

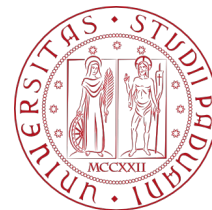
Disinformation, Fake News and Computational Propaganda: Challenges and Opportunities for Machine Learning Research

Giovanni Da San Martino
University of Padova, Italy
giovanni.dasanmartino@unipd.it

Preslav Nakov, Mohamed bin Zayed University of
Artificial Intelligence UAE



ICML
International Conference
On Machine Learning



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**

- Goal: provide a guide to several tasks around the topic of disinformation:
 - Terminology
 - Tasks and Data
 - Approaches

- Topics
 - Definitions and Motivation
 - A Fact-Checking Pipeline
 - Fake News Detection
 - Computational Propaganda
 - Conclusions

Definitions and Motivation



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

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

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



SINCE 1828

JOIN MWU | GAMES | BROWSE THESAURUS | WORD OF THE DAY | WORDS AT PLAY | TIME TRAVELER



DICTIONARY

THESAURUS



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”).

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



SINCE 1828

JOIN MWU | GAMES | BROWSE THESAURUS | WORD OF THE DAY | WORDS AT PLAY | TIME TRAVELER



DICTIONARY

THESAURUS

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

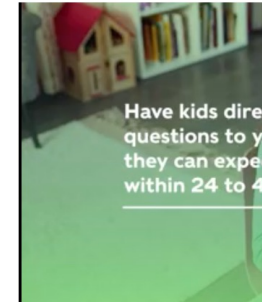
NEWS IN PHOTOS

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 |

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

You may also like



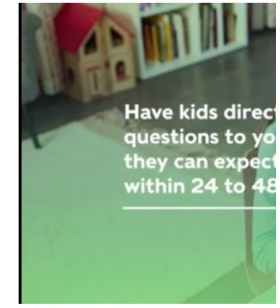
NEWS IN PHOTOS

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 |

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

You may also like





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>

fake news in British



noun

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

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



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.

<https://dictionary.cambridge.org/dictionary/english/fake-news>

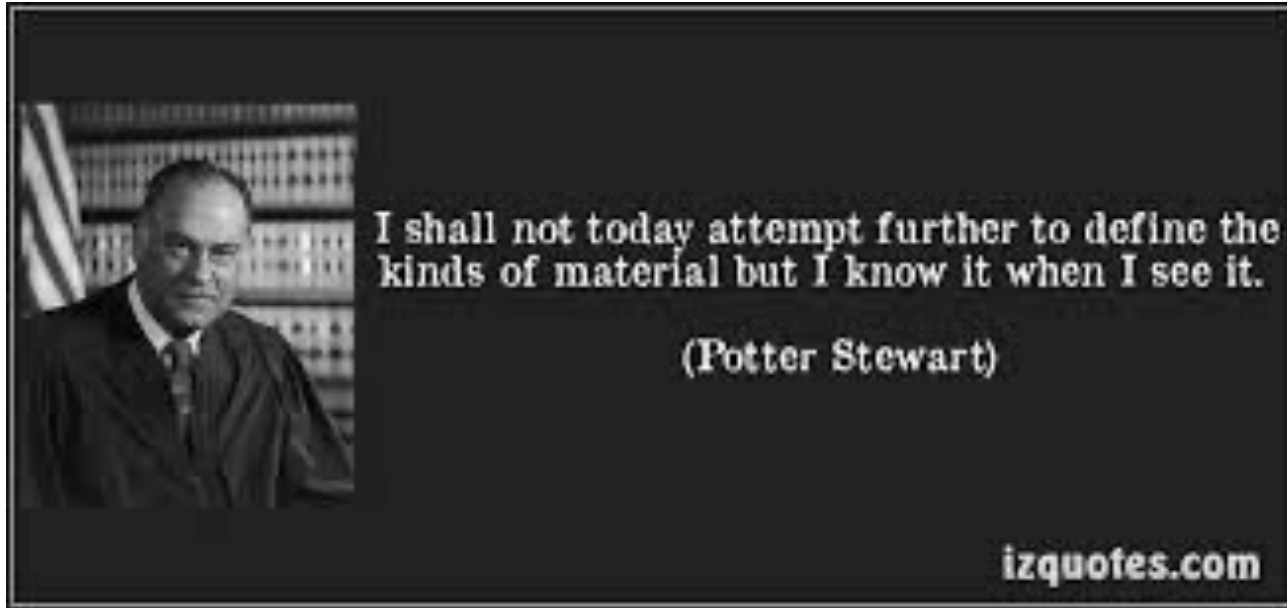
WIKIPEDIA

The Free Encyclopedia

- 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.

https://en.wikipedia.org/wiki/Fake_news

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



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”

БЪЛГАРСКА НАЦИОНАЛНА ТЕЛЕВИЗИЯ **БНТ** BULGARIAN NATIONAL TELEVISION WATCH BNT ONLINE

БНТ 1 БНТ 2 БНТ 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



Варна 6° 8:17

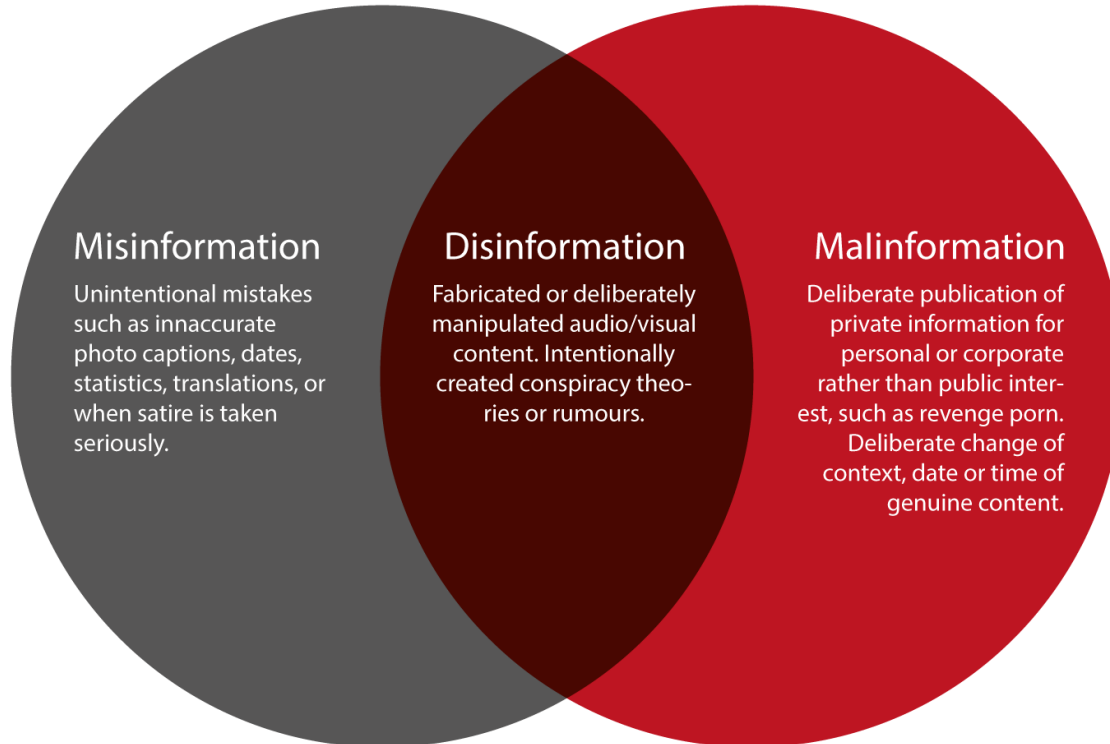
<https://www.bnt.bg/en/a/mariya-gabriel-75-of-young-people-do-not-recognize-fake-news>

Disinformation

TYPES OF INFORMATION DISORDER

FALSENESS

INTENT TO HARM



**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”?**

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

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



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



Getty Images



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>

N 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>



4:13

+ QUEUE

DOWNLOAD

EMBED



WORLD

Viral WhatsApp Messages Are Triggering Mob Killings In India

July 18, 2018 · 9:12 AM ET

LAUREN FRAYER



<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



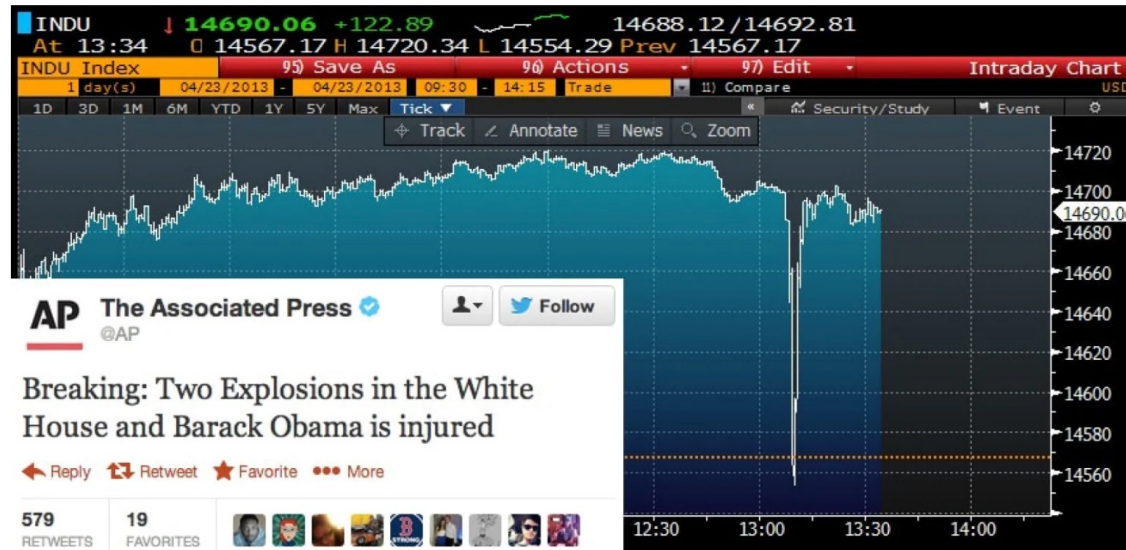
<https://time.com/5197039/un-facebook-myanmar-rohingya-violence/>

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

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

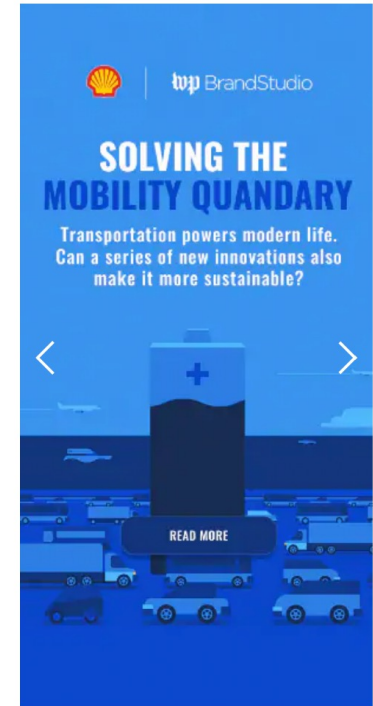
<https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/>

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.



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.”



World Health
Organization



World Health
Organization

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>

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.

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

“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*.

A Fact-Checking Pipeline



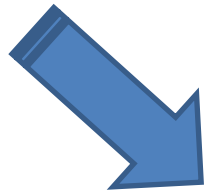
UNIVERSITÀ
DEGLI STUDI
DI PADOVA

TYPES OF INFORMATION DISORDER

FALSENESS

INTENT TO HARM

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



Misinformation

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

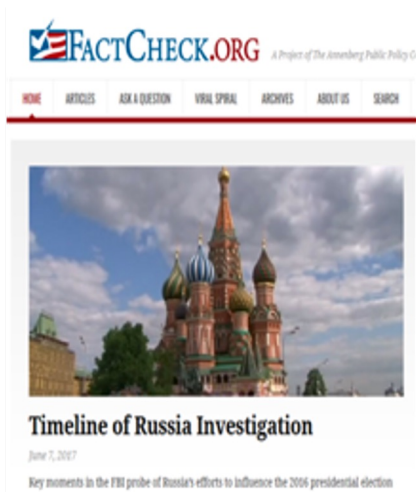
Disinformation

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

Malinformation

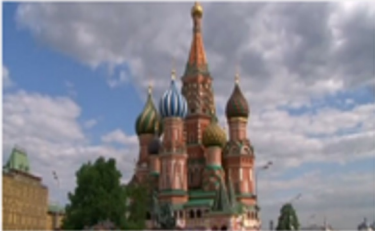
Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn.
Deliberate change of context, date or time of genuine content.

The Fact-Checking Landscape



FACTCHECK.ORG A Project of The Annenberg Public Policy Center

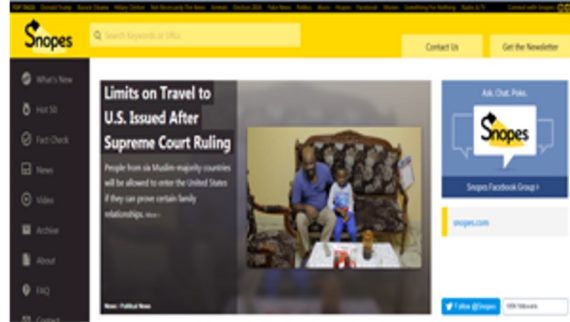
HOME ARTICLES ASK A QUESTION VIRAL SPIN ARCHIVES ABOUT US SEARCH



Timeline of Russia Investigation

June 7, 2017


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



Snopes Search keywords on Snopes Contact Us Get the Newsletter


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 can prove certain family relationships, news...



Snopes Facebook Group

Like Snopes 10K Members




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?



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

Most Recent

- 29TH AUG 2019 Claim that 97% of Labour activists support remain alliance is flawed
- 29TH AUG 2019 Do Netting Hill Carnival and Glastonbury have similar crime rates?
- 28TH AUG 2019 What does proroguing parliament mean?
- 28TH AUG 2019 Can refugees enter the UK legally?
- 27TH AUG 2019 Does a no deal Brexit mean no divorce bill?

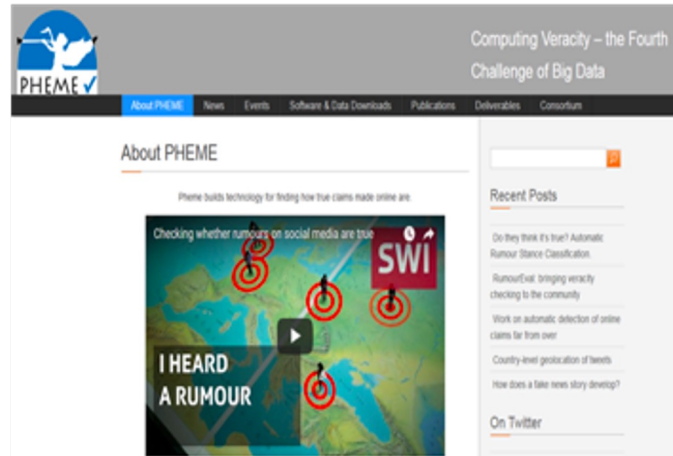


POLITIFACT WINNER OF THE PULITZER PRIZE

My Computer

Search

EDITORS TRUTH ASSESSOR PEOPLE PROMISSES FACTS ON FILE ABOUT US




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 finding how true claims made online are



I HEARD A RUMOUR

Checking whether rumours on social media are true

SWI

Recent Posts

- Do they think it's true? Automatic Rumour Stance Classification
- RumourNet: bringing veracity checking to the community
- Work on automatic detection of online claims for Iran over
- Country-level geolocation of tweets
- How does a fake news story develop?

On Twitter



WeVerify About Us Verification Plugin Results News And Events Blog

Partners/Consortium

HOME / PARTNERS/CONSORTIUM

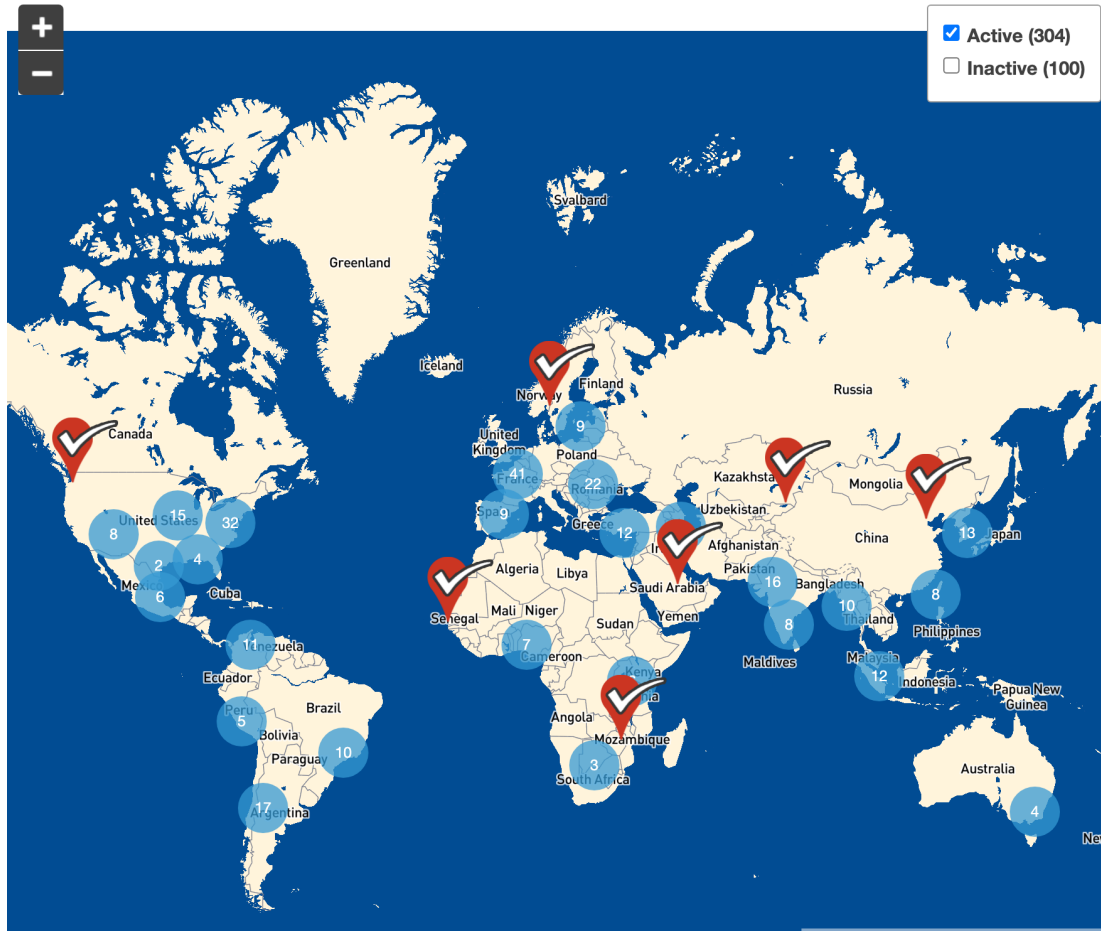
ONTOTEXT will coordinate the overall project as well as lead WP3, where, together with ATC, they will implement the WeVerify open platform for decentralised content verification. Building on its expertise in solution development, platforms targeting media and publishing, and development of open-source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



USFD will lead WP6 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.

