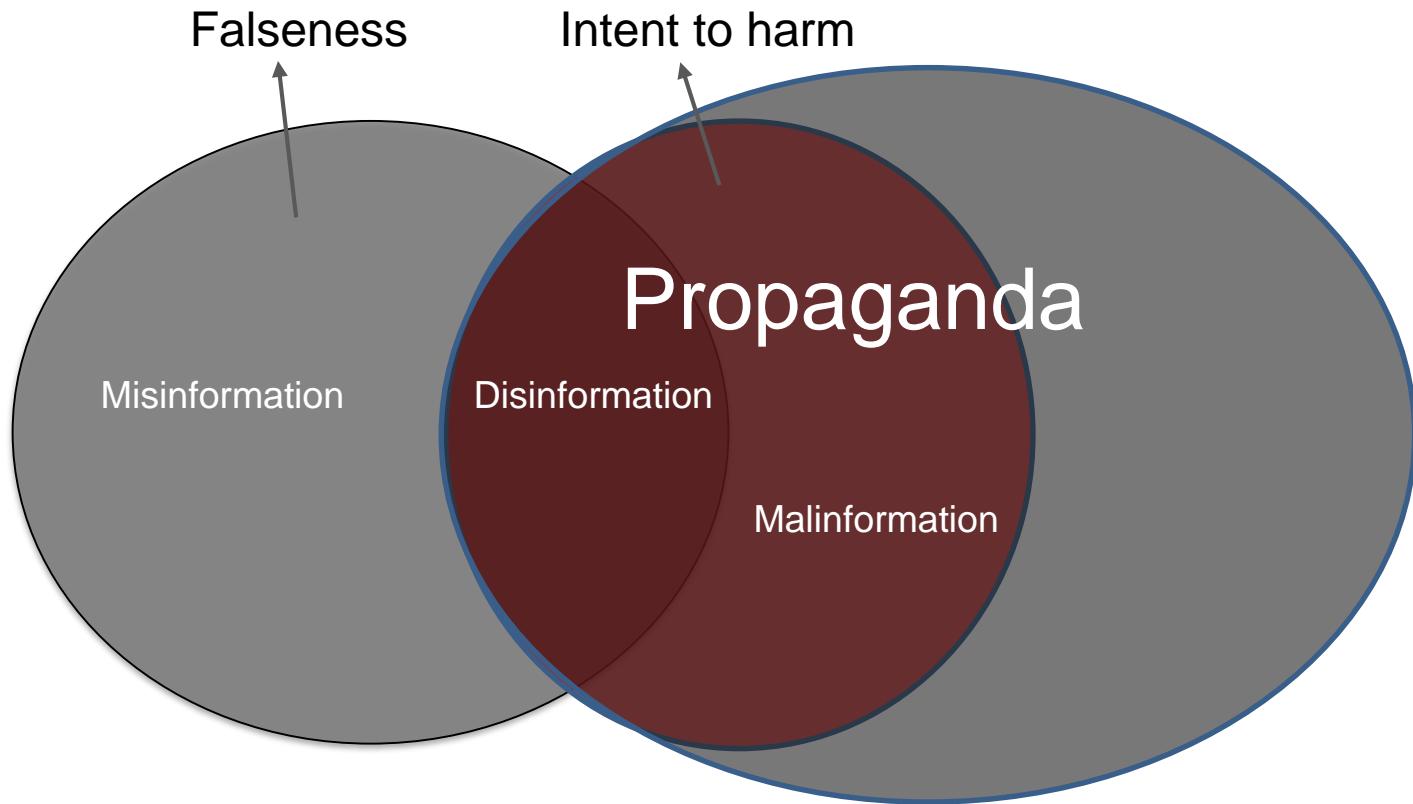


Propaganda campaigns in the pre-Internet era:

- Control of mass media
- Closed Borders (non-anonymous campaigns)
- Require massive resources

TYPES OF INFORMATION DISORDER

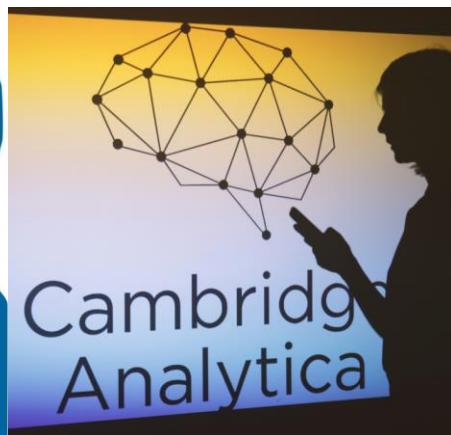




- Propaganda may not have ill-intent

Computational Propaganda

- “The rise of the Internet [...] has opened the **creation and dissemination of propaganda messages**, which were once the province of states and large institutions, to a **wide variety of individuals and groups**.”



Computational Propaganda: Technical Considerations

- Different skills needed
 - **Create persuasive messages that go undetected** (Relatively novel research topic, limited number of research papers available)
 - **Create bots to disseminate the messages**
 - Maximise audience reach
 - microprofiling
 - Monitor the results of the campaigns

Detecting Propagandistic Texts

- Propaganda Detection at Document Level
 - Datasets and Approaches
 - Future Research Directions
- Detection of Propaganda Techniques in Texts
 - Datasets and Approaches
 - Future Research Directions

Detecting Propagandistic Documents

- Annotated news articles for propaganda are scarce
 - journalistic initiatives to review media sources are available (media bias fact check)
 - articles are annotated via distant supervision

MEDIA BIAS/FACT CHECK
The Most Comprehensive Media Bias Resource



US NEWS & WORLD REPORT NEWS »

HOME AD FREE MEMBERSHIP TRANSPARENCY APPS/EXTENSIONS MBFC NEWS CHANGES/CORRECT

AD-FREE LOG IN

Left Left-Center Least Biased Right-Center Right Conspiracy-Pseudoscience Questionable Sources Pro-Science Satire



EN Questionable Sources

EU VS
DiSiNFO

News & Analysis

Disinfo Review

Disinfo Database

Studies & Reports

In The Media

Disinfo Quiz

Document-level Propaganda

Propaganda Detection can be framed as a text classification problem

Related Tasks:

- **Bias Detection¹:** ``Propaganda is another form of extreme bias''
- **Hyperpartisanship², Fake News and Satire Detection**

1. André Ferreira Cruz, Gil Rocha, Henrique Lopes Cardoso, Henrique Lopes Cardoso. On document representations for detection of biased news articles. SAC '20: Proceedings of the 35th Annual ACM Symposium on Applied Computing, March 2020 Pages 892–899

2. Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 Task 4: Hyperpartisan News Detection. In Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval 2019). 829–839.

Propaganda Detection as Text Classification

- **Task:** differentiating real news from satire, hoaxes, and propaganda.
- **Corpus (TSHP-17):** ~22K documents from the English Gigaword (real news) and from seven unreliable news sites (3 satire, 2 hoax, 2 propaganda).
- **Representation:** word n-grams, with $n \in [1, 3]$.
- **Model:** max entropy with L_2 regularization.

Hannah Rashkin et al. *Truth of varying shades: Analyzing language in fake news and political fact-checking*. In EMNLP, pages 2931–2937, 2017

Experiments

kind	sources	articles	training	dev	test	length (tokens)
Trusted	4*	5,750	3,997	1,003	750	522±429.13
Satire	3	5,750	3,981	1,019	750	324±276.31
Hoax	2	5,750	4,014	986	750	262±300.92
Propaganda	2	5,330	3,670	910	750	1,047±1,156.87
Total	11	22,580	15,662	3,918	3,000	529±705.34

Sources	Trusted	Gigaword News*
	Satire	The Onion • The Borowitz Report • Clickhole
	Hoax	American News • DC Gazette
	Propaganda	The Natural News • Activist Report

- Gold labels obtained by distant supervision

- **Representation:**
word n-grams
- **In-domain data (dev):**
 - **F1: 94.48**
 - **Accuracy : 94.44**
- **Out-of-domain data (test):**
 - **F1: 69.26**
 - **Accuracy: 69.73**

Proppy: features

1. Lexical features

Lexicon	Sample words
Wiktionary	
Modal, Action, Manner Adverbs	truly, apparently, accidentally, deliberately
Comparative, Superlative Forms	higher, less, purest, worst
LIWC	
First Person Singular, Second Person	my, I, you, yours
Hear, Money, Negation, Number, See, Sexual, Swear	says, costs, can't, quarter, watch, gay, dumb
Strong/Weak Subjectives (Wilson)	anti-semites, extremist
Hedges (Hyland)	appears, approximately, perhaps
Assertives (Hooper)	admit, hypothesize, certain

For each of the lexicons, the total number of words in the article is a feature

Proppy: features

2. Vocabulary richness features

feature	computation
TTR. Type–token ratio	$ types / tokens $
Hapax legomena. Amount of tokens appearing once in a text.	$ types_i $
Hapax dislegomena. Amount of tokens appearing twice in a text	$ types_j $
Honore's R. Combination of types, tokens, and hapax legomenæ.	$\frac{100 \cdot \log(tokens)}{1 - hapax_legomena / types } \cdot 10^4 \sum_i i^2 types_k - tokens \over tokens ^2$
Yule's characteristic K. Combination of types appearing with different frequencies and tokens. The chance of a word to occur in a text to follow a Poisson distribution	

where $i = 1, j = 2$, and $k = [1, 2, \dots]$ are the different frequencies of types in the text.