The Fact-Checking Landscape

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



[LATEST NEWS](#) [FACT-CHECKING](#) [TECH & CHECK](#) [ABOUT THE LAB](#)

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 (FR)	AFP Factuel
Number of claims	33,261	12,812	16,476	520	1,311	823	125	250	678	266

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

Technology should:

- automate the entire fact-checking process
- make fact-checkers more productive

What do fact checkers want?

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

Check-worthiness

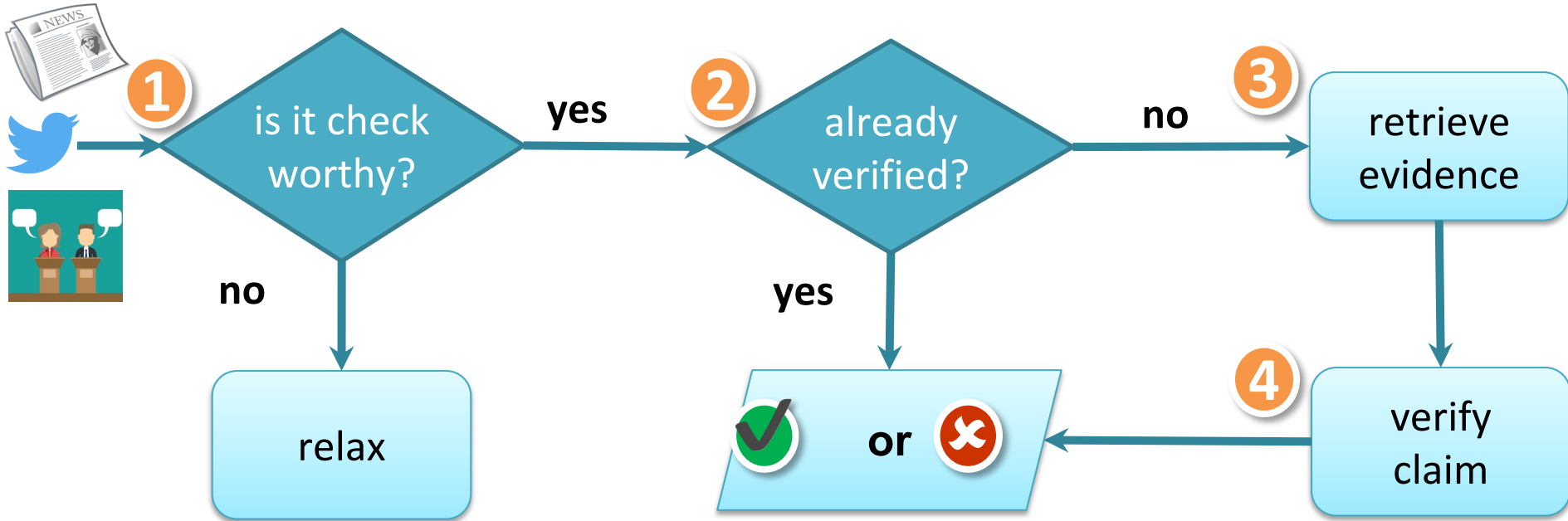
*Verified claim
retrieval*

*Evidence retrieval /
claim verification*

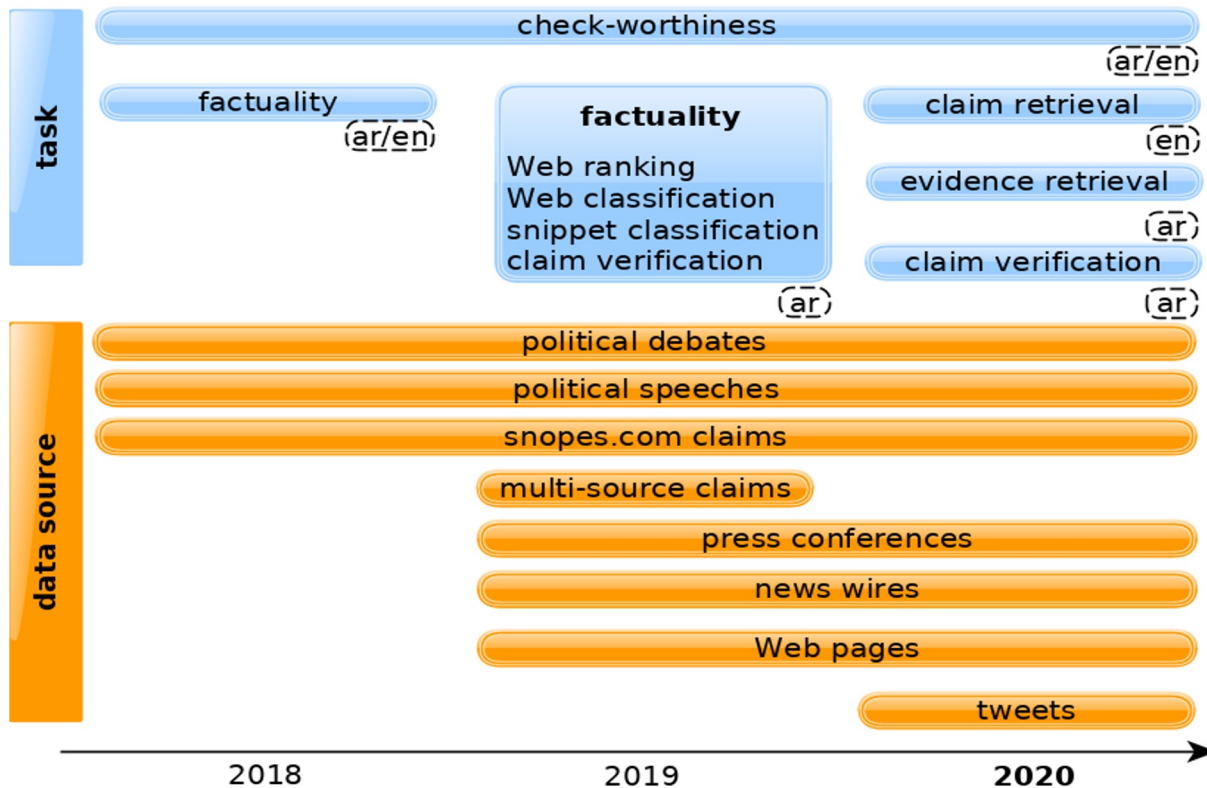
IJCAI-2021: *Automated Fact-Checking for Assisting Human Fact-Checkers.*
Preslav Nakov, David Corney, Maram Hasanain, Firoj Alam, Tamer Elsayed,
Alberto Barrón-Cedeño, Paolo Papotti, Shaden Shaar, Giovanni Da San Martino



The Fact-Checking Flow-Chart

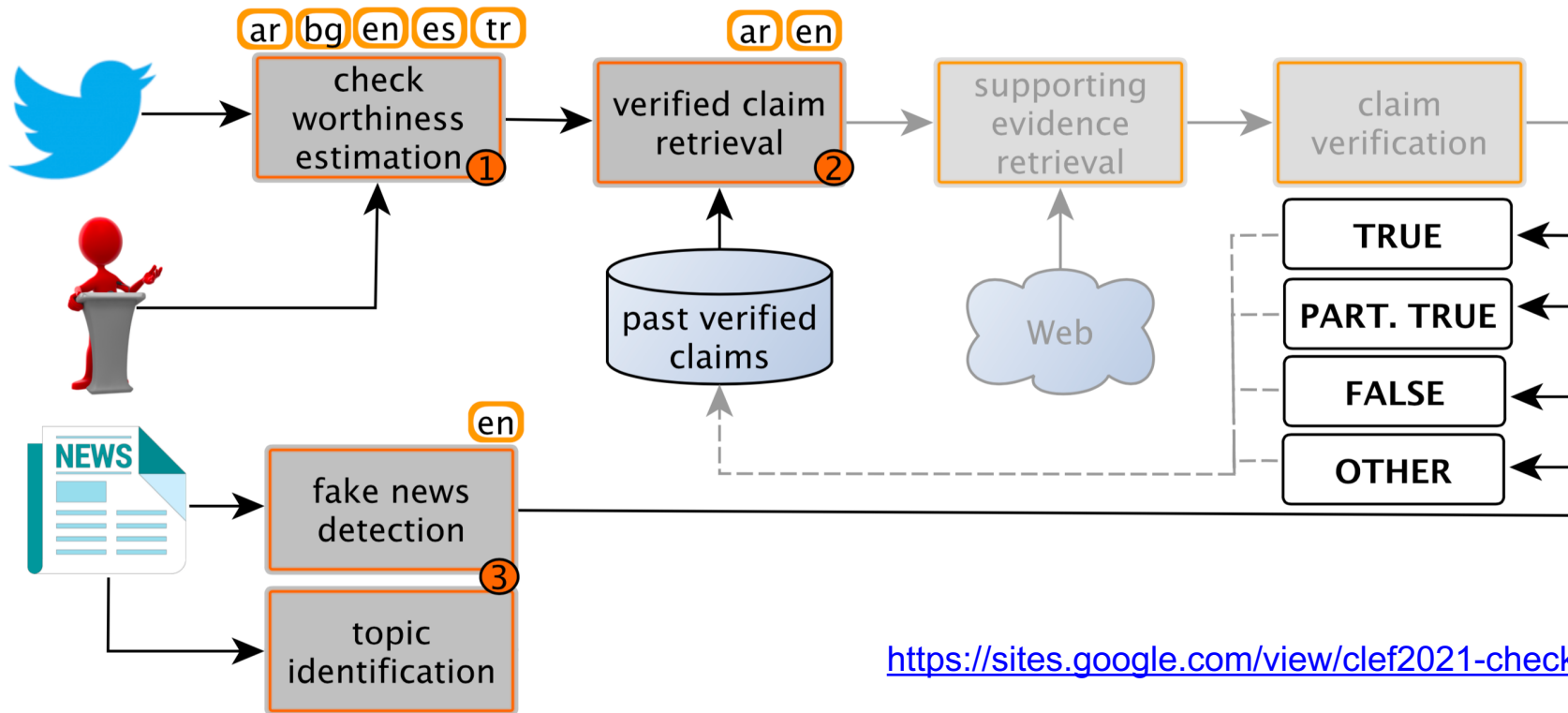


The CheckThat! Lab at CLEF 2018-2020



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

The CheckThat! Lab at CLEF'2021



<https://sites.google.com/view/clef2021-checkthat/>

ECIR-2021: The CLEF-2021 CheckThat! Lab on Detecting Check-Worthy Claims, Previously Fact-Checked Claims, and Fake News
Preslav Nakov, Giovanni Da San Martino, Tamer Elsayed, Alberto Barrón-Cedeño, Rubén Míguez, Shaden Shaar, Firoj Alam, Fatima Haouari, Maram Hasanain, Nikolay Babulkov, Alex Nikolov, Gautam Kishore Shahi, Julia Maria Struß, Thomas Mandl

What is Worth Fact-Checking?

***(In Political Debates and Tweet
Streams)***

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

NATIONAL

DEBATES



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



Donald Trump

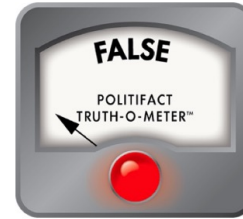
stated on October 22, 2020 in a presidential debate:

"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.

<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 Walker: (07:58)

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

Kristen Walker: (08:27)

The goal 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 I will begin with the fight against the coronavirus. President Trump, the first question is for you. The country is heading into a dangerous new phase. More than 40,000 Americans are in the hospital tonight with COVID. Including record numbers here in Tennessee. And since the two of you last shared a stage, 16,000 Americans have died from COVID. So please be specific. How would you lead the country during this next stage of the coronavirus crisis? Two minutes, uninterrupted.

Kristen Walker: (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 and 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 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 give us Operation Warp Speed, which is the miracle 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 the therapeutic, I guess they would call it. Some people would call cure, but I was in for a short period of time. And I got better without beeing hospitalized. And now they say I'm immune. Whether it's four months or a lifetime, nobody's been able to say that, but I'm immune. More and more people are getting better. We have a problem that's a worldwide problem. This is a worldwide problem, but I've been congratulated by the heads of many countries on what we've been able to do. If you take a look at what we've done in terms of goggles and masks and gloves and everything else, and in particular ventilators we're now making ventilators all over the world, thousands and thousands a million 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 Walker: (11:06)

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: (11:54)

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

Joe Biden: (12:19)

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

Joe Biden: (12:41)

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

Kristen Walker: (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 Walker: (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 were working on very closely with other countries, in particular Europe.

Kristen Walker: (13:35)

Let me follow up with you, and because this is new information, you have

Donald Trump: (18:24)

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, for 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 model 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 Walker: (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, "It's being transparent. The president of China is being transparent. We owe him a debt of gratitude. We have to thank him."

Joe Biden: (16:36)

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 that. He had nothing, he did virtually nothing. And then he gets out of the hospital and he talks about, "Oh, don't worry. It's all going to be over soon." Come on. There's not another serious scientist in the world who thinks it's going to be over soon.

Kristen Walker: (17:03)

President Trump, your reaction?

Donald Trump: (18:26)

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

Kristen Walker: (19:16)

Vice President Biden.

Joe Biden: (19:17)

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

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

Joe Biden: (20:13)

I didn't say either of those things.

Donald Trump: (20:14)

You certainly did. You certainly did.

Kristen Walker: (22:04)

Let's talk about schools. President Trump.

Donald Trump: (22:04)

I think we have to respond, if I might.

Kristen Walker: (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 Democratic Governors, Cuomo in New York, you look at what's going on in California, you look at Pennsylvania, North Carolina, Democrats, Democrats all, they're shut down so tight and they're dying. They're dying. And he supports all these people. All he talks about is shut downs. Now, 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 Walker: (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: (23:05)

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

abuse. We have to open our country, I've said it often, the cure cannot be

The debate has 1000+ sentences!

What is worth fact-checking?

... learned a lot, learned a lot, great doctors, great hospitals, and now recovered. 99.9 of young people recover. 99% of people recover. We have to recover.

Donald Trump: (18:03)
99% of people recover. We have to recover. We can't close up our nation. We have to open our school and we can't close up our nation, or you're not going to have a nation.

Kristen Walker: (18:11)
And, of course, the CDC has said young people can get sick with COVID-19 and can pass it. Vice President Biden, I want to talk broadly about strategy though.

Joe Biden: (18:20)
Can I respond to that?

Kristen Walker: (18:21)
30 seconds, please. Then I have a question.

Joe Biden: (18:23)
Number one, that says that we're learning to live with it. People are learning to live with it. You folks now will have an empty chair at the kitchen table this morning. That man or wife going to bed tonight and reaching over to try to touch, there is out of habit, where their wife or husband was, is gone. Learning to live with it. Come on. We're dying with it, because he's never said. See, you said, "It's dangerous." When's the last time? It's really dangerous still? Are we dangerous. You tell the people it's dangerous now. What should they do about the danger? And you say, "I take no responsibility."

Kristen Walker: (18:56)
Let me talk about your Tuesday-

... able to track. You need to be able to provide all the resources that are needed to do this. And that is not inconsistent with saying 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 Walker: (24:36)
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 unbelievably expensive, 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 Walker: (25:16)
Vice President Biden.

Joe Biden: (28:15)
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. 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

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

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 (i)



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

Donald Trump: (07:37)
How are you doing? How are you?
Kristen Welker: (07:58)
And I do want to say a very good evening to both of you. This debate will cover six major topics. At the beginning of each section, each candidate will have two minutes, uninterrupted, to answer my first question. The Debate Commission will then turn on their microphone only when it is their turn to answer. And the Commission will turn it off exactly when the two minutes have expired. After that, both microphones will remain on. But on behalf of the voters, I'm going to ask you to please speak one at a time.

Kristen Welker: (08:27)
The goal 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 I will begin with the fight against the coronavirus.
President Trump, the first question is for you. The country is heading into a dangerous new phase. More than 40,000 Americans are in the hospital tonight with COVID. Including record numbers here in Tennessee. And since the two of you last shared a stage, 16,000 Americans have died from COVID. So please be specific. How would you lead the country during this next stage of the coronavirus crisis? Two minutes, uninterrupted.

Kristen Welker: (09:03)
... during this next stage of the coronavirus crisis. Two minutes uninterrupted.

Donald Trump: (09:04)
So as you know, 2.2 million people modeled out, were expected to die. We closed up the greatest economy in the world in order to fight this horrible disease that came from China. It's a worldwide pandemic. It's the world. You see the spikes in Europe and many other places. If you notice, the mortality rate is down 85%. The excess mortality way down and much lower than almost any other country. And you're fighting it and we're fighting it hard. There is a spike. There was Florida and it's now gone.

Donald Trump: (09:41)
There was a very big spike in Texas. It's now gone. There was 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 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 got therapeutic. I guess they would call it. Some people could say cure, but I was in for a short period of time. And I got better. I wouldn't be here tonight. And now they say I'm immune. When 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 gloves and everything else, and in particular ventilators we're now making ventilators all over the world, thousands and thousands a million 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: (11:06)
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 per. 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: (11:54)
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've 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 severest, most serious journals in the whole world said for the first time ever that the way this president has responded to this crisis has been absolutely 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 we 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, "It'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 that. He had nothing, he did virtually nothing. And that he gets out of the hospital and he talks about

Donald Trump: (15:28)
Let's talk about responsibility. It's not my fault that it came here. It's China's fault. And you know what? It's not Joe's fault that it came here either. It's China's fault. They kept it from going into the rest of China for the most part, but they didn't keep it from coming out to the world, including Europe and ourselves.

Kristen Welker: (19:16)
Vice President Biden.

Joe Biden: (19:17)
The fact is that when we knew it was coming, when it hit, what happened? What did the President say? He said, "Don't worry, it's going to go away. It'll be gone by Easter. Don't worry. Warm weather. Don't worry. Maybe inject bleach." He said he was kidding when he said that, but a lot of people thought it was serious. A whole range of things the President has said, even today, he thinks we are in control. We're about to lose 200,000 more people.

Kristen Welker: (19:43)
President Trump.

Donald Trump: (19:44)
Look, perhaps just to finish this, I was kidding on that, but just to finish this when I closed he said I shouldn't have closed. And that went on for months. Nancy Pelosi said the same thing. She was dancing on the streets in Chinatown, in San Francisco. But when I closed, he said, "This is a terrible thing, you xenophobic." I think he called me racist even, because I was closing it to China. Now he says I should have closed it earlier. Joe, it doesn't.

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

in a vaccine if it were approved?

Joe Biden: (14:26)
Make sure it's totally transparent. Have the scientists of the world see it. Know it, look at it, go through all the processes. And by the way, this is the same fellow who told you, "This is going to end by Easter" last time. This is the same fellow who told you that, "Don't worry, we're going to end this by the summer." We're about to go into a dark winter, a dark winter and he has no clear plan. And there's no prospect that there's going to be a vaccine available for the majority of the American people before the middle of next year.

Kristen Welker: (14:54)
President Trump, your reaction. He says, you have no plan.

Donald Trump: (14:56)
I don't think we're going to have a dark winter at all. We're opening up our country. We've learned and studied and understand the disease, which we didn't at the beginning. When I closed and banned China from coming in heavily infected and then ultimately Europe, but China was in January, months later, he was saying I was xenophobic. I did it too soon. Now he's saying, "Oh, I should have moved quicker," but he did 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 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.

to recover.
Donald Trump: (16:03)
99% of people recover. We have to recover. We can't close up our nation. We have to open our school and we can't close up our nation, or you're not going to have a nation.

Kristen Welker: (16:11)
And, of course, the CDC has said young people can get sick with COVID-19 and can pass it. Vice President Biden, I want to talk broadly about strategy though.

Joe Biden: (18:20)
Can I respond to that?

Kristen Welker: (18:21)
30 seconds, please. Then I have a question.

Joe Biden: (18:23)
Number one, he says that we're learning to live with it. People are learning to live with it. You folks none will have an empty chair at the kitchen table this morning. That man or wife going to bed tonight and reaching over to try to touch, there out of habit, where their wife or husband was, is gone. Learning to live with it. Come on. We're dying of it, because he's never said. See, you said, "It's dangerous." When's the last time? It is really dangerous still? Are we dangerous. You tell the people it's dangerous now. What should they do about the danger? And you say, "I take no responsibility."

Kristen Welker: (18:56)
Let me talk about your Tuesday-

Vice President you suggested you would support new shutdowns if scientists recommended it. What do you say to Americans who are fearful that the cost of shutdowns, the impact on the economy, the higher rates of hunger, depression, domestic, and substance abuse outweighs the risk of exposure to the virus?

Joe Biden: (20:54)
What I would say is, I'm going to shut down the virus, not the country. It's his ineptitude that caused the country to have to shut down in large part, why businesses have gone under, why schools are closed, why so many people have lost their living, and why they're concerned. Those are other concerns are real. That's why he should have been, instead of in a sand trap at his golf course, he should have been negotiating with Nancy Pelosi and the rest of the Democrats and Republicans about what to do about the acts they were passing for billions of dollars to make sure people had the capacity.

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

Joe Biden: (21:25)
Oh, well, no. I'm not shutting down today, but there are ... Look, you need standards. The standard is, if you have a reproduction rate in a community that's above a certain level, everybody says, "Slow up. More social distancing. Do not open bars and do not open gymsnasiums. Do not open until you get this under control, under more control." But when you do open, give the people the capacity to be able to open, and they have the capacity to do it safely. For example, schools, schools, they need a lot of money to be open. They need to deal with ventilation systems. They need to deal with smaller classes, more teachers, more pods, and he's refused

Donald Trump: (22:04)
Let's talk about schools. President Trump, Excuse me. I take full responsibility. It's not my fault that it came here. It's China's fault. And you know what? It's not Joe's fault that it came here either. It's China's fault. They kept it from going into the rest of China for the most part, but they didn't keep it from coming out to the world, including Europe and ourselves.
Kristen Welker: (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 Democratic Governors, Cuomo in New York, you look at what's going on in California, you look at Pennsylvania, North Carolina, Democrats, Democrats all, they're shut down so tight and they're dying. They're dying. And he supports all these people. All he talks about is shut downs. No, we're not going to shut down. And we have to open our schools. And it's like, as an example, I have a young son, he also tested positive. By the time I spoke to the doctor the second time, he was fine. It just went away. Young people, I guess, it's their immune system.

Kristen Welker: (22:47)
Let me follow up with you, President Trump. You've demanded schools open in person and insist they can do it safely. But just 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: (23:05)
I want to open the schools. The transmittal rate to the teachers is very small. But I want to open the schools. We have to open our country. We

to do this. We can't keep this country on a massive economy. People are suicide. There's depression, alcohol, even before. There's abuse, tremendous I've said it often, the cure cannot be at's what's happening. And he wants you if one person in our massive down.

to walk and chew gum at the same open, but they need resources to open? if you're going to open a business. Inness. You need to have, if you have a sales 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 saying that we're going

to make sure that we're going to open safely. And by the way, all your 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 unbelievably expensive, 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: (25:15)
Vice President Biden.

Joe Biden: (25:15)
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

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

TRUMP: That makes me smart.

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



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/>

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
<hr/> Total number of sentences: 5,415 <hr/>		

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 |
| | Metadata: reaction of the public |
| 3. Mixed | 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
- 3. Mixed** Discourse
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**."

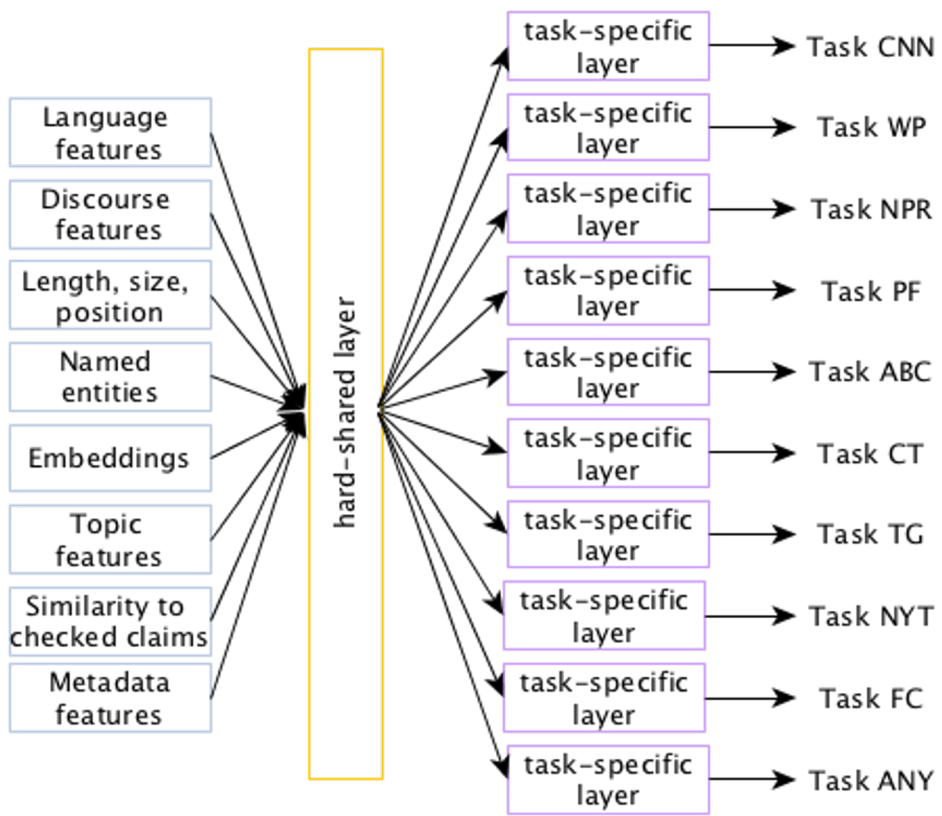
New sentence:

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

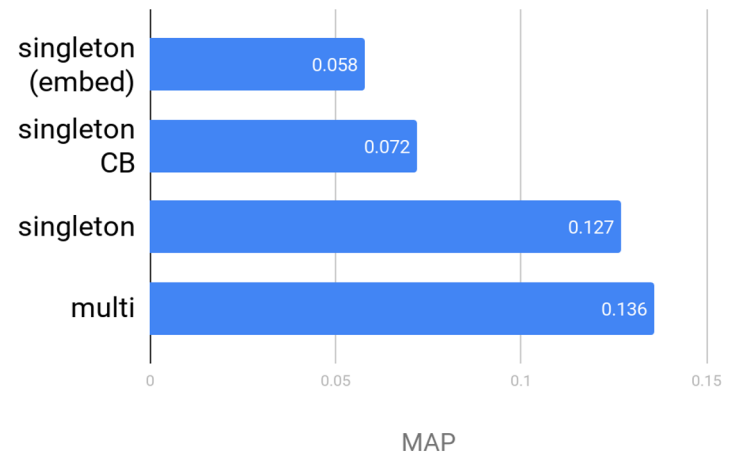
The Role of Context

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



Slavena Vasileva, Pepa Gencheva, Lluís Màrquez, Alberto Barrón-Cedeño, Preslav Nakov:
It Takes Nine to Smell a Rat: Neural Multi-Task Learning for Check-Worthiness Prediction. RANLP 2019

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.

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

DATE: 2017-11-23 NAME: Question Time SENTENCE: What it is is this, it takes people in Germany four days to produce what people in the UK take five days to produce. SENTENCE_ID: b09g1rz4_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.
Digital Threats: Research and Practice 2(2):1-16.

← **Eric Trump** ✓
6,630 Tweets




Eric Trump ✓
@EricTrump

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

⋮ Follow

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

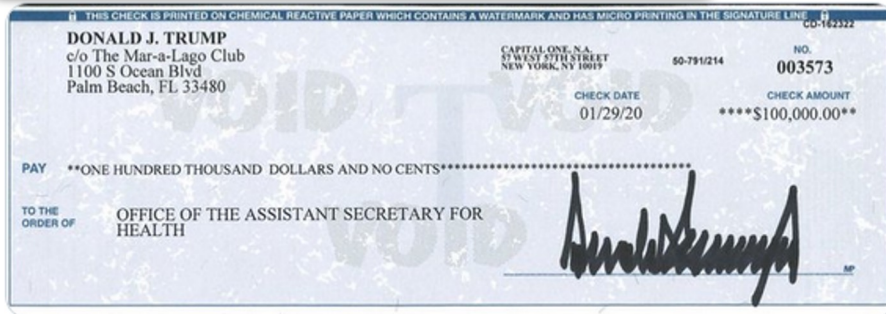


527 1.9K 8.6K



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?

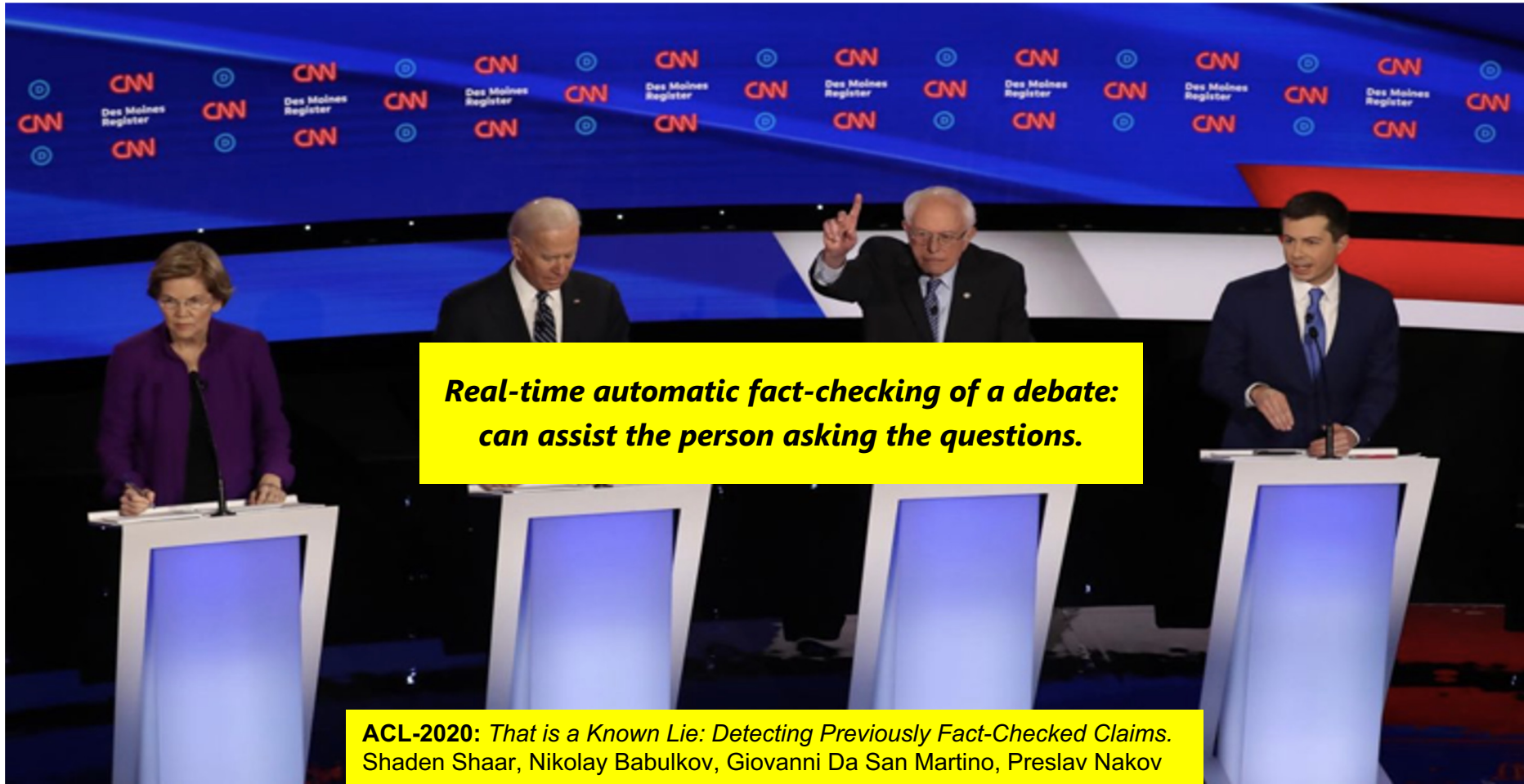
- Q0: viral?
- Q1. contains verifiable factual claim?
- Q2. likely to be false?
- Q3. of interest to the general public?
- Q4: harmful to society, person, company, or product(s)?
- Q5. worth fact-checking?

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

Detecting Previously Fact-Checked Claims

Detecting Previously Fact-Checked Claims



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

ACL-2020: That is a Known Lie: Detecting Previously Fact-Checked Claims.
Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, Preslav Nakov

Detecting Previously Fact-Checked Claims



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

ACL-2020: *That is a Known Lie: Detecting Previously Fact-Checked Claims.*
Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, Preslav Nakov

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

'Strong and stable' - Why politicians keep repeating themselves

By Tom Moseley
Political reporter

27 April 2017

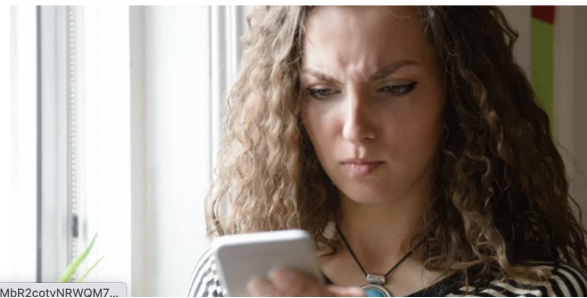


The prime minister sticks relentlessly to her "strong and stable" slogan

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



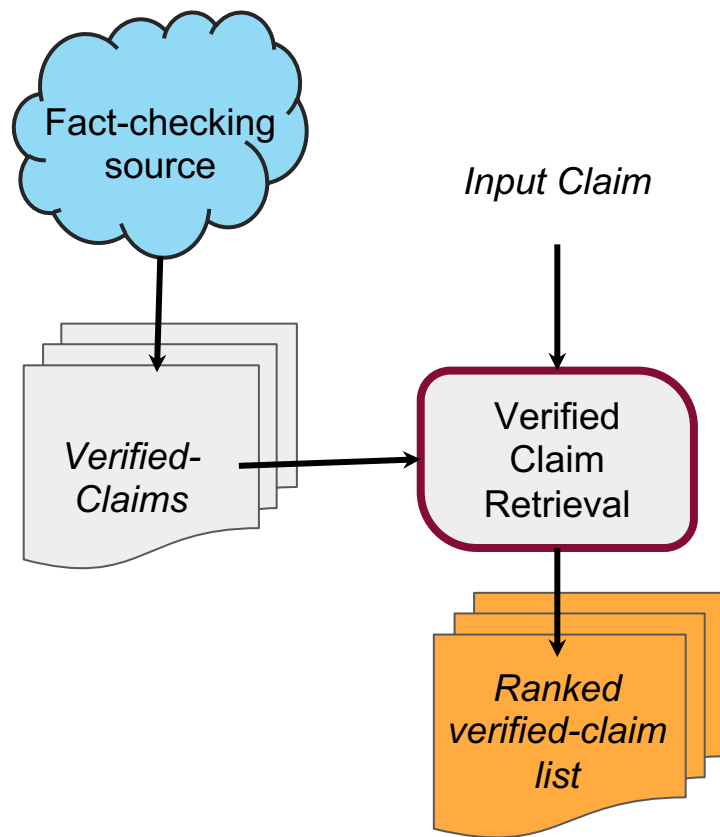
READ THIS NEXT

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 <i>Politifact</i>
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.