Proppy: features

3. Readability features

feature	computation
Flesch–Kincaid grade level. US grade level necessary to understand a text.	$0.39 \cdot \frac{ tokens }{ syllables } + 11.9 \cdot \frac{ syllables }{ tokens } - 15.59$
Flesch reading ease. A scale in range [0, 100] representing the complexity of a text. The latter is the easiest	$206.835 - 1.015 \cdot \frac{ tokens }{ sentences } - 84.6 \cdot \frac{ syllables }{ tokens }$
Gunning fog index. Amount of the years of formal education necessary to understand a text.	$0.4 \left(\frac{ tokens }{ sentences } + 100 \cdot \frac{ tokens_c }{ tokens } \right)$

$tokens_c$ stands for complex tokens; those with three syllables or more.

4. Style features:

- TF-IDF weighted Character 3-grams

Proppy: features

5. NELA* features :

- **Structure** : POS counts, linguistic (LIWC), clickbaits (Chakraborty et al. 2016).
- **Sentiment**: sentiment(Hutto and Gilbert 2014), emotion (Recasens et al. 2013) and (LIWC) , happiness (Mitchell et al. 2013).
- **Topic-dependent**: bio, relativity, personal concerns (LIWC)
- **Morality**: Moral (Haidt et al. 2009) and (Lin et al. 2017)
- **Bias**: bias (Recasens et al. 2013) and (Mukherjee et al. 2015), subjectivity (Pang et al. 2004).

**(B. Horne, S. Khedr, S. Adal, “Sampling the news producers: A large news and feature data set for the study of the complex media landscape” AAAI-18)*

Proppy: experiments and results

4-Class Classification (TSHP-17)

only with	in-domain		out-of-domain		all but	in-domain		out-of-domain	
	F ₁	Acc	F ₁	Acc		F ₁	Acc	F ₁	Acc
word <i>n</i> -grams	94.46	94.41	69.18	69.67	word <i>n</i> -grams	97.58	97.58	67.13	67.77
lexicon	59.17	59.92	47.48	48.00	lexicon	96.77	96.76	65.32	66.57
voc. richness	47.63	50.25	31.35	34.90	voc. richness	96.82	96.81	65.65	66.83
readability	35.17	39.43	26.19	29.13	readability	96.77	96.76	65.47	66.70
char <i>n</i> -grams	97.15	97.14	63.98	65.00	char <i>n</i> -grams	95.43	95.38	69.56	70.53
nela	87.07	87.06	63.81	65.33	nela	96.89	96.89	63.96	65.27
					all	96.77	96.76	65.35	66.60

2-Class Classification (prop./non-prop.)

Features	TSHP-17		TSHP-17 corpus out-of-domain
	in-domain	Features	
word <i>n</i> -grams	90.76	word <i>n</i> -grams	50.68
lexicon	68.74	lexicon	61.54
voc. richness	55.62	voc. richness	54.29
readability	40.16	readability	45.68
char <i>n</i> -grams	96.22	char <i>n</i> -grams	52.51
nela	82.27	nela	
word <i>n</i> -grams + char <i>n</i> -grams	97.21	word <i>n</i> -grams + char <i>n</i> -grams	64.00
char <i>n</i> -grams + lexicon	97.14	char <i>n</i> -grams + lexicon	63.66
char <i>n</i> -grams + nela	96.64	char <i>n</i> -grams + nela	52.89
readability + nela	82.30	readability + nela	53.66
char <i>n</i> -grams + lexicon + voc. richness + nela	96.97	char <i>n</i> -grams + lexicon + voc. richness + nela	64.14
word & char <i>n</i> -grams + lexicon + voc. richness + nela	97.10	word & char <i>n</i> -grams + lexicon + voc. richness + nela	63.47

Proppy: Corpus

- Qprop-18

Label	Sources	Articles	Train	Dev	Test	Length (tokens)
Propagandistic	10	5,737	4,021	575	1,141	1084.46 ± 890.59
Non-propagandistic	94	45,557	31,972	4,564	9,021	620.31 ± 518.92
Total	104	51,294	35,993	5,139	10,162	672.22 ± 590.98

- Collected using GDELT + MBFC

Results: Two-Class Classification on TSHP-17 and QProp-18

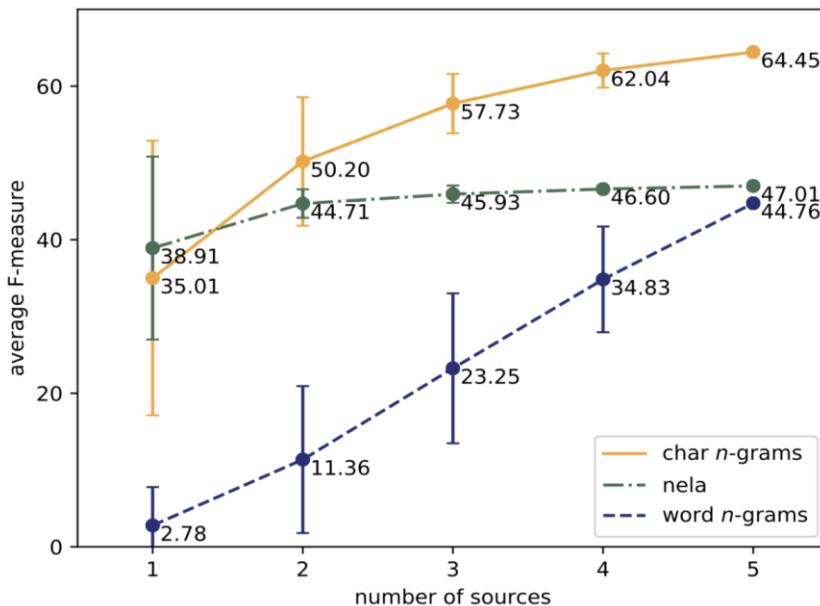
Features	TSHP-17	QProp	
	in-domain	Dev	Test
word n -grams	90.76	74.42	75.55
lexicon	68.74	46.55	44.87
voc. richness	55.62	29.45	29.72
readability	40.16	21.96	21.50
char n -grams	96.22	82.93	82.13
nela	82.27	54.60	50.98
word n -grams + char n -grams	97.21	78.37	79.01
char n -grams + lexicon	97.14	83.02	81.94
char n -grams + nela	96.64	83.21	82.75
readability + nela	82.30	75.34	76.83
char n -grams + lexicon + voc. richness + nela	96.97	83.17	82.89
word & char n -grams + lexicon + voc. richness + nela	97.10	79.04	79.50

Qprop-18: 2-class classification problem, examples from the same source are found both in dev and in test

Experiment: Learning Propaganda vs. Learning the Source

Test set (fixed): selected all examples from 5 propagandistic sources

Training: randomly selecting n propagandistic sources, random sampling the non-propagandistic ones such that the distribution is similar to the one of the full dataset



Conclusions and Future Directions

- Distant supervision may induce the learning algorithm to model the sources of the media
 - Representations based on writing style and complexity can generalize better than current approaches based on word-level representations
 - Adversarial approaches^{1,2?}
- Lack explainability

¹ Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1):2096–2030.

² Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR '15*, pages 815–823, Boston, MA, USA

Detection of Propaganda Techniques

- Novel research direction: detection of propaganda techniques
- Broader effects?
- Telling people what to think does not usually change their mind, warning them they might be manipulated might
- Disinformation is due to lack of analytical thinking: “Lazy, not biased...”



Fine-Grained Propaganda Analysis

- Propaganda is conveyed through a series of rhetorical and psychological techniques
- The set of propaganda techniques differs between scholars¹, from
 - 7 of Miller² to
 - ~70 in Wikipedia³

- Different skills needed
 - Create persuasive messages that go undetected
 - ...

¹ Robyn Torok. 2015. *Symbiotic radicalisation strategies: Propaganda tools and neuro linguistic programming*. In *Proceedings of the Australian Security and Intelligence Conference*, pages 58–65, Perth, Australia.

² Clyde R. Miller. 1939. *The Techniques of Propaganda*. From “How to Detect and Analyze Propaganda,” an address given at Town Hall. The Center for learning.

³ http://en.wikipedia.org/wiki/Propaganda_techniques

Fine-Grained Propaganda Analysis

- Propaganda is conveyed through a series of rhetorical and psychological techniques

A circular diagram illustrating 15 propaganda techniques, each represented by a colored word cloud:

- reductio ad Hitlerum (blue)
- flag-waving (purple)
- red herring (blue)
- minimisation (purple)
- exaggeration (red)
- black-and-white fallacy (blue)
- appeal to prejudice (purple)
- intentional vagueness (red)
- cognitive dissonance (blue)
- loaded language (purple)
- name calling (blue)
- obfuscation (purple)
- appeal to authority (red)
- causal oversimplification (blue)
- labeling (purple)
- bandwagon (red)
- whataboutism (purple)
- thought-terminating cliches (red)