Datasets: Statistics

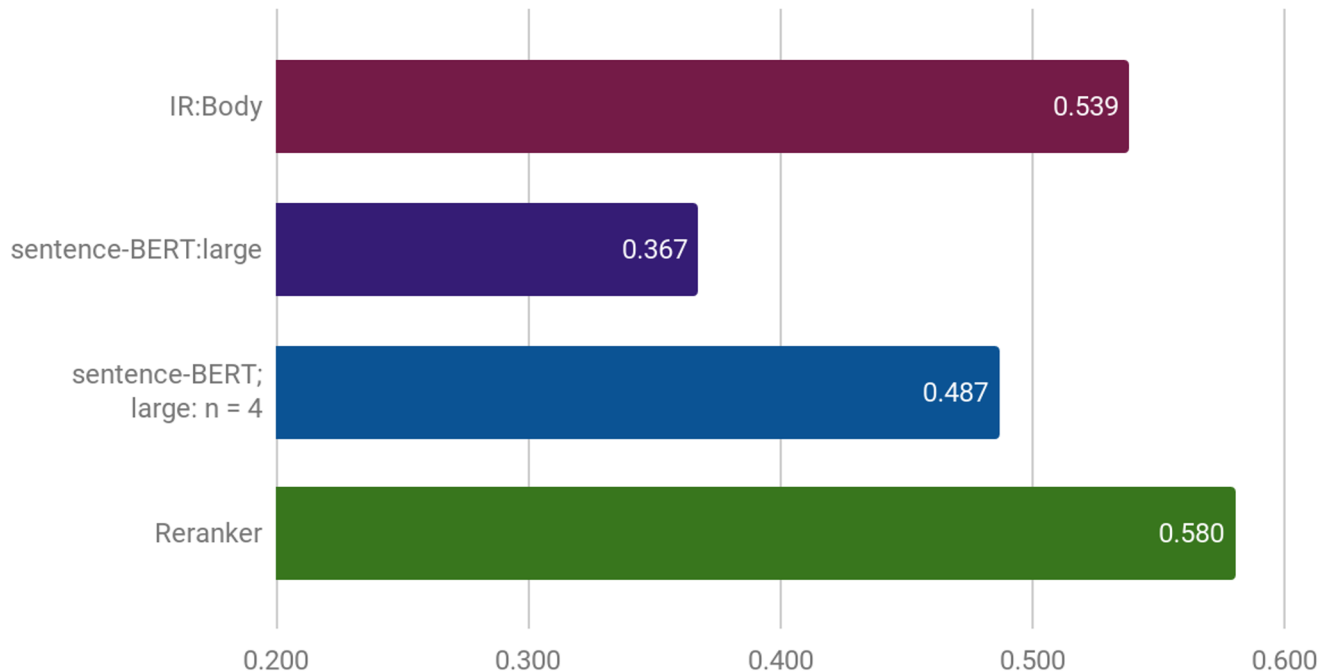
	Politifact	Snopes
<i>Input–VerClaim</i> pairs	768	1,000
– training	614	800
– testing	154	200
Total # of verified claims	16,636	10,396

Experiments:

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

Experiments (Overall): Politifact

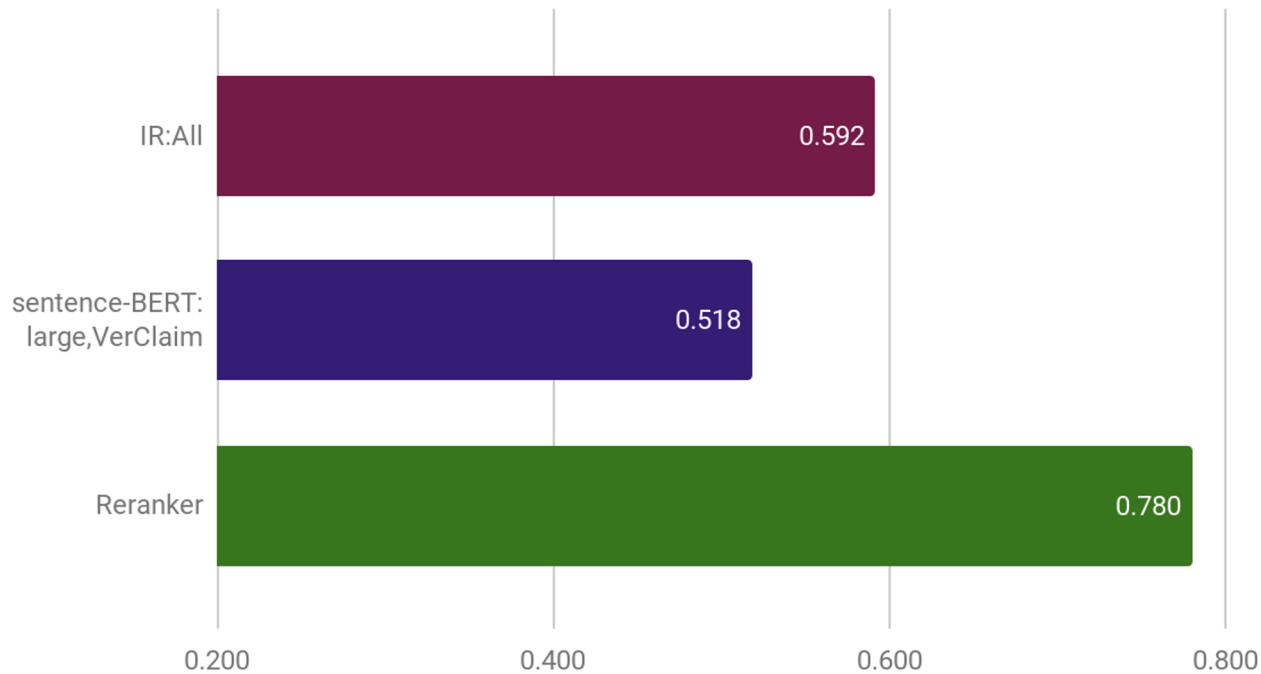
Overall Experiments (Politifact) - MAP@3



ACL-2020: *That is a Known Lie: Detecting Previously Fact-Checked Claims.*
Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, Preslav Nakov

Experiments (Overall): Snopes

Overall Experiments (Snopes) - MAP@3



ACL-2020: *That is a Known Lie: Detecting Previously Fact-Checked Claims.*
Shaden Shaar, Nikolay Babulov, Giovanni Da San Martino, Preslav Nakov

Extension: Using Visual Information

- Important information might not be in the text
- OCR is used to extract text from the images (76%-86% cases)
- The image itself is used to match claims and articles

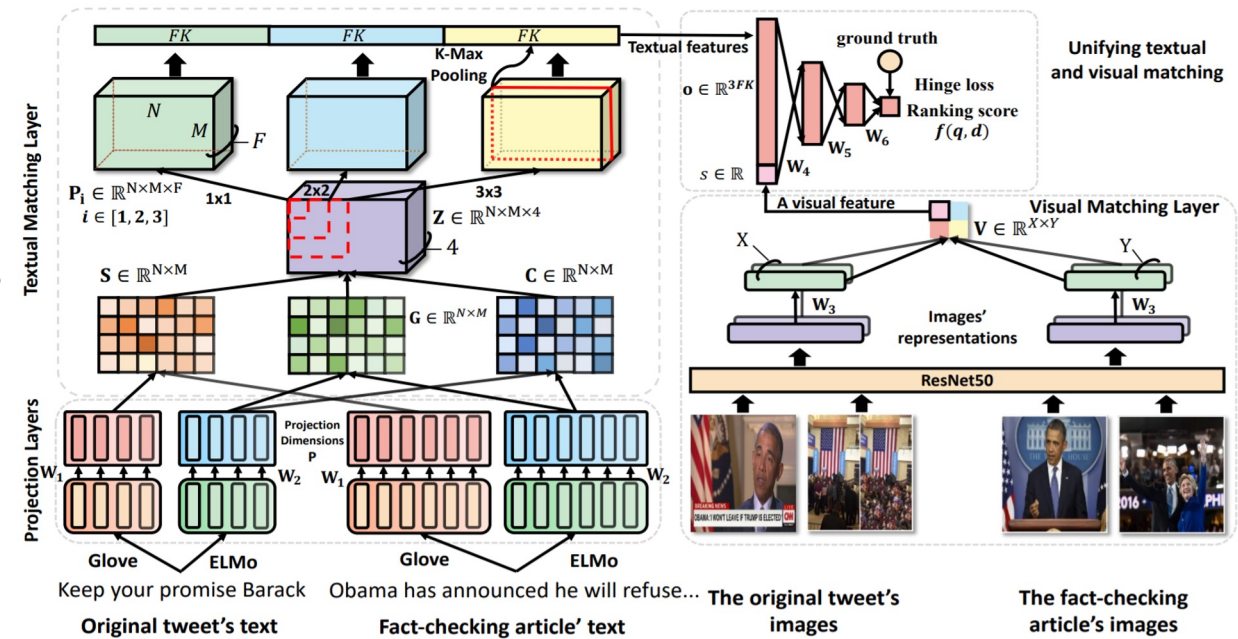


Extension: Using Visual Information

Text - S: word overlap (Glove); G: resolve polysemy; C matches typos (Elmo)

Video - max between
all pairwise image
similarities

Performance of text+images
increases of 4.8%
(normalized discounted
cumulative gain) over
baselines



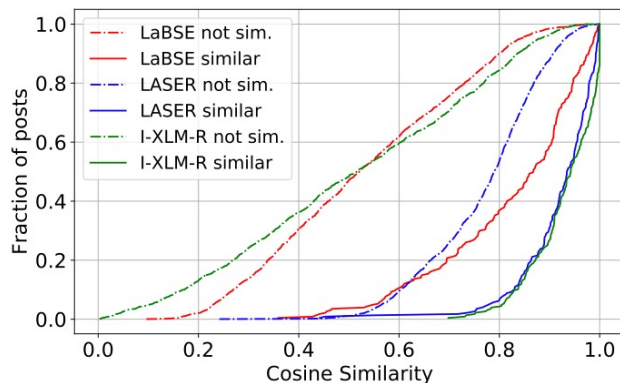
Claim Matching Beyond English to Scale Global Fact-Checking

Item #1	Item #2	Label
पाकिस्तान में गनपॉइंट पर हुई एक डकैती को बताया जा रहा है मुंबई की घटना	कराची पाकिस्तान में घटित लूट को मुंबई का बताया जा रहा है।	Very Similar
பாகிஸ்தானில் உள்ள இந்திய தூதர் உடனடியாக டெல்லி திரும்ப மத்திய அரசு உத்தரவு	*செய்திகள்24/7* *FLASH* *பாகிஸ்தானில் உள்ள இந்திய தூதர் டெல்லி திரும்ப மத்திய அரசு உத்தரவு என தகவல்..*	Very Similar
Barber's salon poses the biggest risk factor for Corona! This threat is going to remain for a long duration. *At an average a barber's napkin touches 5 noses minimum* The US health dept chief J Anthony said that salons have been responsible for almost 50% deaths.	*The biggest danger is from the barbershop itself*. This danger will remain for a long time. *Barber rubs the nose of at least 4 to 5 people with a towel,* The head of the US Department of Health J. Anthony has said that 50 percent of the deaths in the US have happened in the same way that came in saloons.	Very Similar
ഇവിടുത്തെ മാമ മാധ്യമങ്ങൾ Live കാണിച്ചില്ലേലും ദേശീയ മാധ്യമങ്ങൾ ചെയ്തു കേട്ടു	ഇവിടുത്തെ മാധ്യമങ്ങൾ കാണിച്ചില്ലേലും ദേശീയ മാധ്യമങ്ങൾ ചെയ്തു കേട്ടു	Very Similar
Guys important msg:- There is the news of military bsf & cisf coming to Mumbai and having a seven days to 2 weeks curfew...	*Just received information* Entire Mumbai and pune will be under Military lockdown for 10 days starts from Saturday...	Somewhat Similar
Don't believe this FAKE picture of PM Modi; here's the truth	Don't believe this FAKE picture of Virat Kohli; here's the fact check	Very Dissimilar

Claim Matching Beyond English to Scale Global Fact-Checking

Language	Items	κ	Majority Yes
Bengali (bn)	1093	0.30	30%
English (en)	1000	0.60	54%
Hindi (hi)	1000	0.59	41%
Malayalam (ml)	1025	0.63	69%
Tamil (ta)	948	0.63	21%

Lang.	Pairs	κ	Annotators	V. Sim.
bn	644	0.64–0.68	2–3	6%
en	398	0.69	4	15%
hi	399	0.90	3	21%
ml	604	0.91	3	7%
ta	298	0.85	2	11%

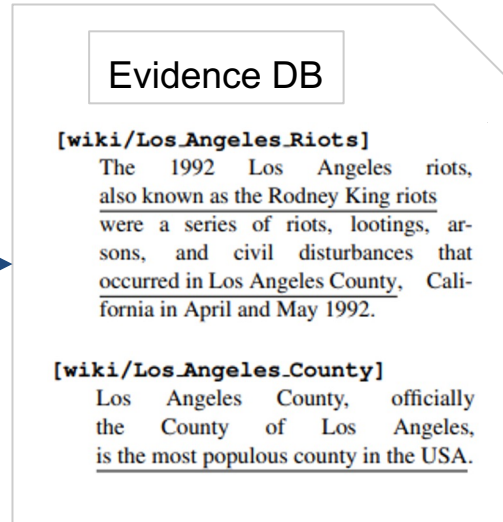


Fact-Checking

Evidence-Based Fact-Checking

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

look up

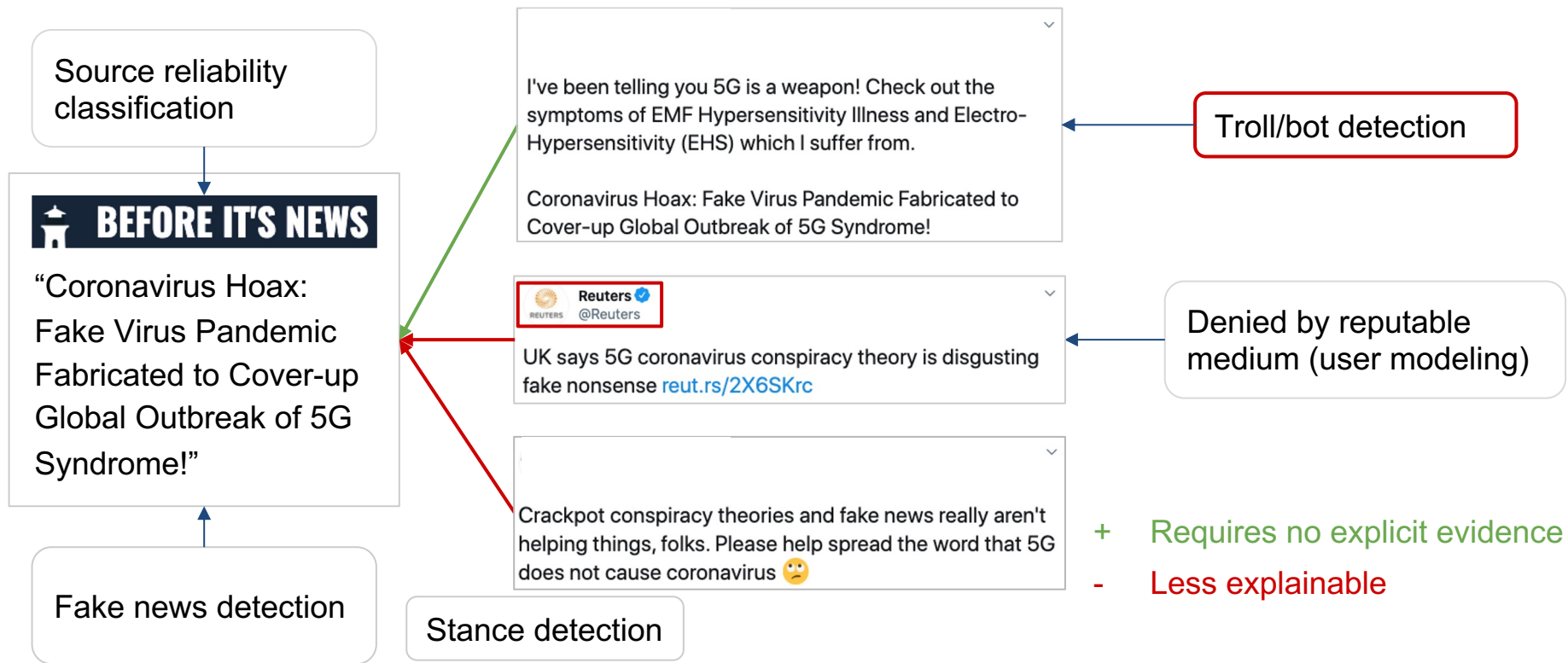


analyse

Verdict: Supported

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

Contextual Fake News Detection



Topic-Aware Evidence Reasoning and Stance-Aware Aggregation

Verdict: *SUPPORTS*

Claim

A high *school student* named Cole Withrow was *charged* for *leaving* an unloaded *shotgun* in his *vehicle* while parking at *school*.