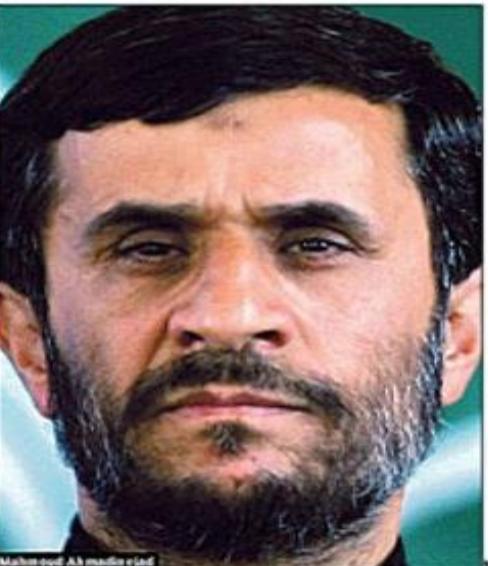
DAILY NEWS

50¢

NEW YORK'S HOMETOWN NEWSPAPER

NYDailyNews.com

THE EVIL HAS LANDED



Mahmoud Ahmadinejad

STORIES
ON
PAGES 4-7
EDITORIAL
PAGE 22

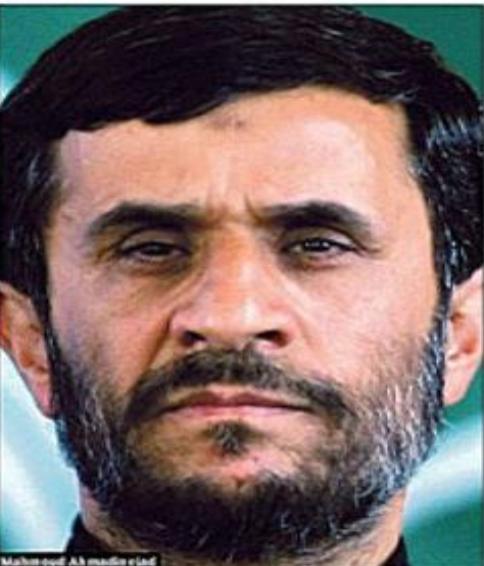
DAILY NEWS

50¢

NEW YORK'S HOMETOWN NEWSPAPER

NYDailyNews.com

THE EVIL HAS LANDED

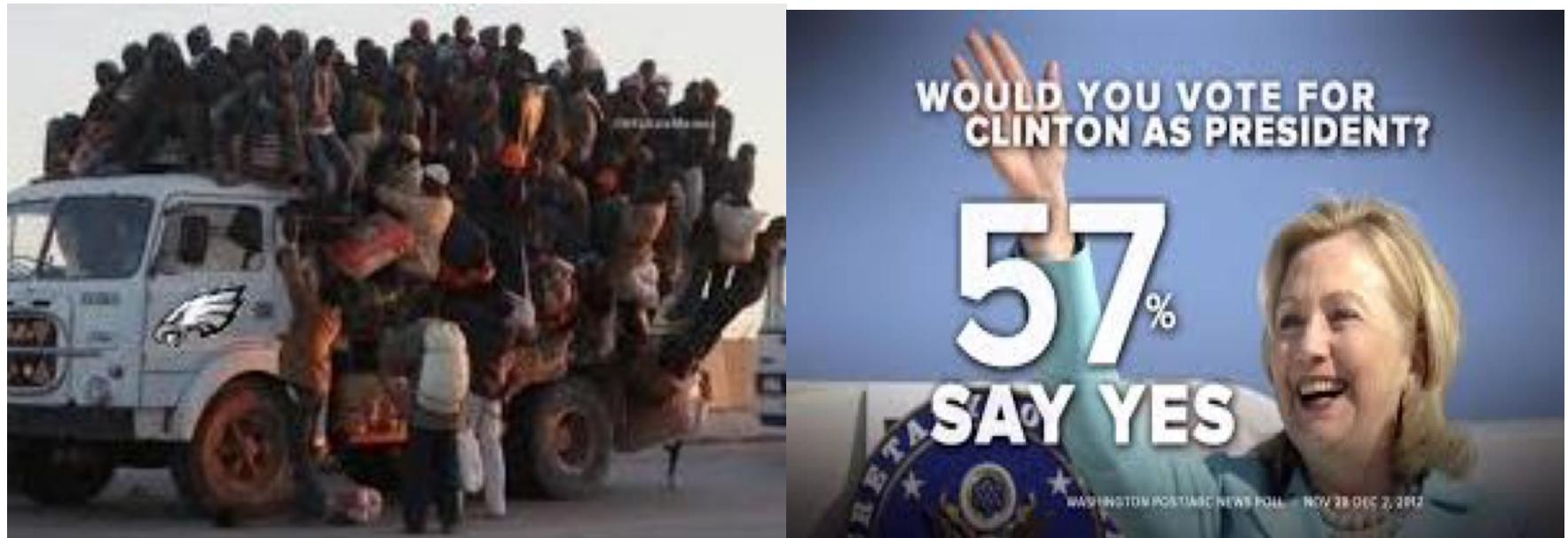


Mahmoud Ahmadinejad

STORIES
ON
PAGES 4-7
EDITORIAL
PAGE 22

Name Calling

Bandwagon



Attempting to persuade the target audience to join in and take the course of action because "everyone else is taking the same action."



"We are in the middle of the sixth mass extinction, with more than 200 species getting extinct every day"

Greta Thunberg



"We are in the middle of the sixth mass extinction, with more than 200 species getting extinct every day"

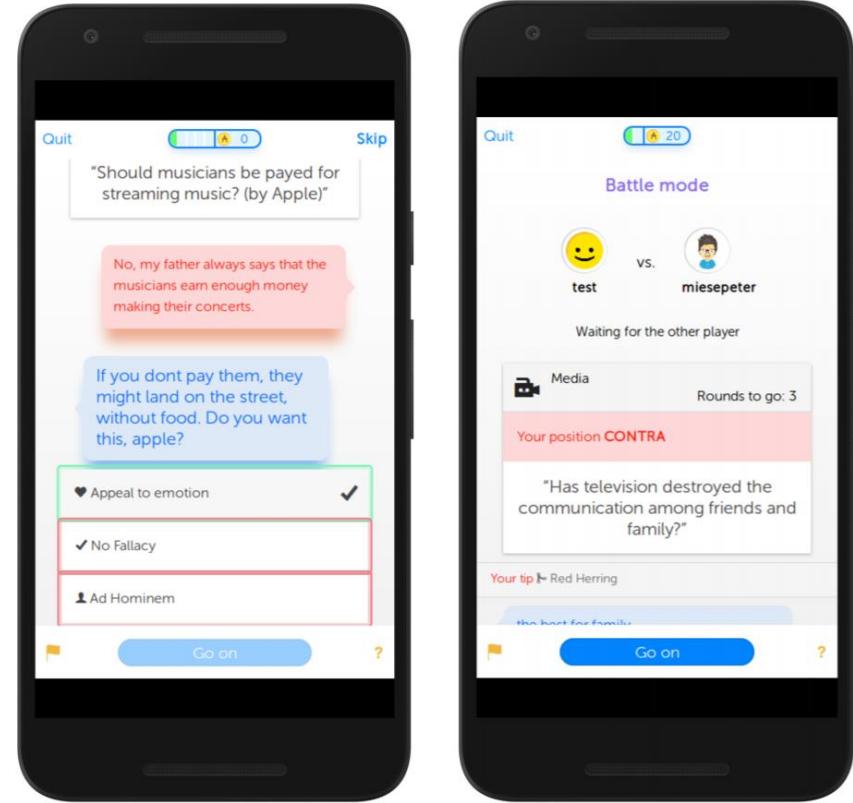
Greta Thunberg

Appeal to Fear

Propaganda Techniques Corpora

Argotario

- A game to educate people to recognize and create fallacies.
Users learn
 - to recognize fallacies in others' arguments
 - to write fallacious arguments
- A byproduct of Argotario is a corpus with 1.3k arguments annotated with five fallacies, including *ad hominem*, *red herring* and *irrelevant authority*
 - in English and German



Ivan Habernal, Raffael Hannemann, Christian Pollak, Christopher Klamm, Patrick Pauli, and Iryna Gurevych. 2017. Argotario: Computational argumentation meets serious games. In Proceedings of CLEF 2017, pages 7–12, Copenhagen, Denmark.

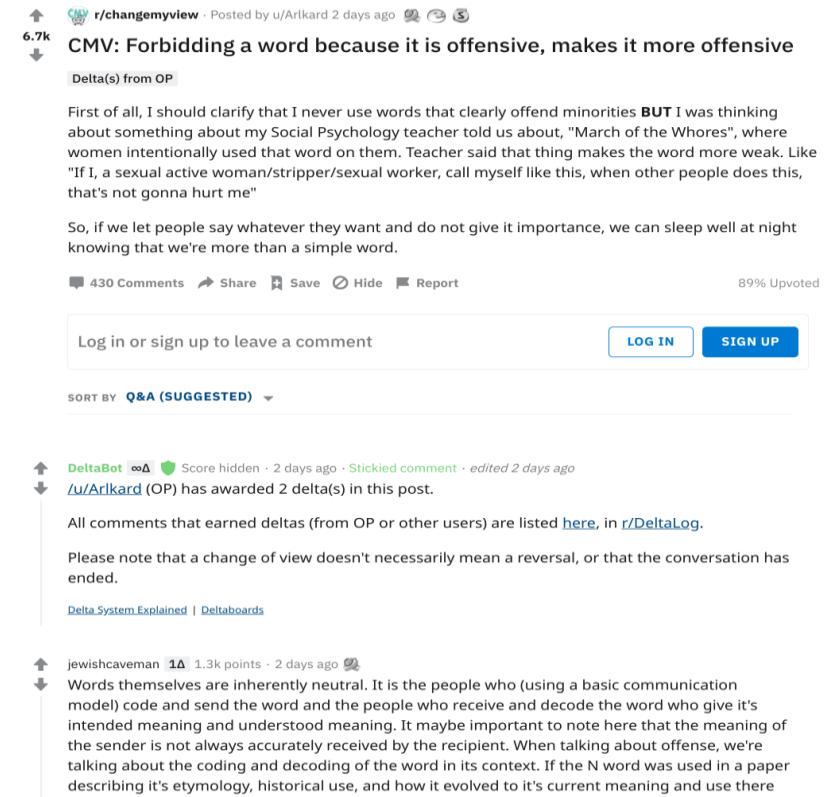
Change My View Corpus

- Change My View (CMV) is an online moderated platform for argumentation posted on reddit

- A user posts an opinion
 - other users provide their arguments to change his/her point of view
 - the original poster award points to the convincing arguments

- Moderators remove all *ad hominem* (attack to the person) arguments
 - authors collected 3 396 threads

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. Before name-calling: Dynamics and triggers of *ad hominem* fallacies in web argumentation. In Proceedings of NAACL-HLT '18, pages 386–396, New Orleans, LA, USA.



A screenshot of a post from the subreddit r/changemyview. The post has 6.7k upvotes and 430 comments. The title of the post is "CMV: Forbidding a word because it is offensive, makes it more offensive". The post content discusses the psychological effect of banning words like "whore" and how it can make the word stronger. Below the post, there is a comment section with a login/signup button, a sorting dropdown set to "Q&A (SUGGESTED)", and a note about DeltaBot awarding 2 delta(s). There are also replies from users like jewishcaveman and u/Arkkard.

6.7k r/changemyview · Posted by u/Arkkard 2 days ago

CMV: Forbidding a word because it is offensive, makes it more offensive

First of all, I should clarify that I never use words that clearly offend minorities BUT I was thinking about something about my Social Psychology teacher told us about, "March of the Whores", where women intentionally used that word on them. Teacher said that thing makes the word more weak. Like "If I, a sexual active woman/stripper/sexual worker, call myself like this, when other people does this, that's not gonna hurt me"

So, if we let people say whatever they want and do not give it importance, we can sleep well at night knowing that we're more than a simple word.

430 Comments Share Save Hide Report 89% Upvoted

Log in or sign up to leave a comment

SORT BY Q&A (SUGGESTED)

DeltaBot Score hidden · 2 days ago · Sticked comment · edited 2 days ago
u/Arkkard (OP) has awarded 2 delta(s) in this post.

All comments that earned deltas (from OP or other users) are listed [here](#), in [r/DeltaLog](#).

Please note that a change of view doesn't necessarily mean a reversal, or that the conversation has ended.

jewishcaveman 1Δ 1.3k points · 2 days ago

Words themselves are inherently neutral. It is the people who (using a basic communication model) code and send the word and the people who receive and decode the word who give it's intended meaning and understood meaning. It maybe important to note here that the meaning of the sender is not always accurately received by the recipient. When talking about offense, we're talking about the coding and decoding of the word in its context. If the N word was used in a paper describing it's etymology, historical use, and how it evolved to it's current meaning and use there

Change My View Corpus

- Ad hominem argument classification
- human upper bound estimated by re-annotating the data via crowdsourcing

- 10-fold cross-validation words are

Model	Accuracy
Human upper bound estimate	0.878
2 Stacked Bi-LSTM	0.782
CNN	0.810

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. In Proceedings of NAACL-HLT '18, pages 386–396, New Orleans, LA, USA.

The screenshot shows a Reddit post from the subreddit r/changemyview. The post has 6.7k upvotes and 430 comments. The title of the post is "CMV: Forbidding a word because it is offensive, makes it more offensive". The post includes a text block, a comment section, and a sidebar for logging in or signing up.

First of all, I should clarify that I never use words that clearly offend minorities **BUT** I was thinking about something about my Social Psychology teacher told us about, "March of the Whores", where women intentionally used that word on them. Teacher said that thing makes the word more weak. Like "If I, a sexual active woman/stripper/sexual worker, call myself like this, when other people does this, that's not gonna hurt me"

So, if we let people say whatever they want and do not give it importance, we can sleep well at night knowing that we're more than a simple word.

430 Comments Share Save Hide Report 89% Upvoted

Log in or sign up to leave a comment [LOG IN](#) [SIGN UP](#)

SORT BY [Q&A \(SUGGESTED\)](#)

DeltaBot Score hidden · 2 days ago · Stickied comment · edited 2 days ago
u/Arlkard (OP) has awarded 2 delta(s) in this post.

All comments that earned deltas (from OP or other users) are listed [here](#), in [r/DeltaLog](#).

Please note that a change of view doesn't necessarily mean a reversal, or that the conversation has ended.

[Delta System Explained](#) | [Deltabards](#)

jewishcaveman 1Δ 1.3k points · 2 days ago
Words themselves are inherently neutral. It is the people who (using a basic communication model) code and send the word and the people who receive and decode the word who give it's intended meaning and understood meaning. It maybe important to note here that the meaning of the sender is not always accurately received by the recipient. When talking about offense, we're talking about the coding and decoding of the word in its context. If the N word was used in a paper describing it's etymology, historical use, and how it evolved to it's current meaning and use there

Propaganda Techniques Corpus

536 news articles from 48 sources (450k words, 400 man hours to annotate it) annotated at fragment level with 18 techniques

		Stereotyping_name_calling_or_labeling
1	Manchin says Democrats acted like babies at the SOTU	
2	Democrat West Virginia Sen. Joe Manchin says his colleagues' refusal to stand or applaud during President Donald Trump's State of the Union speech was disrespectful and a signal that	
		Black-and-white Fallacy
		the party is more concerned with obstruction than it is with progress.
		Loaded_language
4	In a glaring sign of just how stupid and petty things have become in Washington these days, Manchin was invited on Fox News Tuesday morning to discuss how he was one of the only Democrats in the chamber for the State of the Union speech	
		Exaggeration
		Loaded_language
		not looking as though Trump killed his grandma.
6	As Manchin noted, many Democrats bolted as soon as Trump's speech ended in an apparent effort to signal	
		Exaggeration
		they can't even stomach being in the same room as the president

Propaganda Technique	inst	avg. length
loaded language	2,547	23.70 ± 25.30
name calling, labeling	1,294	26.10 ± 19.88
repetition	767	16.90 ± 18.92
exaggeration, minimization	571	45.36 ± 35.55
doubt	562	123.21 ± 97.65
appeal to fear/prejudice	367	93.56 ± 74.59
flag-waving	330	61.88 ± 68.61
causal oversimplification	233	121.03 ± 71.66
slogans	172	25.30 ± 13.49
appeal to authority	169	131.23 ± 123.2
black-and-white fallacy	134	98.42 ± 73.66
thought-terminating cliches	95	34.85 ± 29.28
whataboutism	76	120.93 ± 69.62
reductio ad hitlerum	66	94.58 ± 64.16
red herring	48	63.79 ± 61.63
bandwagon	17	100.29 ± 97.05
obfusc., int. vagueness, confusion	17	107.88 ± 86.74
straw man	15	79.13 ± 50.72
all	7,485	46.99 ± 61.45

Tasks

- **FLC** - detect the text-fragments in which a propaganda technique is used and identify the technique.
- **SI** is a lighter version of the task in which only the span has to be identified.
- **TC** - given a propagandistic span of text, identify the technique(s) in it - **SI + TC = FLC**
- **SLC** a binary task at sentence-

1	Manchin says Democrats acted like babies at the SOTU	Stereotyping name calling or labeling
2	Democrat West Virginia Sen. Joe Manchin says his colleagues' refusal to stand or applaud during President Donald Trump's State of the Union speech was disrespectful and a signal that	Black-and-white Fallacy
	the party is more concerned with obstruction than it is with progress.	
4	In a glaring sign of just how stupid and petty things have become in Washington these days, Manchin was invited on Fox News Tuesday morning to discuss how he was one of the only Democrats in the chamber for the State of the Union speech	Loaded language
	not looking as though Trump killed his grandma.	Exaggeration Loaded language
6	As Manchin noted, many Democrats bolted as soon as Trump's speech ended in an apparent effort to signal	Exaggeration
	they can't even stomach being in the same room as the president	

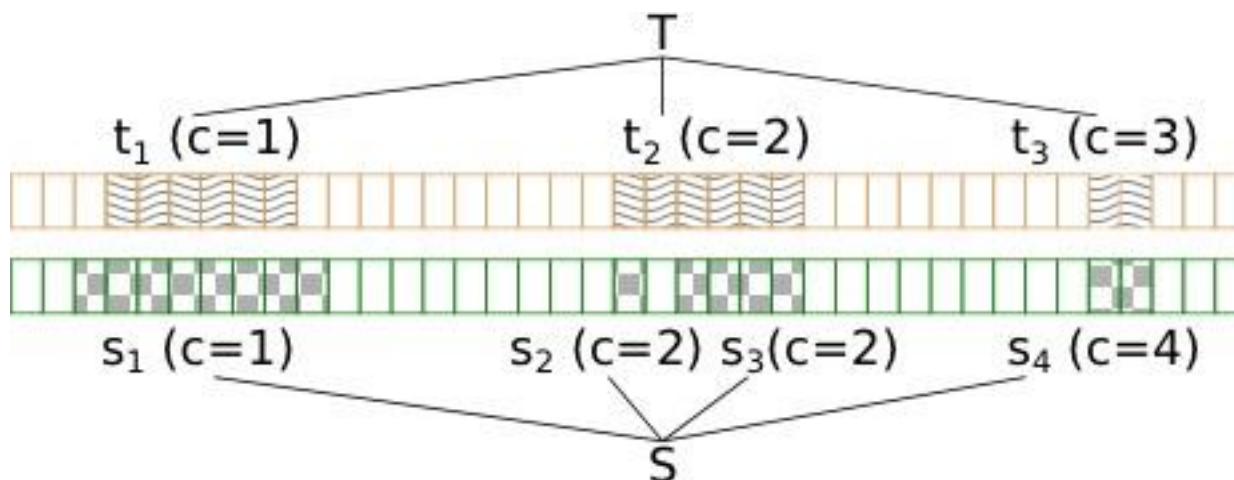
propaganda
non-propaganda
propaganda
propaganda
non-propaganda
non-propaganda