-support-

-support-

-unrelated-

-unrelated-

E1 (gold)

Family *friend* Kim Boykin said Withrow, an Eagle Scout and honors *student*, accidentally left his *gun* in the *car* aftersketeer *shooting* over the *weekend*.

E2 (gold)

Others in the Princeton High *community* agree that Withrow's *punishment* is too harsh, especially after *charges* weren't filed when a loaded *gun* was found in an assistant principal's *car* two *years* ago.

E3 (non-gold)

"Please know that with *student* and personnel *issues*, We carefully *balance* all *factors* to arrive at a *fair* and just *Outcome*." she said in a *statement*.

E4 (non-gold)

He locks his *vehicle*, goes inside and tries to do the right thing.

Topic-Aware Evidence Reasoning and Stance-Aware Aggregation

Verdict: *SUPPORTS*

Claim

A high *school student* named Cole Withrow was *charged* for *leaving* an unloaded *shotgun* in his *vehicle* while parking at *school*.

-support-

-support-

-unrelated-

-unrelated-

E1 (gold)

Family *friend* Kim Boykin said Withrow, an Eagle Scout and honors *student*, accidentally left his *gun* in the *car* aftersketeer *shooting* over the *weekend*.

E2 (gold)

Others in the Princeton High *community* agree that Withrow's *punishment* is too harsh, especially after *charges* weren't filed when a loaded *gun* was found in an assistant principal's *car* two *years* ago.

E3 (non-gold)

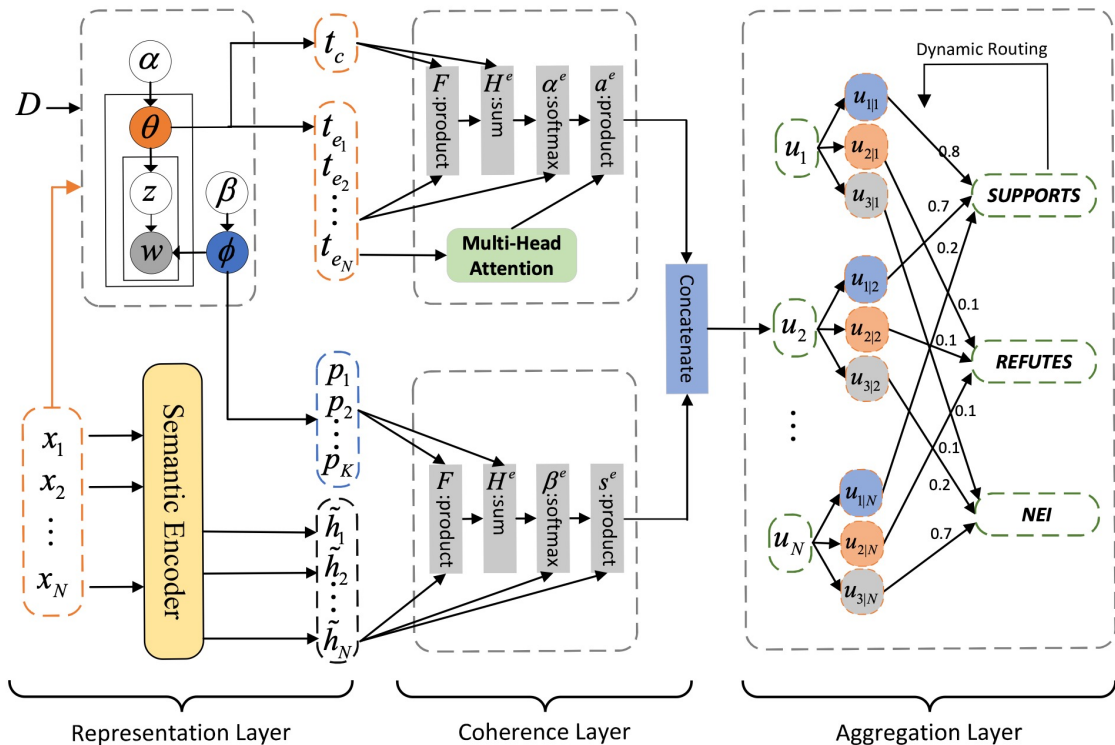
"Please know that with *student* and personnel *issues*, We carefully *balance* all *factors* to arrive at a *fair* and just *Outcome*." she said in a *statement*.

E4 (non-gold)

He locks his *vehicle*, goes inside and tries to do the right thing.

Topic-Aware Evidence Reasoning and Stance-Aware Aggregation

- 1) checking topical consistency between the claim and evidence;
- 2) maintaining topical coherence among multiple pieces of evidence;
- 3) ensuring semantic similarity between the global topic information and the semantic representation of evidence;
- 4) aggregating evidence based on their implicit stances to the claim



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

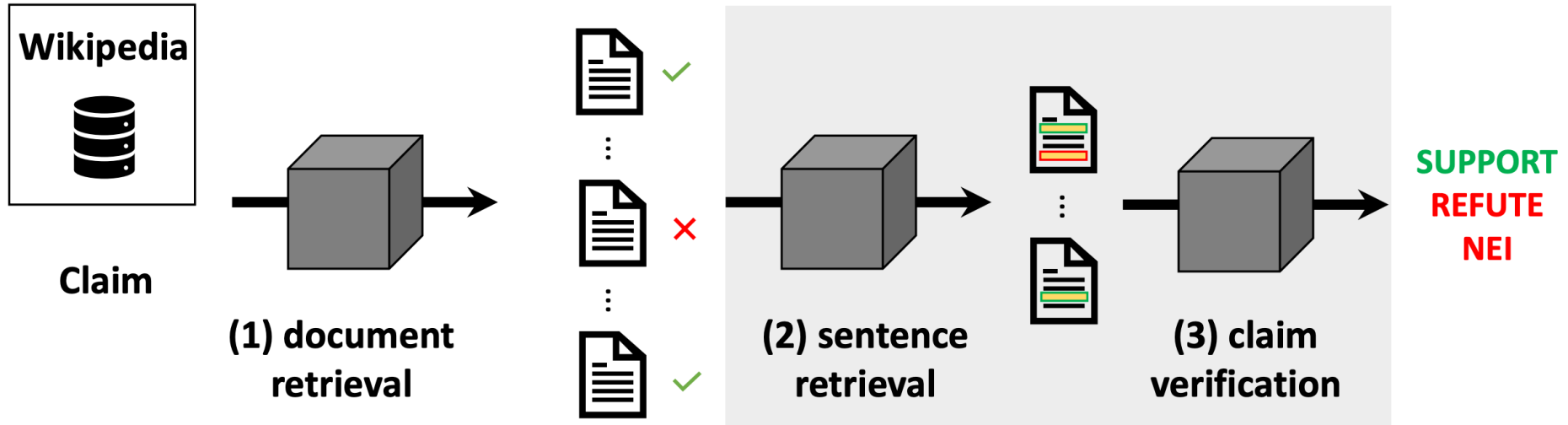
[LEARN MORE](#)

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]

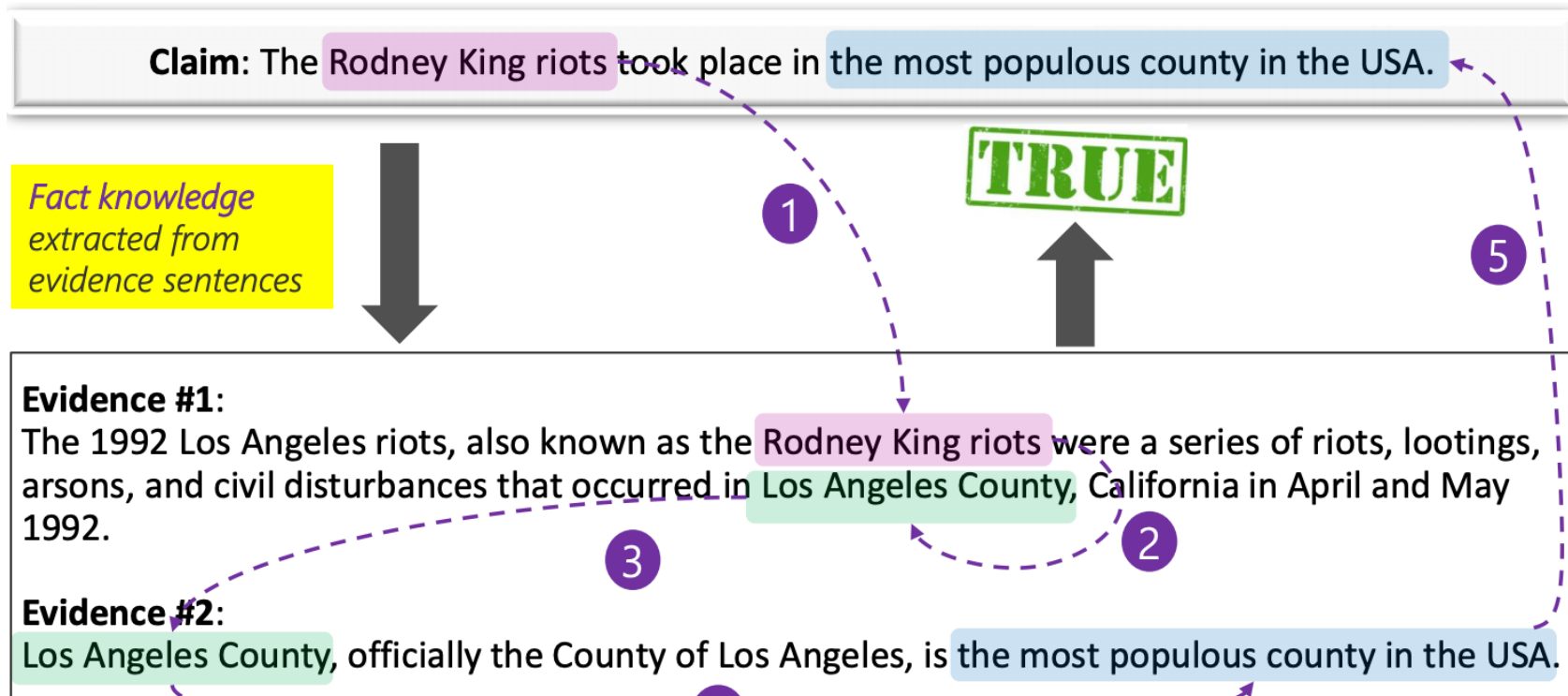
The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[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



Solving FEVER Using Semantic-Level Graphs

Evidence #1:

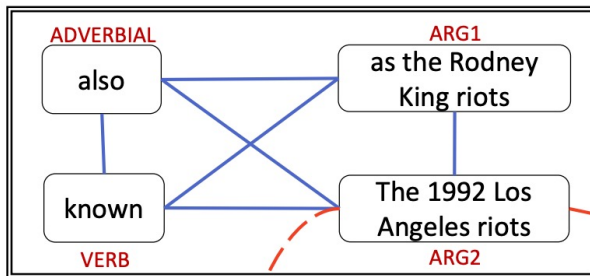
The 1992 Los Angeles riots, also known as the Rodney King riots were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

Evidence #2:

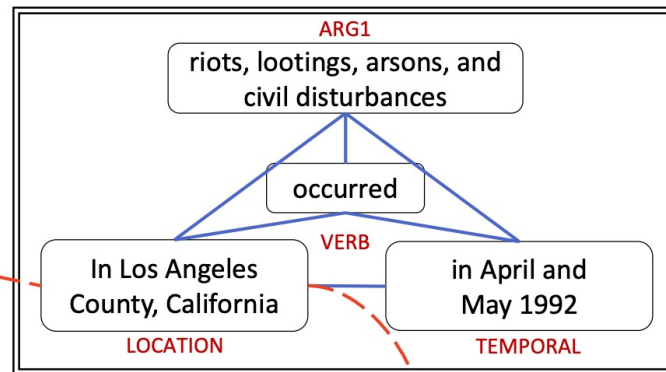
Los Angeles County, officially the County of Los Angeles, is the most populous county in the USA.

Graph Construction

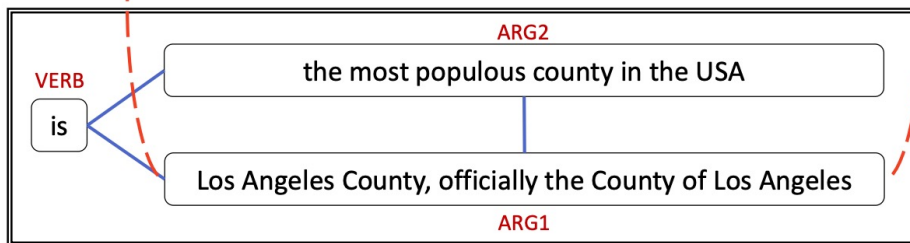
SRL results with verb "known"



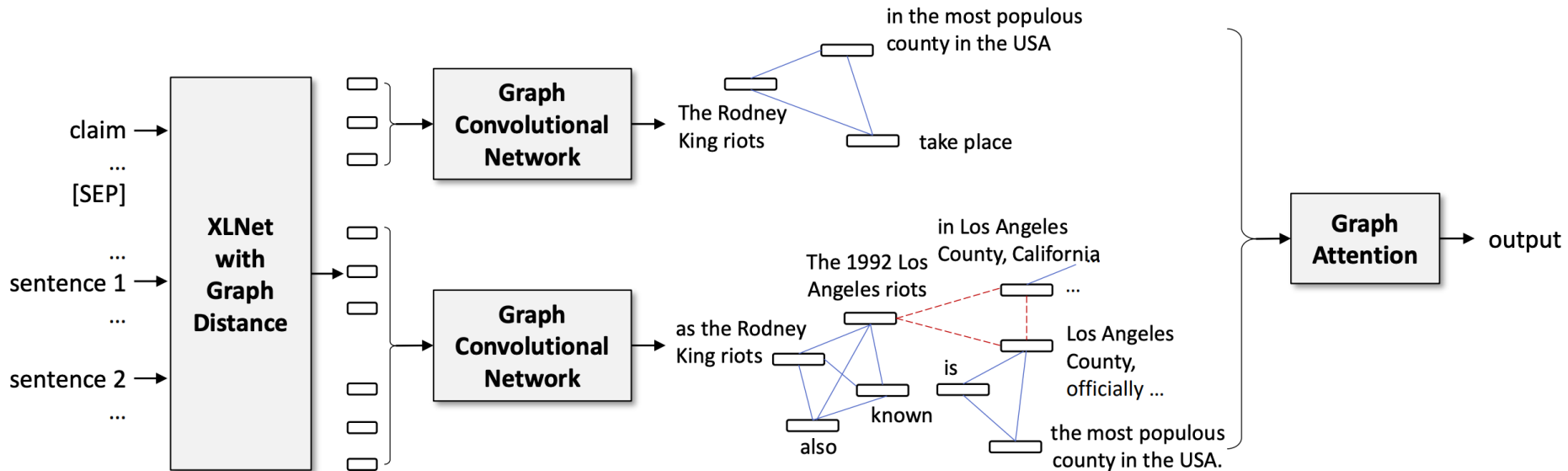
SRL results with verb "occurred"



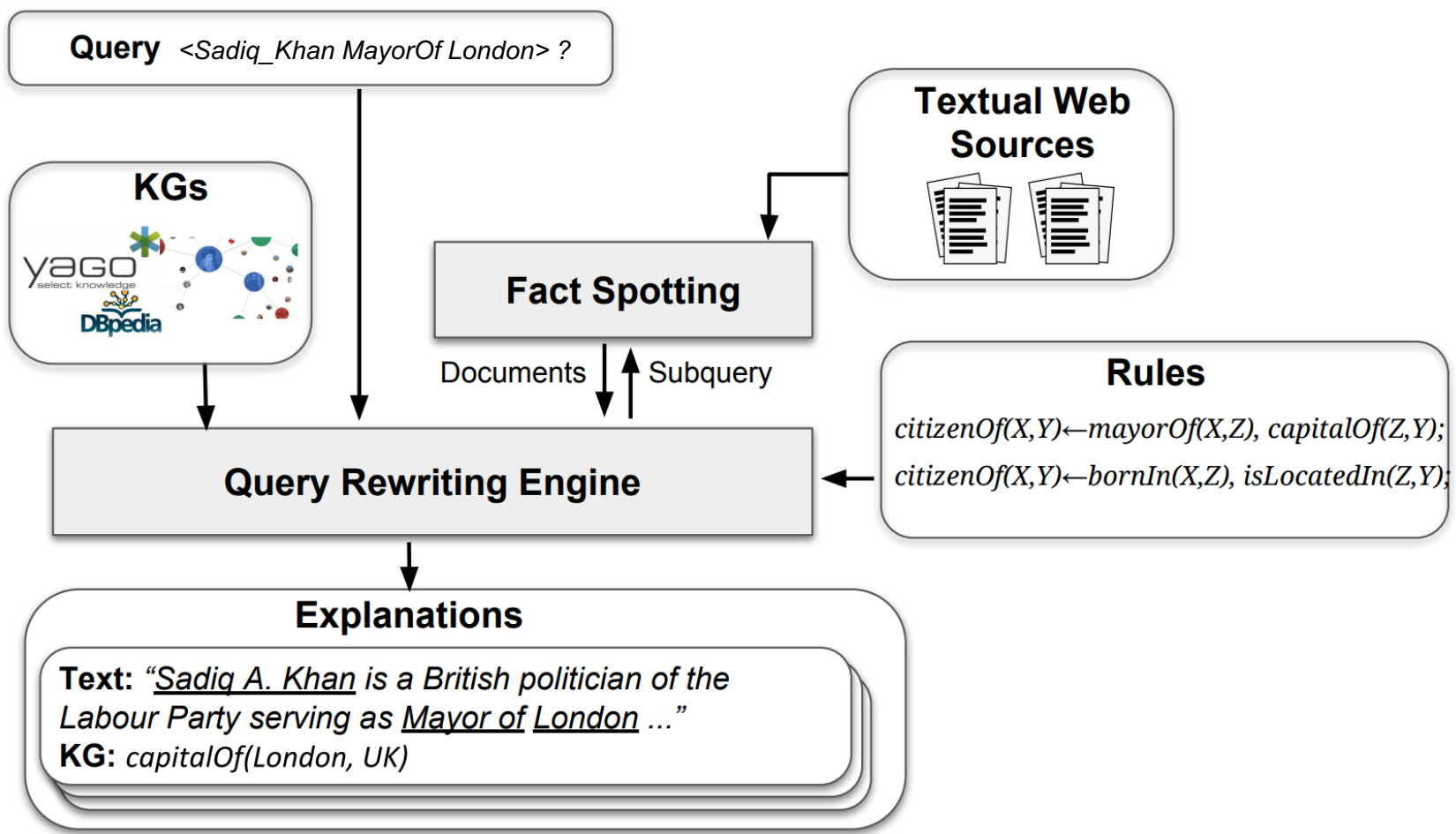
SRL results with verb "is"