- 1 propaganda
- 2 non-propaganda
- 3 propaganda
- 4 propaganda
- 5 non-propaganda
- 6 non-propaganda

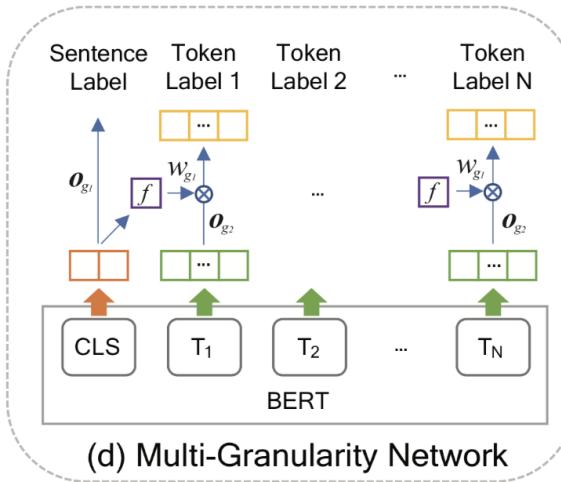
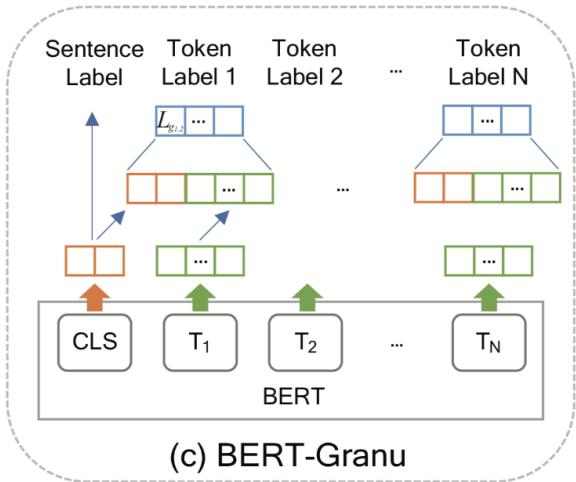
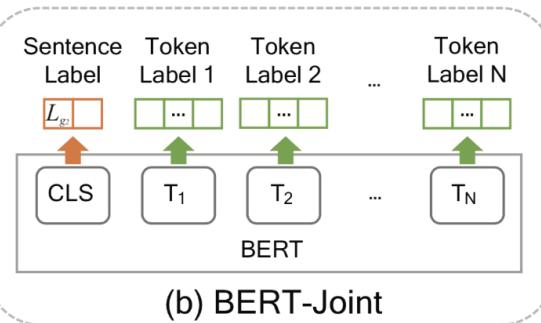
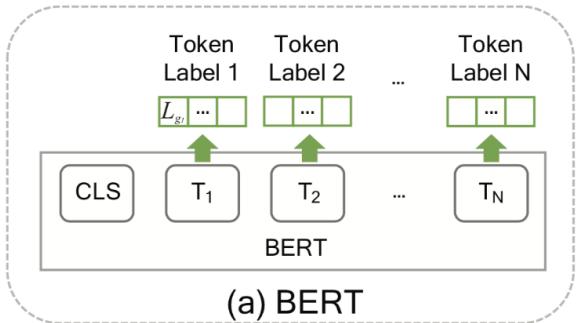
Evaluation Measures

- **SLC, TC**: standard F_1 measure
- **SI, FLC** - we adapted a measure for NER to account for overlapping gold spans

$$C(s, t, h) = \frac{|(s \cap t)|}{h} \delta(l(s), l(t)), \quad P(S, T) = \frac{1}{|S|} \sum_{\substack{s \in S, \\ t \in T}} C(s, t, |s|), \quad R(S, T) = \frac{1}{|T|} \sum_{\substack{s \in S, \\ t \in T}} C(s, t, |t|),$$



Baseline Models



Results: Fragment-Level

Model	Spans			Full Task		
	P	R	F ₁	P	R	F ₁
BERT	39.57	36.42	37.90	21.48	21.39	21.39
Joint	39.26	35.48	37.25	20.11	19.74	19.92
Granu	43.08	33.98	37.93	23.85	20.14	21.80
Multi-Granularity						
ReLU	43.29	34.74	38.28	23.98	20.33	21.82
Sigmoid	44.12	35.01	38.98	24.42	21.05	22.58

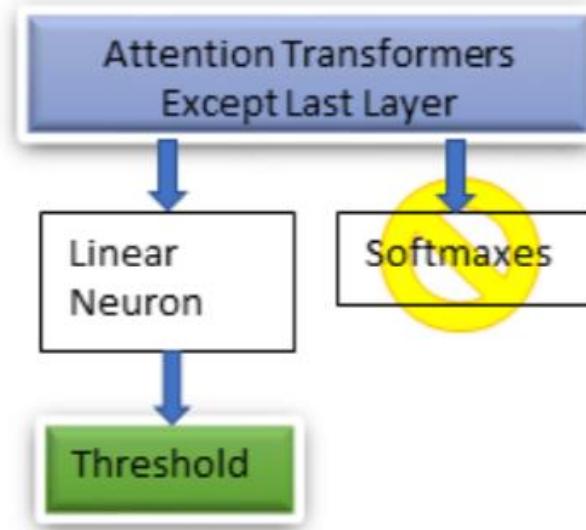
Results: Sentence-Level

Model	Precision	Recall	F1
All-Propaganda	23.92	1.00	38.61
BERT	63.20	53.16	57.74
BERT-Granu	62.80	55.24	58.76
BERT-Joint	62.84	55.46	58.91
MGN Sigmoid	62.27	59.56	60.71
MGN ReLU	60.41	61.58	60.98

A Demo of the system is available at <https://www.tanbih.org/prta>

Task SLC - NLP4IF'19 Shared Task

- Mapes et al. combines 12 head attention transformers models,
- Each model is trained with a different random seed to get a different view on the data
- Participated in the NLP4IF'19 shared task and outperformed all other approaches with an F1=63.23



Backpropagate Error

Task FLC - NLP4IF'19 Shared Task

- Yoosuf and Yang, 2019 used a 20-way word-level classification based on an uncased base model of BERT with 12 Transformer layers
- oversampling of the least represented classes proved to be crucial for the final performance.
- Overall F1 performance is **24.88**
- Analysis of the model shows that it pays special attention to adjectives and adverbs.

More approaches for the tasks SLC and FLC are described here:

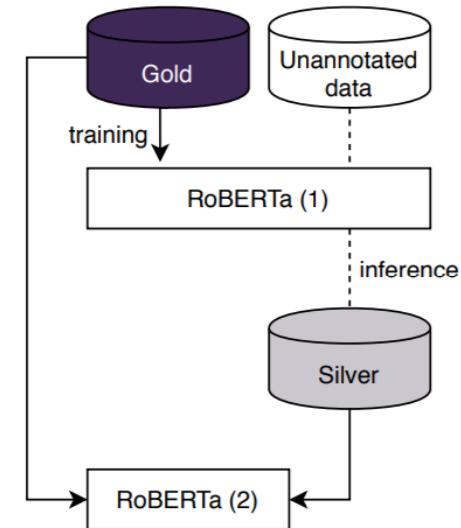
<https://www.aclweb.org/anthology/events/nlp4if-2019/>

Task SI - SemEval2020 Shared Task

- Each token is represented by
 - the encoding of a feedforward neural network (either BERT, GPT-2, XLNet, XLM, RoBERTa, or XLM-RoBERTa)
 - part of speech and named entity embeddings
- and fed to two bi-LSTM network predicting
 - the BIO tag and classification of each token (using task FLC annotations)
 - the sentence level classification (using task SLC annotations)
- Experimental result on test: **F1 = 51.74**

Task TC - SemEval2020 Shared Task

- Jurkiewicz et al., 2020 applied self-supervision using RoBERTa: they generated 500k annotated sentences and then used them for training
- The final classifier is an ensemble of models trained on the original corpus and a model trained also on silver data.
- They win the task with a score of **F1= 63.74**



Further approaches for tasks SI and TC are described here:
<https://propaganda.qcri.org/semeval2020-task11/workshop.html>

Conclusions and Future Work

- Propaganda detection, especially the detection of propaganda techniques is a novel task.
- Although results are encouraging, there is a significant margin for improvement
- More annotated data is needed (but annotations are costly!)
- Tackling each technique in isolation could be an effective way to advance the state of the art

A novel shared task has just started “Prediction of persuasion techniques in text and images at SemEval 2021”, in which the techniques need to be detected in the textual and visual content of memes.

<https://semeval.github.io/SemEval2021/tasks>

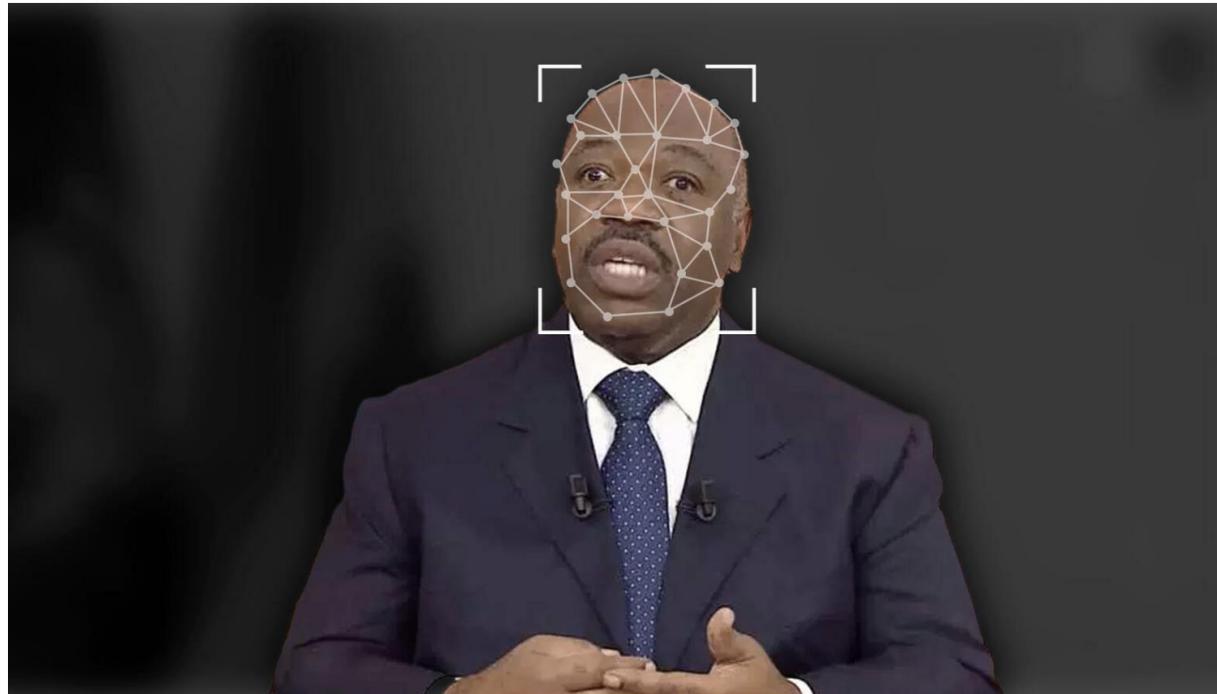
Neural Fake News

A Military Coup in Gabon Inspired by a Potential Deepfake Video is Our Political Future



Ty Joplin

Published May 8th, 2019 - 09:54 GMT



Ali Bongo (Youtube, Rami Khoury/Al Bawaba)

A **video** of Gabon's President, Ali Bongo, appeared on Jan 1, 2019, sparking a wave of confused reactions inside the country.

<https://www.albawaba.com/news/military-coup-gabon-inspired-potential-deepfake-video-our-political-future-1284760>

GANs for Fake Person Generation

<https://thispersondoesnotexist.com/>



Imagined by a GAN ([generative adversarial network](#))
[StyleGAN2](#) (Dec 2019) - Karras et al. and Nvidia
Don't panic. Learn how it works [\[1\]](#) [\[2\]](#) [\[3\]](#)
[Help](#) this AI continue to dream | Contact me
Code for training your own [\[original\]](#) [\[simple\]](#)
[Art](#) • [Cats](#) • [Horses](#) • [Chemicals](#)
[Another](#) ×

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, Timo Aila:
Analyzing and Improving the Image Quality of StyleGAN. CVPR 2020: 8107-8116

GPT-2: A Fake News Generator?

SYSTEM PROMPT (HUMAN-WRITTEN)	<p><i>A train carriage containing controlled nuclear materials was stolen in Cincinnati today. Its whereabouts are unknown.</i></p>
MODEL COMPLETION (MACHINE-WRITTEN, FIRST TRY)	<p>The incident occurred on the downtown train line, which runs from Covington and Ashland stations.</p> <p>In an email to Ohio news outlets, the U.S. Department of Energy said it is working with the Federal Railroad Administration to find the thief.</p> <p>"The theft of this nuclear material will have significant negative consequences on public and environmental health, our workforce and the economy of our nation," said Tom Hicks, the U.S. Energy Secretary, in a statement. "Our top priority is to secure the theft and ensure it doesn't happen again."</p> <p>The stolen material was taken from the University of Cincinnati's Research Triangle Park nuclear research site, according to a news release from Department officials.</p> <p>The Nuclear Regulatory Commission did not immediately release any information.</p> <p>According to the release, the U.S. Department of Energy's Office of Nuclear Material Safety and Security is leading that team's investigation.</p> <p>"The safety of people, the environment and the nation's nuclear stockpile is our highest priority," Hicks said. "We will get to the bottom of this and make no excuses."</p>

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever:
Language Models are Unsupervised Multitask Learners. Technical report, OpenAI, 2019



Global development Football Tech Business Environment Obituaries

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse



▲ The AI wrote a new passage of fiction set in China after being fed the opening line of Nineteen Eighty-Four by George Orwell (pictured). Photograph: Mondadori/Getty Images

The creators of a revolutionary AI system that can write news stories and works of fiction - dubbed "deepfakes for text" - have taken the unusual step of not releasing their research publicly, for fear of potential misuse.

OpenAI, a nonprofit research company backed by Elon Musk, Reid Hoffman, Sam Altman, and others, says its new AI model, called GPT2 is so good and the risk of malicious use so high that it is breaking from its normal practice of

Advertisement



<https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction>

Fake News Generation Can Now Be Automated

GPT-2: <https://taktotransformer.com>

Custom prompt

The coronavirus was created in a secret laboratory in Antarctica.

GENERATE ANOTHER

Completion

The coronavirus was created in a secret laboratory in Antarctica. The virus is similar to the SARS virus but is not the same. The scientists who created the virus thought it would be a "perfect mechanism to explain why SARS happened" according to reports, Dr. Ben Neuman told CNN.

The coroner of the SARS cases at L'Aquila hospital said earlier this week that one of the victims may have been exposed to the virus while on a cruise ship.

Thirty eight people from L'Aquila's Villa Salsina del Tronto died after the island's popular bar the Pre-Aesthetic suffering from respiratory problems.

Grover: <https://grover.allenai.org>

Headline

The coronavirus was created in a secret laboratory in Antarctica.

Give your article a headline to write against

Article

The story is well researched. In a nutshell, it's that scientists decided at the end of 2006 that there were no viruses in Antarctica that could cause fungal diseases or, so far as anyone knows, perhaps anything of great global importance. So they did exactly what government scientists sometimes do when they're worried: They set up a secret laboratory and created one.

What did they come up with? Made a fungal pathogen that resembled what might have caused SARS. The secret is that the virus was not contained in Antarctica: It was moved there in a container labeled "mono- and coca beans." From the Journal:

Take a quiz: <https://quiz.newsyoucantuse.com/>

Grover: Defending Against Neural Fake News

...Using Fake News Generators...

The screenshot shows a news article from The New York Times Science section. The title is "Link Found Between Vaccines and Autism" by Paul Waldman, dated May 29, 2019. The text discusses a study from the University of California San Diego School of Medicine and the Centers for Disease Control and Prevention reporting a link between vaccination and autism. A large watermark reading "Fake News" is overlaid on the image.

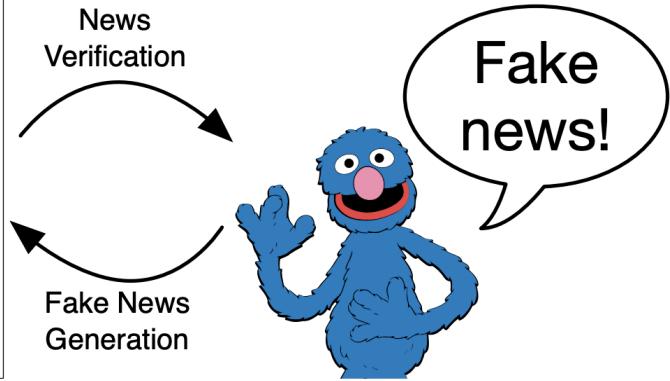


Figure 1: In this paper, we explore GROVER, a model which can detect *and generate* neural fake news. Humans find the articles difficult to distinguish from “real news” without high levels of scrutiny.

Defending Against Neural Fake News: Looking for Multi-Modal Inconsistencies

nytimes.com

What's Next for Britons after Brexit?

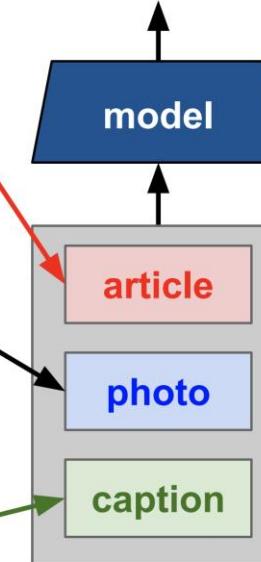
August 28, 2019 - Anne Smith

In September, voters overwhelming rejected a plan from Prime Minister Theresa May's team for the United Kingdom to stay in the European Union. On March 29, Britain will officially exit the union after years of campaigning and serious negotiations. The EU's chief Brexit negotiator, Michel Barnier, has warned that there could be no future trade deals with the United Kingdom if there is a "no deal." The transition period will allow the United Kingdom and the European Union to work out a new plan for their relationship. But we may not know ...