Solving FEVER Using Semantic-Level Graphs



Fact-Checking a Claim Using Knowledge Graphs



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).
```

Output →

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

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

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

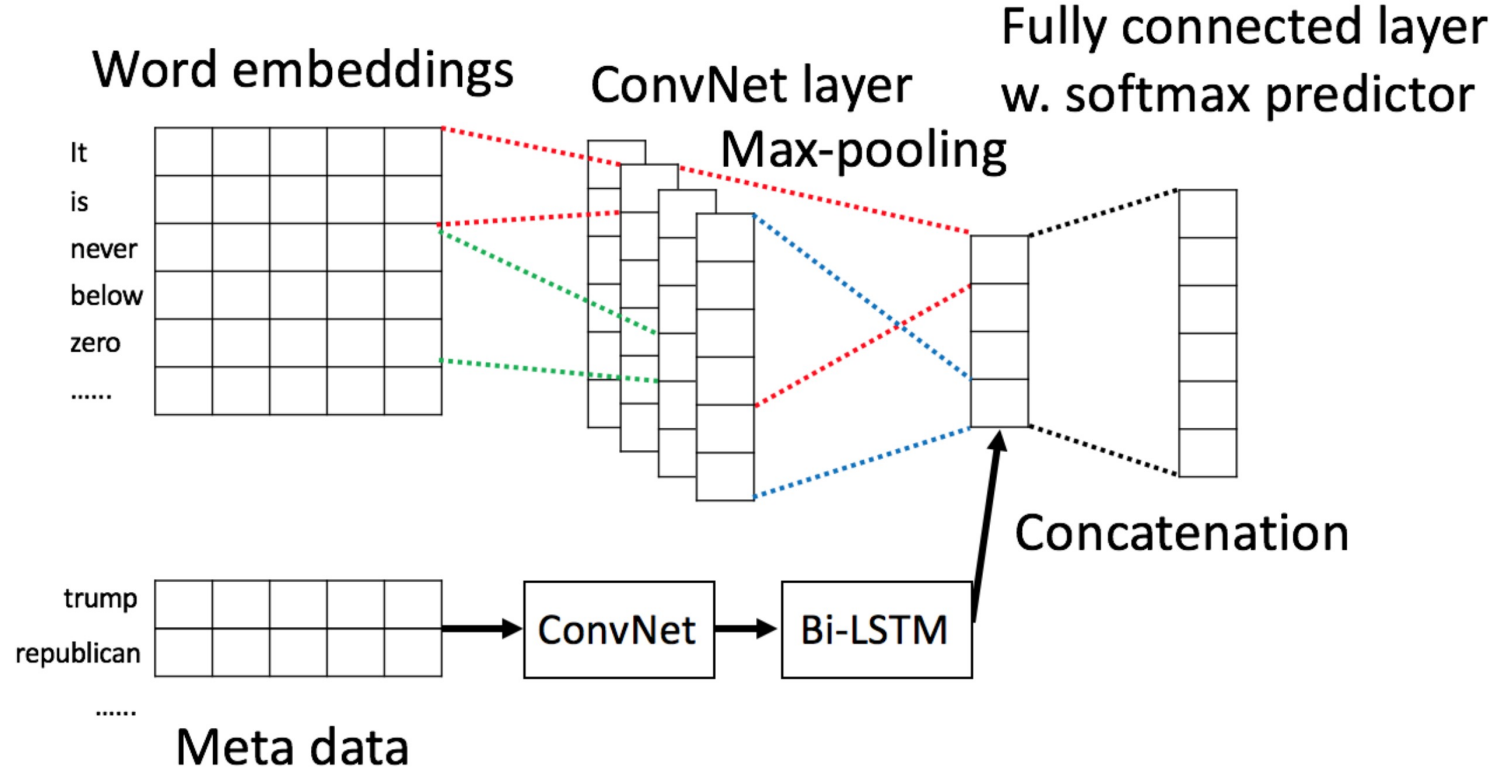
Speaker: Nancy Pelosi

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 dont back that up. We rule this statement False.

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



Fact-Checking a Claim Using Language Models

Fact-Checking

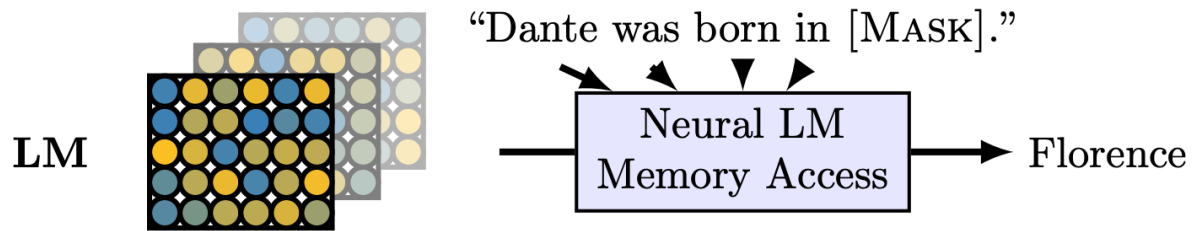
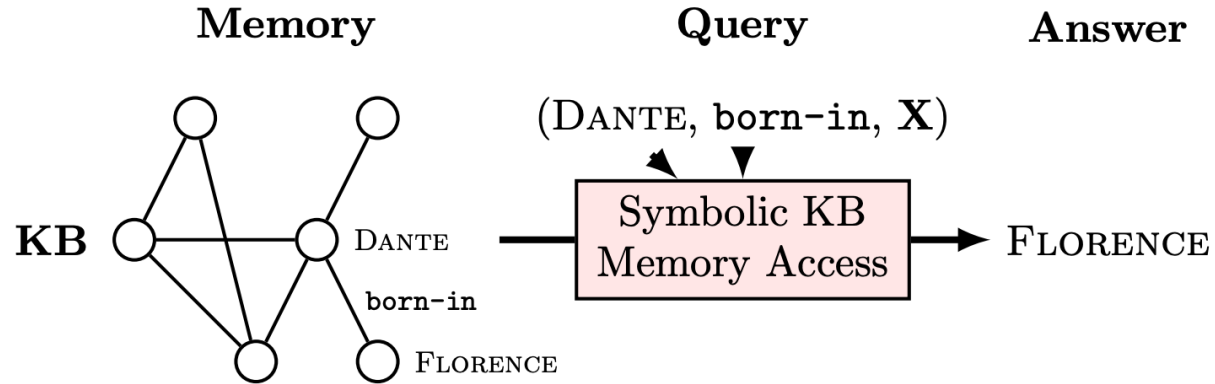
Explainable

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

Non-explainable

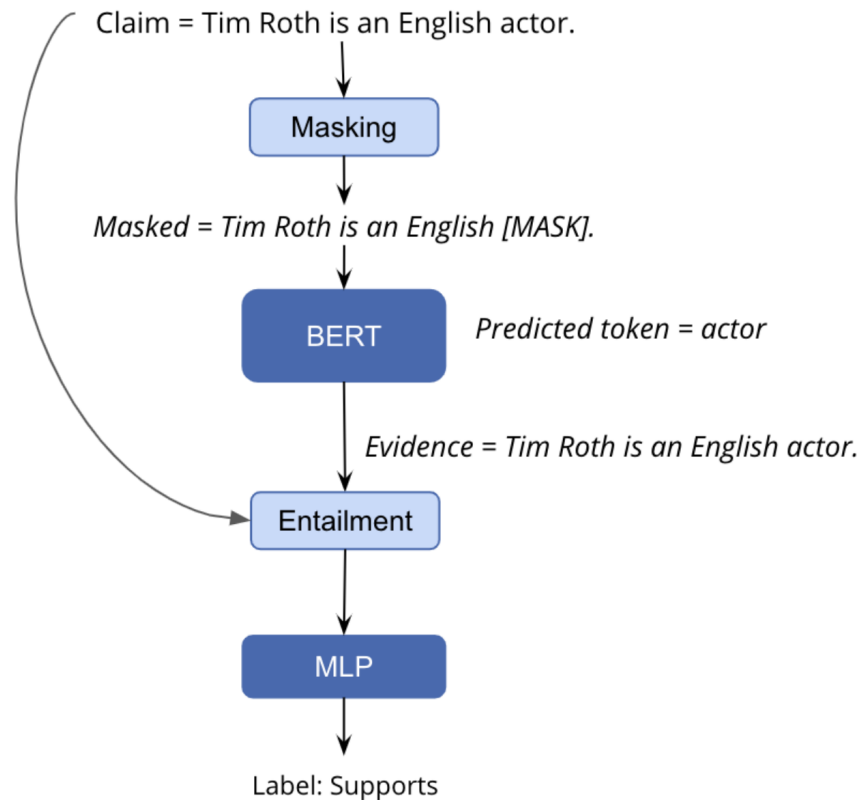
- Claim text
- ~~Perplexity~~
- Language models
- ~~Web~~
- ~~Social media~~

LMs as Knowledge Bases



e.g. ELMo/BERT

LMs as Fact-Checkers



FEVER@ACL-2020: *Language Models as Fact Checkers?*

Nayeon Lee, Belinda Z. Li, Sinong Wang, Wen-tau Yih, Hao Ma, Madian Khabsa

Fact-Checking Claims About Images

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

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

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



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

FACTCHECK.ORG

A tagging system for fact-checks

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

ClaimReview

<https://schema.org/ClaimReview>

[Thing](#) > [CreativeWork](#) > [Review](#) > [ClaimReview](#)

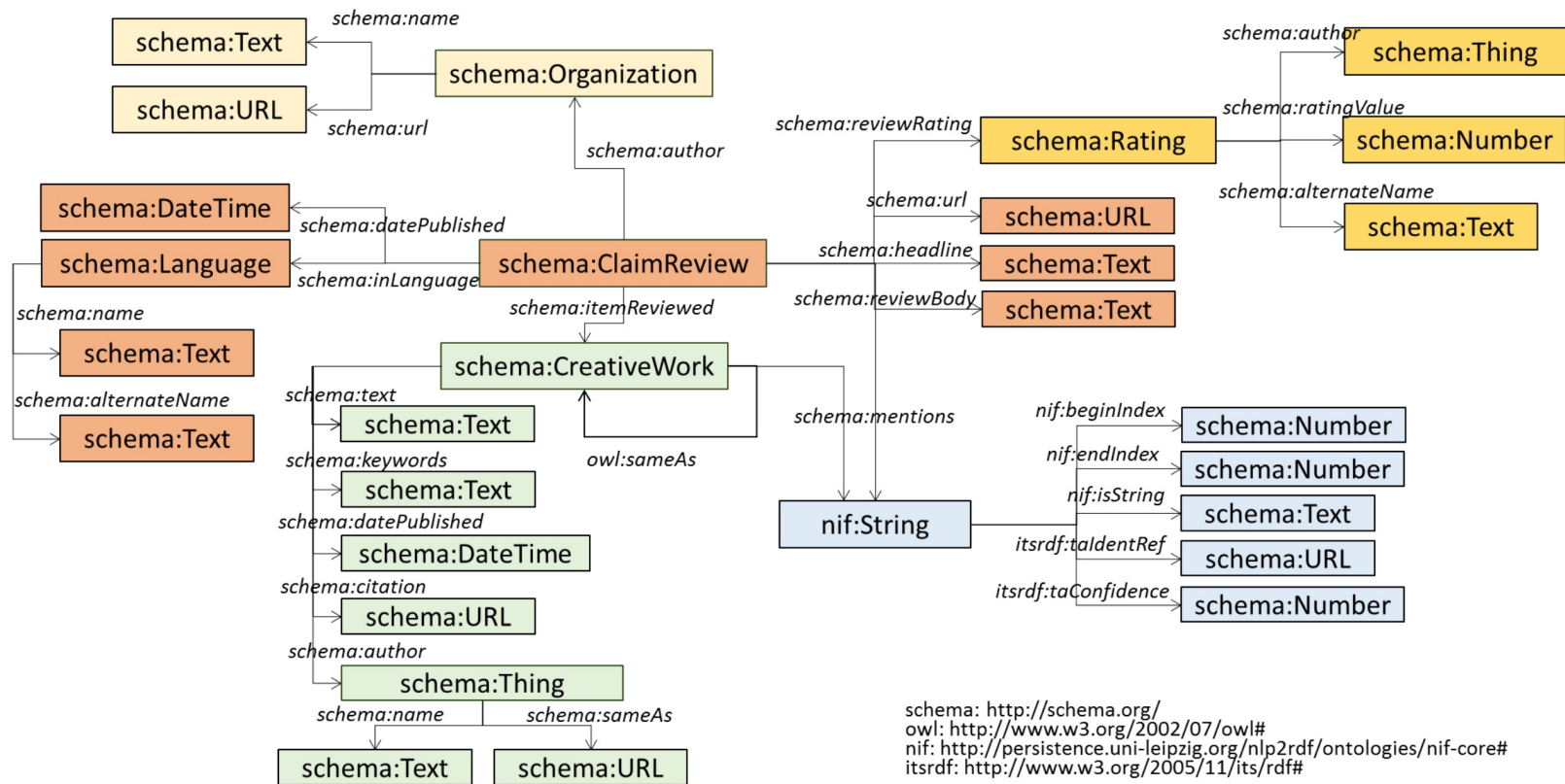
A fact-checking review of claims made (or reported) in some creative work (referenced via itemReviewed).

[\[more...\]](#)

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 .

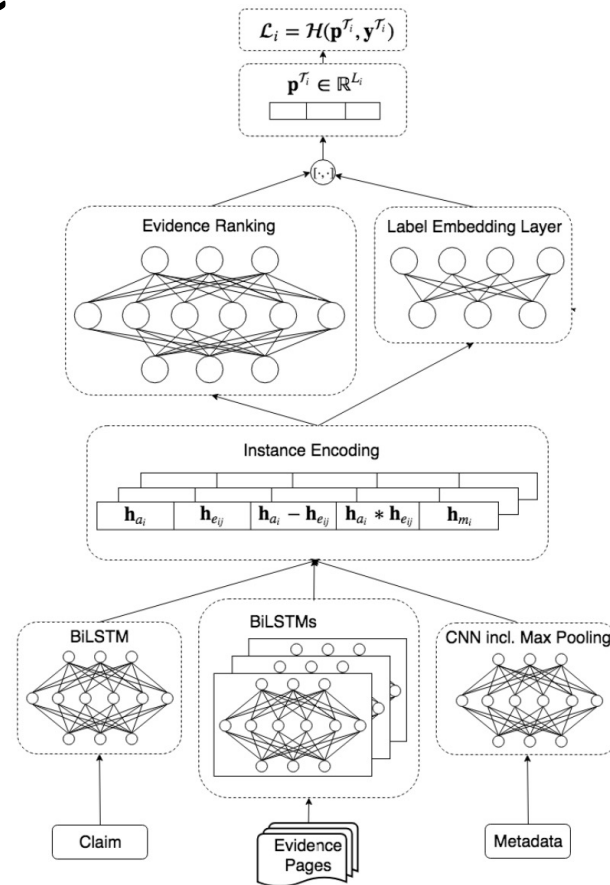
ClaimsKG

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



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

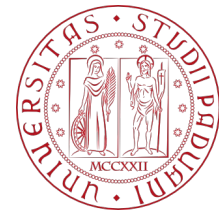


MultiMC

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
clck	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
goop	2943	6	0, 1, 2, 3, 4, 10
hoer	1310	7	facebook scams, true messages, bogus warning, satirical 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 pinnochios, 3 pinnochios, 2 pinnochios, 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

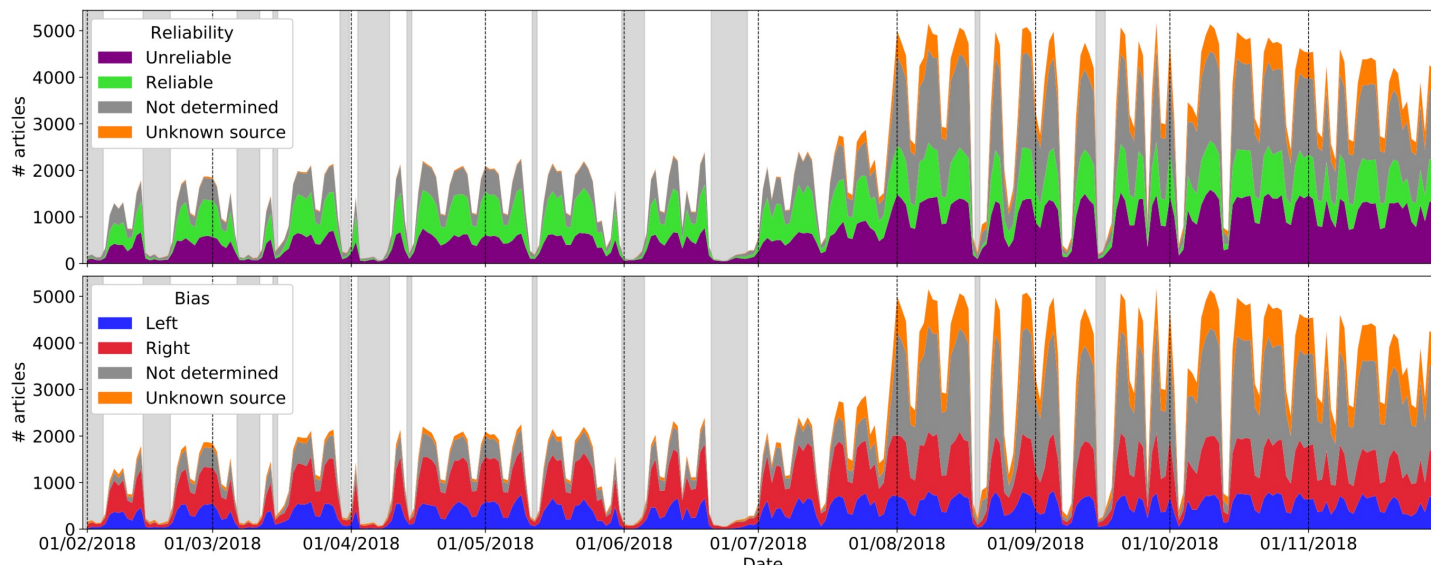


UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Fake News Detection: Datasets

NELA-GT-2018 Fake News Dataset

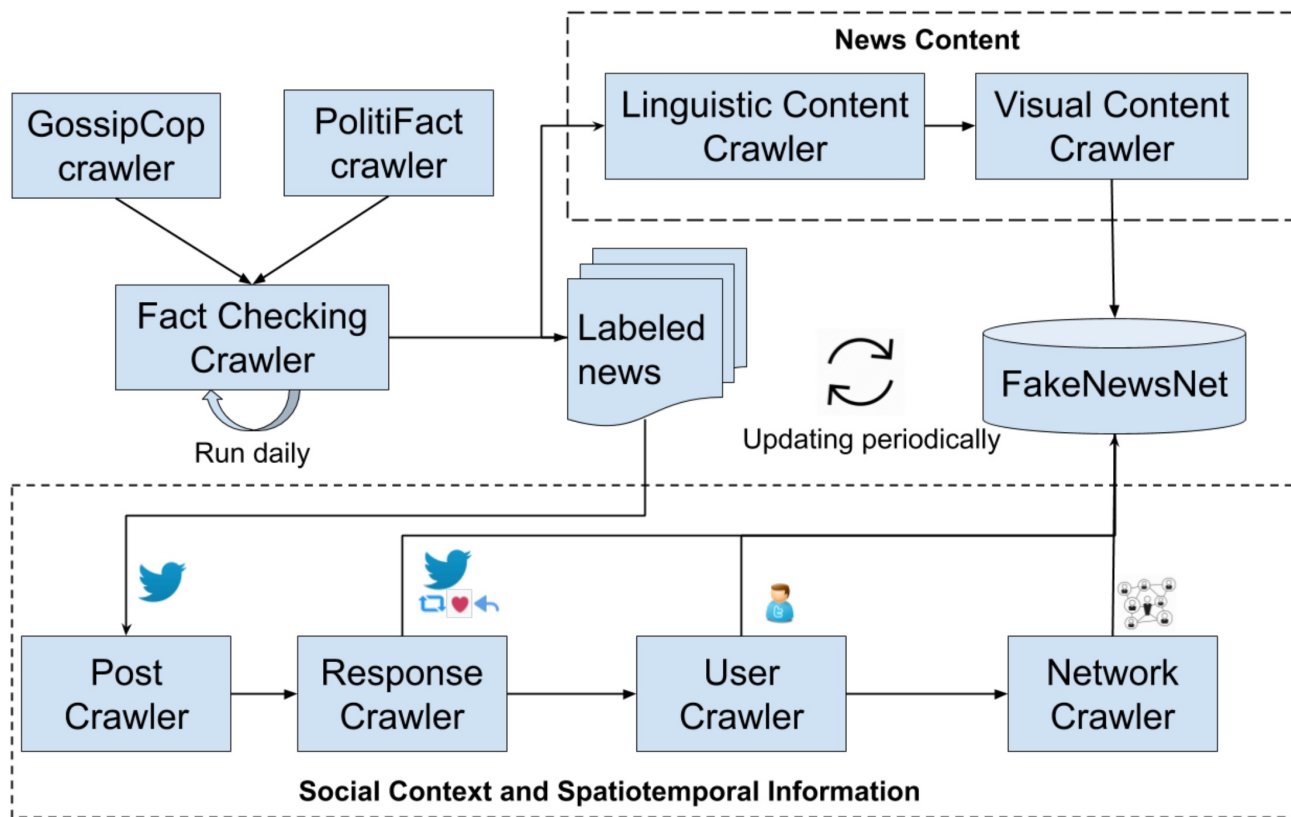
- **713k articles; 194 news and media outlets**
- **distant supervision for annotations**



Jeppe 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

FakeNewsNet



FakeNewsNet

	Category	Features	PolitiFact		GossipCop	
			Fake	Real	Fake	Real
News Content	<i>Linguistic</i>	# News articles	432	624	5,323	16,817
		# News articles with text	420	528	4,947	16,694
	<i>Visual</i>	# News articles with images	336	447	1,650	16,767
Social Context	<i>User</i>	# 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
	<i>Post</i>	# Tweets posting news	164,892	399,237	519,581	876,967
	<i>Response</i>	# 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
	<i>Network</i>	# 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	<i>Spatial</i>	# User profiles with locations	217,379	719,331	429,547	220,264
		# Tweets with locations	3,337	12,692	12,286	2,451
	<i>Temporal</i>	# 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.

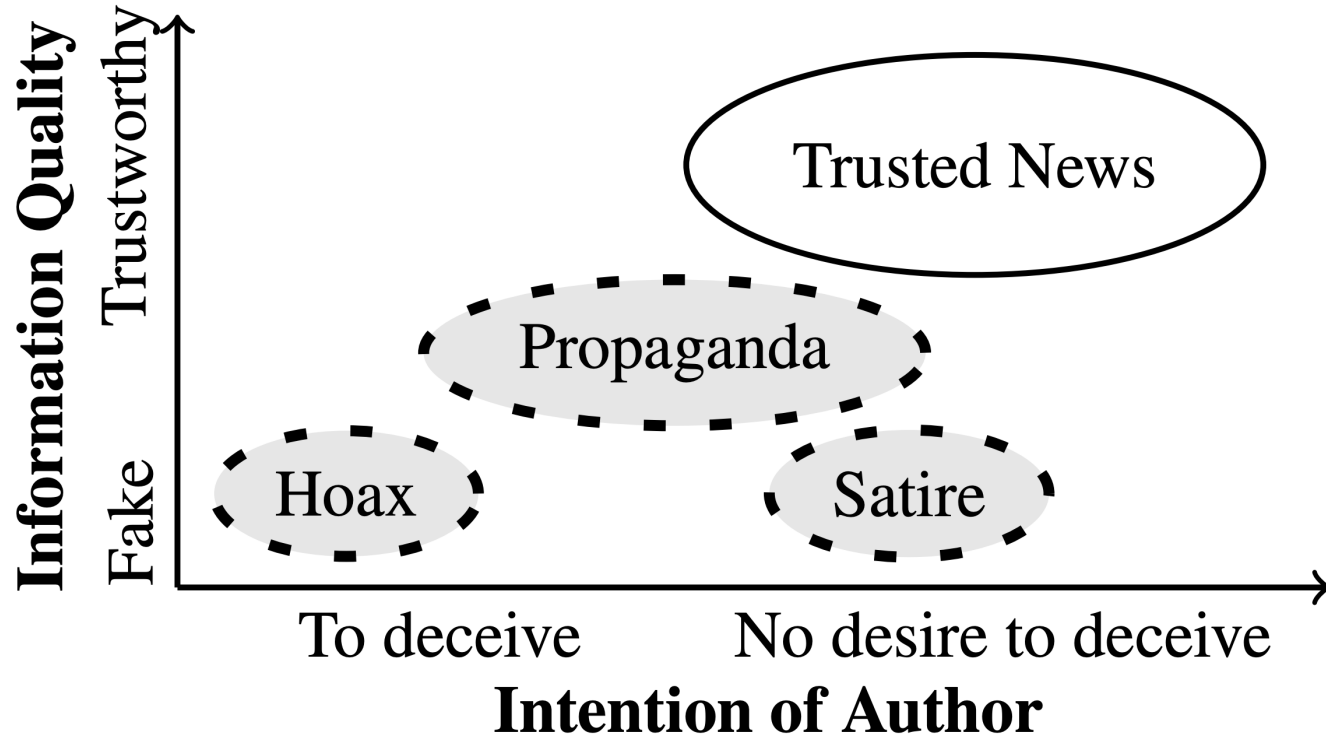
Big data, 8(3), 171-188, 2020.

FakeNewsNet

Dataset	News Content		Social Context				Spatiotemporal Information	
	<i>Linguistic</i>	<i>Visual</i>	<i>User</i>	<i>Post</i>	<i>Response</i>	<i>Network</i>	<i>Spatial</i>	<i>Temporal</i>