Parliament was scheduled to reconvene on Oct 9, but Mr. Johnson said he planned to extend its break.

Human or
machine-generated?



The COVID-19 Infodemic

'Hundreds dead' because of Covid-19 misinformation

By Alistair Coleman
BBC Monitoring

⌚ 12 August

Reality Check



REUTERS

Facebook, Twitter and chat apps like WhatsApp have proved fertile ground for unfounded rumours about Covid-19

At least 800 people may have died around the world because of coronavirus-related misinformation in the first three months of this year, researchers say.

<https://www.bbc.com/news/world-53755067>

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**

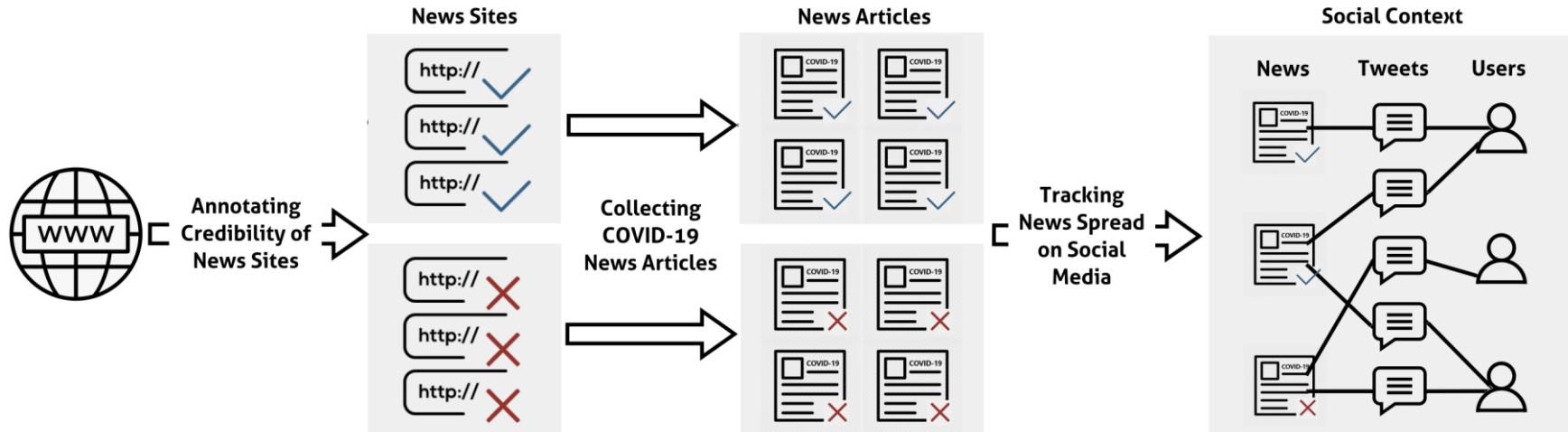
February 12, 2020

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/the-coronavirus-is-the-first-true-social-media-infodemic/>

ReCOVery (COVID-19 Fake News)

distant supervision for annotations



	Reliable	Unreliable	Total
News articles	1,364	665	2,029
w/ images	1,354	663	2,017
w/ social information	1,219	528	1,747
Tweets	114,402	26,418	140,820
Users	78,659	17,323	93,761

Xinyi Zhou, Apurva Mulay, Emilio Ferrara, Reza Zafarani:

ReCOVery: A Multimodal Repository for COVID-19 News Credibility Research. CIKM 2020: 3205-3212

ReCOVery (COVID-19 Fake News)

distant supervision for annotations



(a) Reliable News¹⁵

The screenshot shows a news article from HAF (Healthcare Action Fund) titled "Researchers: 100% Covid-19 Cure Rate Using Intravenous Chlorine Dioxide". The article is dated May 21, 2020. The text discusses preliminary data from a clinical trial involving more than 100 COVID-19 patients in Ecuador, which resulted in a claimed 100% cure rate within four days. It mentions that the tests were carried out by the Asociacion Ecuatoriana de Medicos Expertos en Medicina Integrativa, a group of integrative medicine practitioners. The text also notes that Ecuador has been hit particularly hard by the coronavirus and that the current "standard of care" promoted by Western medicine, which largely relies on ventilators, has been killing the vast majority of critical patients while failing to address the real root of the problem.

(b) Unreliable News¹⁶

ReCOVery (COVID-19 Fake News)

distant supervision for annotations

	Reliable	Unreliable	Total
News articles	1,364	665	2,029
w/ images	1,354	663	2,017
w/ social information	1,219	528	1,747
Tweets	114,402	26,418	140,820
Users	78,659	17,323	93,761

Modeling the Perspective of Journalists, Fact-Checkers, Social Media Platforms, Policy Makers, and the Society



خالد جاسم ■ قاعد في البيت
@khalidjassem74 · Mar 16
غير صحيح ما يتم تداوله حالياً عن قيام مروحيات عسكرية برش مبيدات ضد فيروس كورونا في سماء جميع أنحاء البلاد،
هذا اشاعة ..

Military helicopters spray pesticides against Coronavirus.
This is rumor!

(a) rumor in Arabic and English



Truth Caller
@TCzeroed

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



(b) spreading panic



سarin -
@SARENsoraqi

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



M.Krishnikaa
@krishnikakrishn

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



SwaggyPeterson
@SwaggyPetersen

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



Chris Hayes
@chrishayes

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



lewis bassett
@le_bassett

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



Think free
@thinkfree55

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

Can We Win the War on “Fake News”?

Complex Problem, No Easy Solution

- **Need for cooperation**
 - social media, technology companies
 - governments
 - international organizations
 - civil society: 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/2019/05/europe/finland-fake-news-intl/>

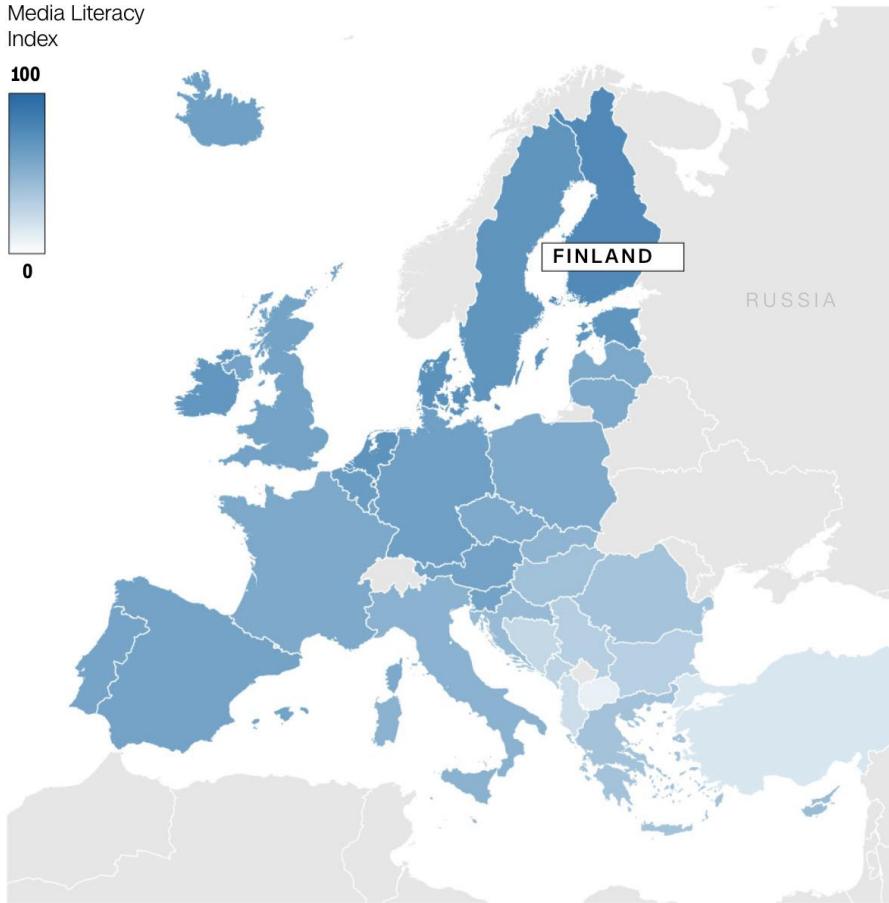
Media literacy across Europe

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

Media Literacy Index

100

0



Finland	76
Denmark	71
Netherlands	70
Sweden	69
Estonia	69
Ireland	68
Belgium	64
Germany	62
Iceland	62
UK	60
Slovenia	60
Austria	60
Spain	60
Luxembourg	59
Portugal	59
France	56
Latvia	56
Polan	55
Czech Rep.	55
Lithuania	55
Italy	50
Slovakia	48
Malta	47
Croatia	44
Cyprus	43
Hungary	40
Greece	39
Romania	38
Serbia	31
Bulgaria	30
Montenegro	28
Bosnia	25
Albania	22
Turkey	16
Macedonia	10

310

Final Thoughts

Final Thoughts

“Fake News” is...