BuzzFeedNews	✓							
LIAR	✓							
BS Detector	✓							
CREDBANK	✓		✓	✓			✓	✓
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
Comparitives	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” and effective methods for debunking it has gained renewed

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

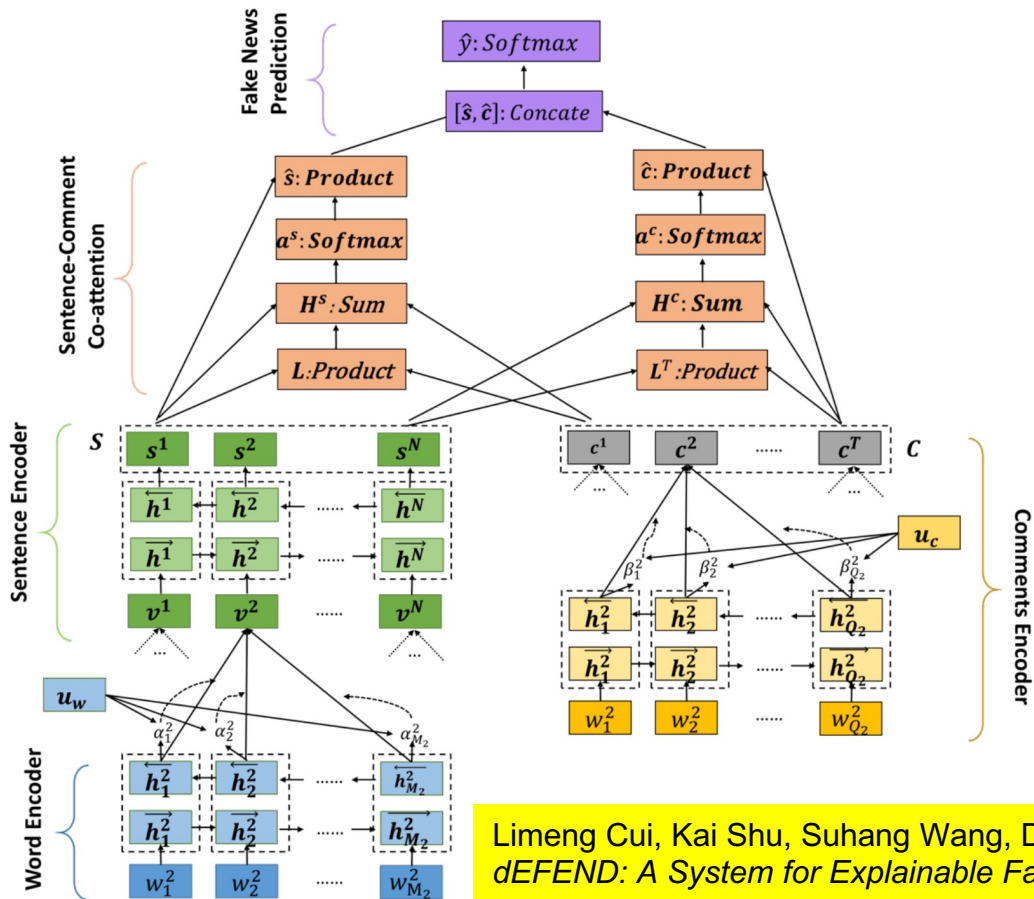
If you had done your research, you would know that the president does not have the power to give citizenship. This would have to done as an act of congress... (0.0160)

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)

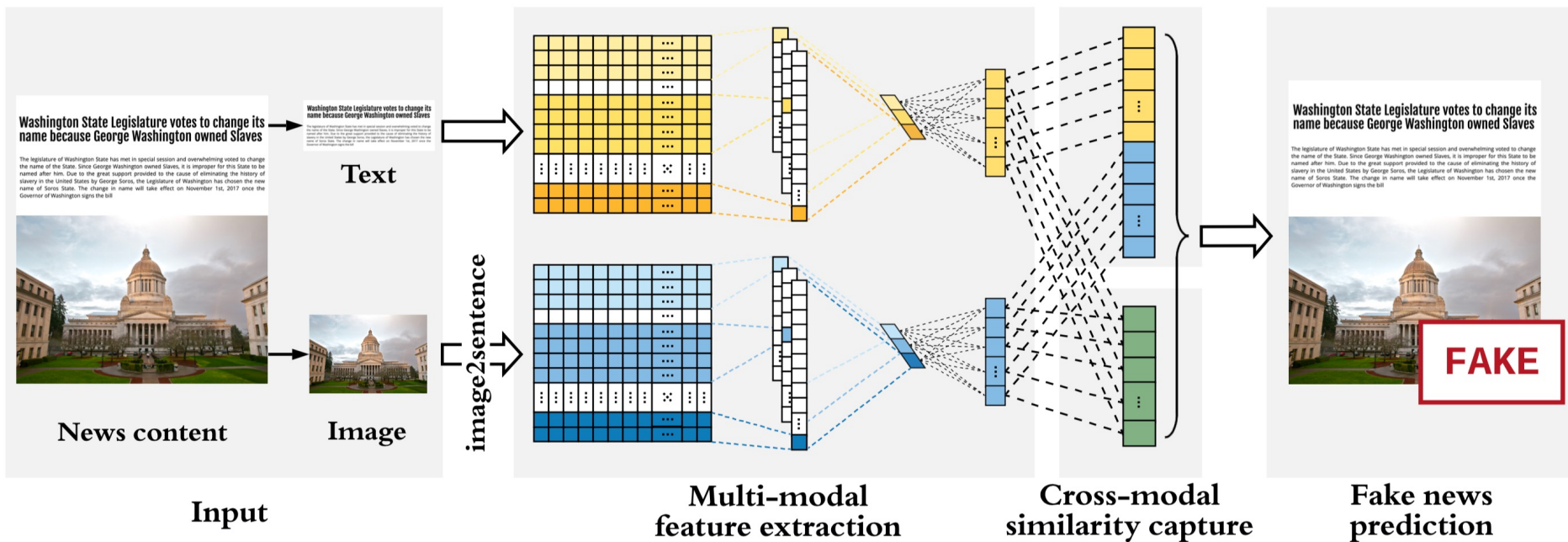
Walkaway from their (0.0080)

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



Xinyi Zhou, Jindi Wu, Reza Zafarani:
SAFE: Similarity-Aware Multi-modal Fake News Detection. PAKDD (2) 2020: 354-367

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:

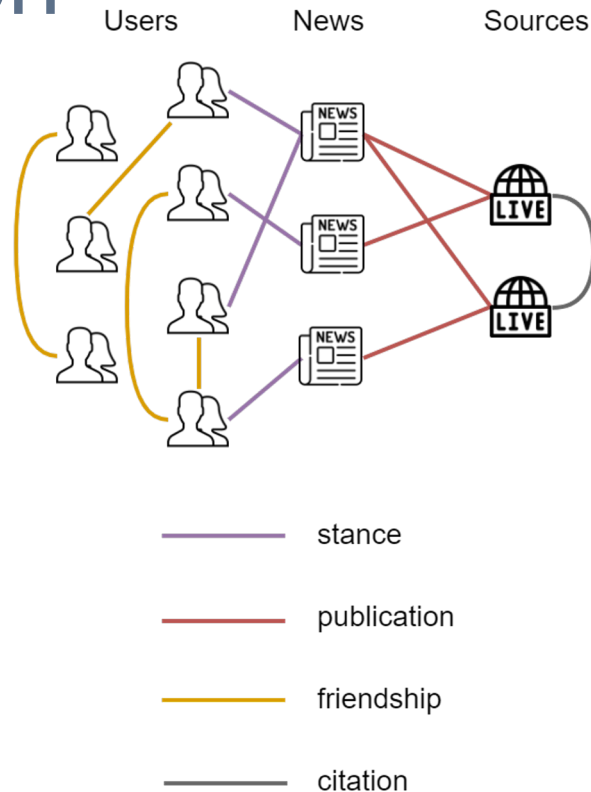
SAFE: Similarity-Aware Multi-modal Fake News Detection. PAKDD (2) 2020: 354-367

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



CIKM-2020: 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

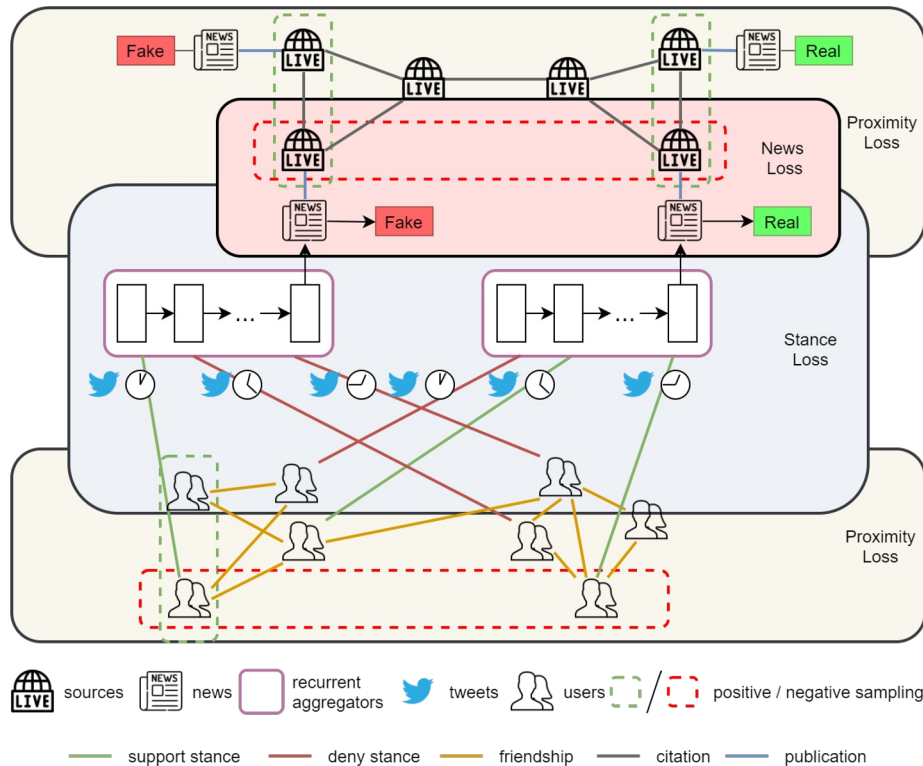
FANG: Factual News Graph

Representations are jointly optimized on

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

Linear combination of the 3 losses

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

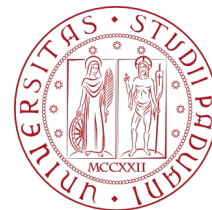


CIKM 2020: Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, Min-Yen Kan:

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

A Survey on Multimodal Disinformation Detection, Firoj Alam, Stefano Cresci, Tanmoy Chakraborty, Fabrizio Silvestri, Dimitar Dimitrov, Giovanni Da San Martino, Shaden Shaar, Hamed Firooz, Preslav Nakov (ArXiv 2021)

Online Propaganda Detection



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

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

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

Detection of Persuasion Techniques

- Novel research direction: detection of persuasion 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
 - ...

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

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

Persuasion Techniques

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

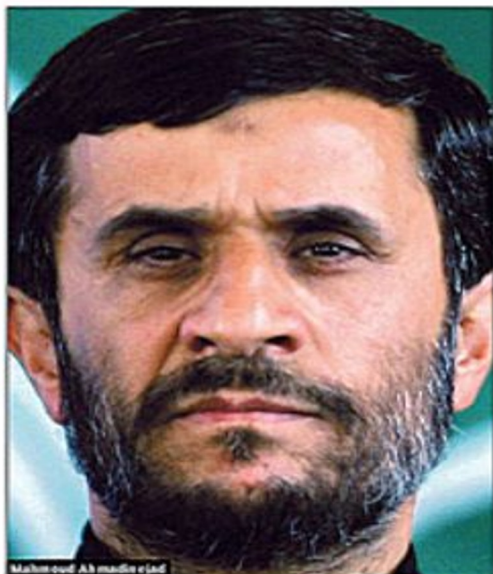
DAILY NEWS

50¢

NEW YORK'S HOMETOWN NEWSPAPER

NYDailyNews.com

THE EVIL HAS LANDED



Mahmoud A3 madia ejad

STORIES
ON
PAGES 4-7
EDITORIAL
PAGE 22



Name Calling



?

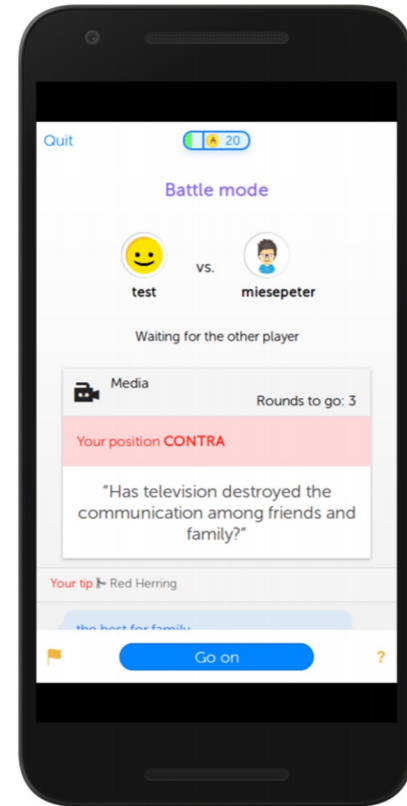
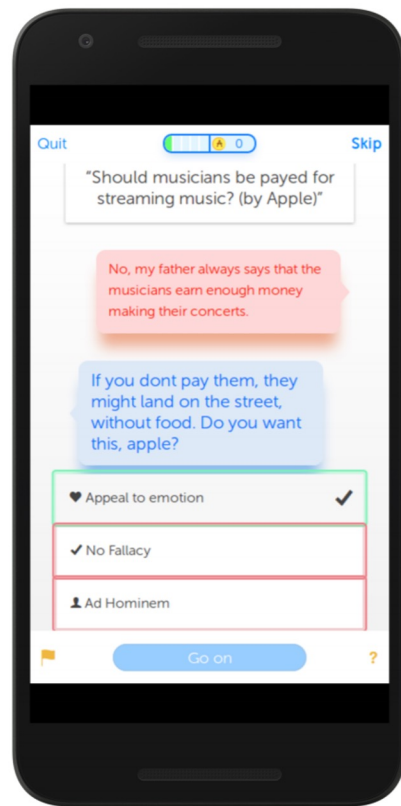
Bandwagon



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



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 with 3,866 *ad hominem* in total

↑ [CMV](#) r/changemyview · Posted by u/ArIkard 2 days ago · 6.7k ↑

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

Delta(s) from OP

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 · [Sticked comment](#) · edited 2 days ago

↓ [u/ArIkard](#) (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](#) | [Deltaboards](#)

↑ jewishcaveman · 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 encoded with word2vec pre-trained embeddings

<u>Model</u>	<u>Accuracy</u>
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.

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

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/ArIkard (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](#) | [Deltaboards](#)

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

1	Manchin says Democrats acted like <u>babies</u> 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 <u>the party is more concerned with obstruction than it is with progress.</u>	Black-and-white_Fallacy