- weaponized
 - extreme personalization
- spam on steroids
 - spreads much widely
 - faster than true news
- politicized
 - better: *disinformation*

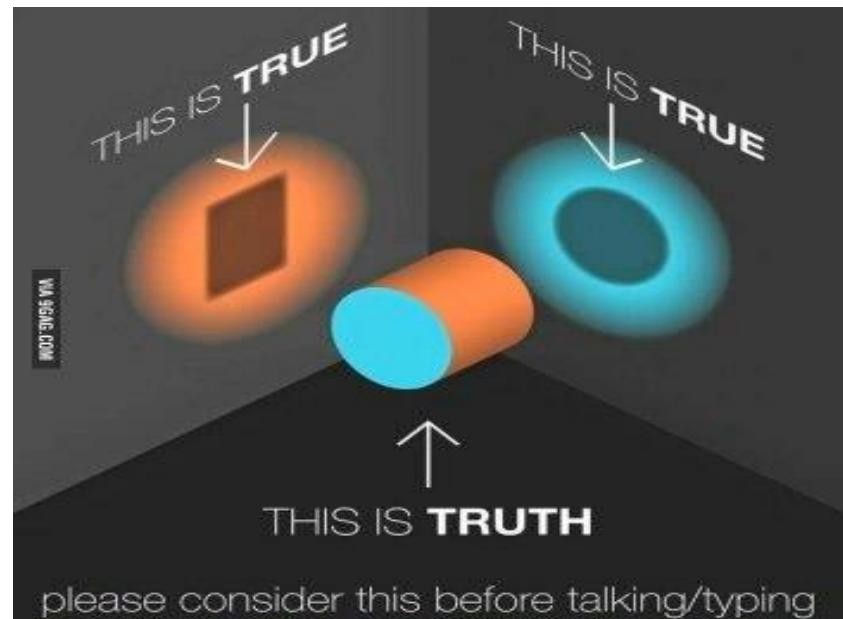
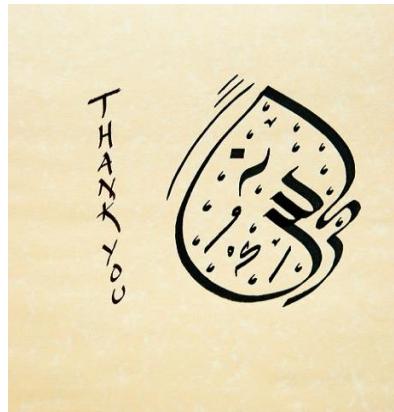
Technology can help

- limit the spread
- raise awareness
- help human fact-checkers

BUT: we should be careful!

- time-proof solutions
- support free speech

Absolute Truth vs. Alternative Facts



References (1)

- Atanas Atanasov, Gianmarco De Francisci Morales and Preslav Nakov: Understanding the Roles of Political Trolls in Social Media. CoNLL 2019
- Pepa Atanasova, Lluís Màrquez, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, Preslav Nakov: Overview of the CLEF-2018 CheckThat! Lab on Automatic Identification and Verification of Political Claims. Task 1: Check-Worthiness. CLEF (Working Notes) 2018
- Pepa Atanasova, Preslav Nakov, Georgi Karadzhov, Mitra Mohtarami, Giovanni Da San Martino: Overview of the CLEF-2019 CheckThat! Lab: Automatic Identification and Verification of Claims. Task 1: Check-Worthiness. CLEF (Working Notes) 2019
- Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James R. Glass, Preslav Nakov: Multi-Task Ordinal Regression for Jointly Predicting the Trustworthiness and the Leading Political Ideology of News Media. NAACL-HLT (1) 2019: 2109-2116
- Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James R. Glass, Preslav Nakov: Predicting Factuality of Reporting and Bias of News Media Sources. EMNLP 2018: 3528-3539
- Ramy Baly, Mitra Mohtarami, James R. Glass, Lluís Màrquez, Alessandro Moschitti, Preslav Nakov: Integrating Stance Detection and Fact Checking in a Unified Corpus. NAACL-HLT (2) 2018: 21-27
- Alberto Barrón-Cedeño, Giovanni Da San Martino, Israa Jaradat, Preslav Nakov: Proppy: A System to Unmask Propaganda in Online News. AAAI 2019: 9847-9848
- Carlos Castillo, Marcelo Mendoza, Barbara Poblete: Information credibility on twitter. WWW 2011: 675-684
- Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, Preslav Nakov: Fine-Grained Analysis of Propaganda in News Articles. EMNLP 2019
- Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, Yelena Mejova: Seminar Users in the Arabic Twitter Sphere. SocInfo (1) 2017: 91-108

References (2)

- Yoan Dinkov, Ahmed Ali, Ivan Koychev, Preslav Nakov: Predicting the Leading Political Ideology of YouTube Channels Using Acoustic, Textual, and Metadata Information. INTERSPEECH 2019
- Yoan Dinkov, Ivan Koychev, Preslav Nakov: Detecting Toxicity in News Articles: Application to Bulgarian. RANLP 2019
- William Ferreira, Andreas Vlachos: Emergent: a novel data-set for stance classification. HLT-NAACL 2016: 1163-1168
- Pepa Gencheva, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Ivan Koychev: A Context-Aware Approach for Detecting Worth-Checking Claims in Political Debates. RANLP 2017: 267-276
- Andreas Hanselowski, Avinesh P. V. S., Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, Iryna Gurevych: A Retrospective Analysis of the Fake News Challenge Stance-Detection Task. COLING 2018: 1859-1874
- Naeemul Hassan, Chengkai Li, Mark Tremayne: Detecting Check-worthy Factual Claims in Presidential Debates. CIKM 2015: 1835-1838
- Benjamin D. Horne, Sara Khedr, Sibel Adali: Sampling the News Producers: A Large News and Feature Data Set for the Study of the Complex Media Landscape. ICWSM 2018: 518-527
- Benjamin D. Horne, Sibel Adali: This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. CoRR abs/1703.09398 (2017)
- Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Màrquez, Preslav Nakov: ClaimRank: Detecting Check-Worthy Claims in Arabic and English. NAACL-HLT (Demonstrations) 2018: 26-30
- Georgi Karadzhov, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Ivan Koychev: Fully Automated Fact Checking Using External Sources. RANLP 2017: 344-353
- Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, Martin Potthast: SemEval-2019 Task 4: Hyperpartisan News Detection. SemEval@NAACL-HLT 2019: 829-839
- Yixin Nie, Haonan Chen, Mohit Bansal: Combining Fact Extraction and Verification with Neural Semantic Matching Networks. AAAI 2019: 6859-6866

References (3)

- Jespe Nørregaard, Benjamin D. Horne, Sibel Adali: NELA-GT-2018: A Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles. ICWSM 2019: 630-638
- Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, Meeyoung Cha: Detecting Rumors from Microblogs with Recurrent Neural Networks. IJCAI 2016: 3818-3824
- Todor Mihaylov, Georgi Georgiev, Preslav Nakov: Finding Opinion Manipulation Trolls in News Community Forums. CoNLL 2015: 310-314
- Todor Mihaylov, Ivan Koychev, Georgi Georgiev, Preslav Nakov: Exposing Paid Opinion Manipulation Trolls. RANLP 2015: 443-450
- Todor Mihaylov, Preslav Nakov: Hunting for Troll Comments in News Community Forums. ACL (2) 2016
- Todor Mihaylov, Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Georgi Georgiev, Ivan Koychev: The dark side of news community forums: opinion manipulation trolls. Internet Research 28(5): 1292-1312 (2018)
- Tsvetomila Mihaylova, Georgi Karadzhov, Pepa Atanasova, Ramy Baly, Mitra Mohtarami, Preslav Nakov: SemEval-2019 Task 8: Fact Checking in Community Question Answering Forums. SemEval@NAACL-HLT 2019: 860-869
- Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Mitra Mohtarami, Georgi Karadzhov, James R. Glass: Fact Checking in Community Forums. AAAI 2018: 5309-5316
- Mitra Mohtarami, Ramy Baly, James R. Glass, Preslav Nakov, Lluís Màrquez, Alessandro Moschitti: Automatic Stance Detection Using End-to-End Memory Networks. NAACL-HLT 2018: 767-776
- Preslav Nakov, Tsvetomila Mihaylova, Lluís Màrquez, Yashkumar Shiroya, Ivan Koychev: Do Not Trust the Trolls: Predicting Credibility in Community Question Answering Forums. RANLP 2017: 551-560
- Dean Pomerleau and Delip Rao: The Fake News Challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news. 2017. <http://www.fakenewschallenge.org/>.

References (4)

- Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Credibility Assessment of Textual Claims on the Web. CIKM 2016: 2173-2178
- Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Where the Truth Lies: Explaining the Credibility of Emerging Claims on the Web and Social Media. WWW (Companion Volume) 2017: 1003-1012
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, Benno Stein: A Stylometric Inquiry into Hyperpartisan and Fake News. ACL (1) 2018: 231-240
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever: Language Models are Unsupervised Multitask Learners. Technical report, OpenAi, 2019
- Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, Yejin Choi: Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. EMNLP 2017: 2931-2937
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, Arpit Mittal: FEVER: a Large-scale Dataset for Fact Extraction and VERification. NAACL-HLT 2018: 809-819
- James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, Arpit Mittal: The Fact Extraction and VERification (FEVER) Shared Task. FEVER@EMNLP workshop 2019: 1-9
- Slavena Vasileva, Pepa Gencheva, Lluís Márquez, Alberto Barrón-Cedeño, Preslav Nakov: It Takes Nine to Smell a Rat: Neural Multi-Task Learning for Check-Worthiness Prediction. RANLP 2019
- William Yang Wang: "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. ACL (2) 2017: 422-426
- Dimitrina Zlatkova, Preslav Nakov, Ivan Koychev: Fact-Checking Meets Fauxtography: Verifying Claims About Images. EMNLP 2019
- Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, Yejin Choi: Defending Against Neural Fake News. CoRR abs/1905.12616 (2019)