4	In a glaring sign of just how <u>stupid and petty</u> 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 <u>not looking as though Trump killed his grandma.</u>	Loaded_language Exaggeration Loaded_language
6	As Manchin noted, many Democrats bolted as soon as Trump's speech ended in an apparent effort to signal <u>they can't even stomach being in the same room as the president</u>	Exaggeration

PTC Corpus

- 536 news articles in English from 48 sources (450k words) annotated at fragment level with 18 techniques (400 man hours of work) 
- Later extended to 2049 articles in 9 languages¹ (French, German, Italian, Polish, Russian, Greek, Spanish, Georgian)

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

¹ Jakub Piskorski, Nicolas Stefanovitch, Nikolaos Nikolaidis, Giovanni Da San Martino and Preslav Nakov *Multilingual Multifaceted Understanding of Online News in Terms of Genre, Framing, and Persuasion Techniques*

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-level: a sentence is considered as propagandistic if it contains one or more propagandistic fragments.

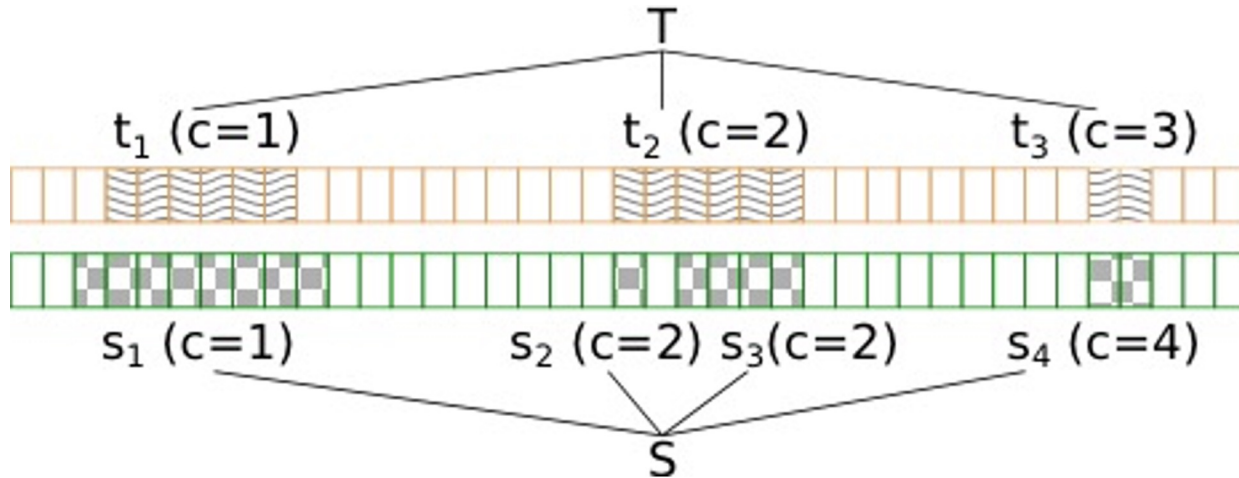
1	Manchin says Democrats acted like babies at the SOTU	Stereotyping_name_calling_or_labeling
2	Democrat West Virginia Sen. Joe Manchin says his colleagues' refusal to stand or applaud during President Donald Trump's State of the Union speech was disrespectful and a signal that the party is more concerned with obstruction than it is with progress.	Black-and-white_Fallacy
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 not looking as though Trump killed his grandma.	Loaded_language Exaggeration Loaded_language
6	As Manchin noted, many Democrats bolted as soon as Trump's speech ended in an apparent effort to signal they can't even stomach being in the same room as the president	Exaggeration

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

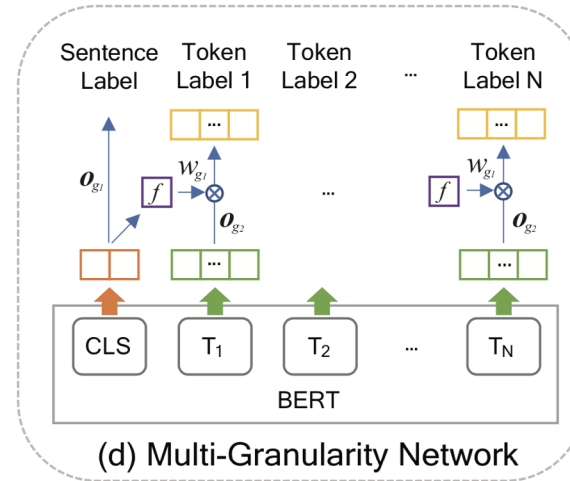
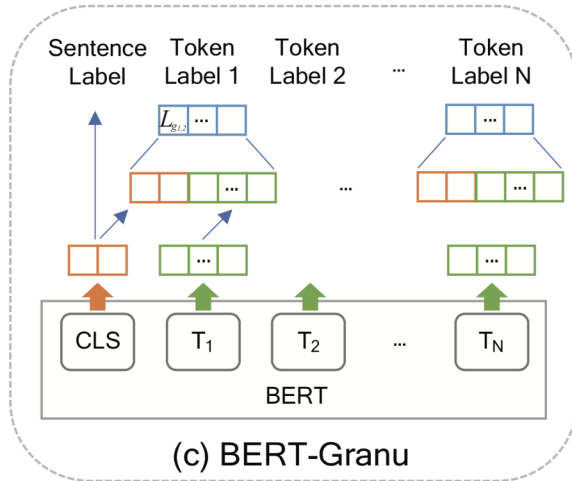
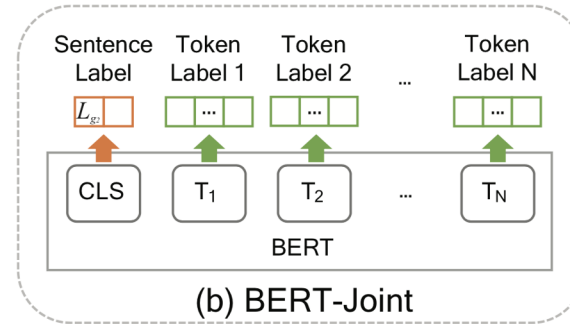
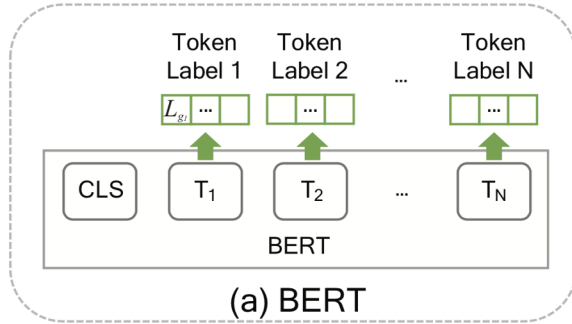
Evaluation Measures

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

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



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 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.

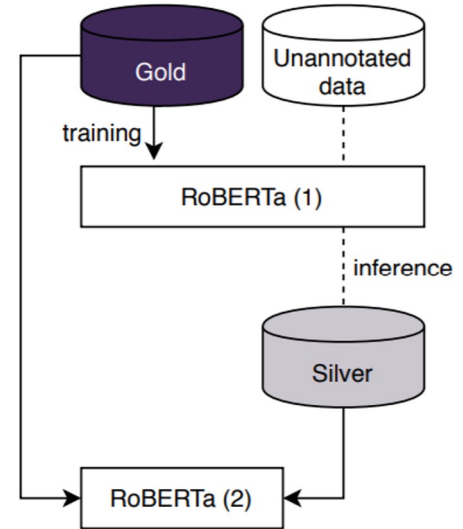
More approaches for the tasks SLC and FLC are described here:
<https://www.aclweb.org/anthology/events/nlp4if-2019/>

Task S1 - 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>

Persuasion Techniques in Memes

- Most communication in social media is multimodal, mixing textual with visual content
- SemEval 2021 task 6: 950 memes annotated with 22 techniques¹
- New Data (10K memes!) and a new shared task are coming soon!²

¹ Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, Giovanni Da San Martino: Detecting Propaganda Techniques in Memes. ACL/IJCNLP (1) 2021: 6603-6617
² SemEval 2024 Task 4: <https://semeval.github.io>



Conclusions and Future Work

- Propaganda detection, especially the detection of propaganda techniques is a novel task.
- Although results are encouraging, there 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

Malicious Users

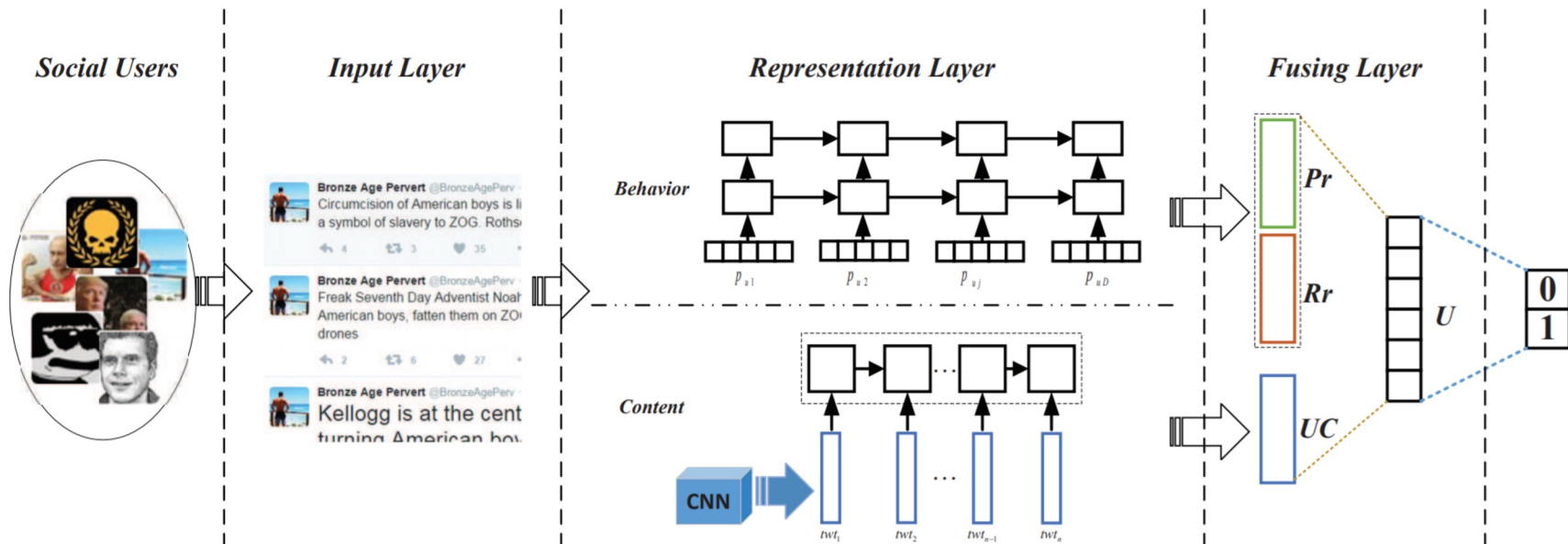
Detecting Fake Accounts: Botometer



- Given a Twitter account, Botometer extracts over 1,000 features relative to the account
- Produces a classification score called bot score: the higher the score, the greater the likelihood that the account is controlled by software,
- Botometer additionally reports six sub-scores, each produced by a model based on a distinct subset of features: user meta-data, friends, content, sentiment, network, and timing
- **Drawbacks:**
 - lack of reliable ground truth
 - malicious actors evolve to avoid detection

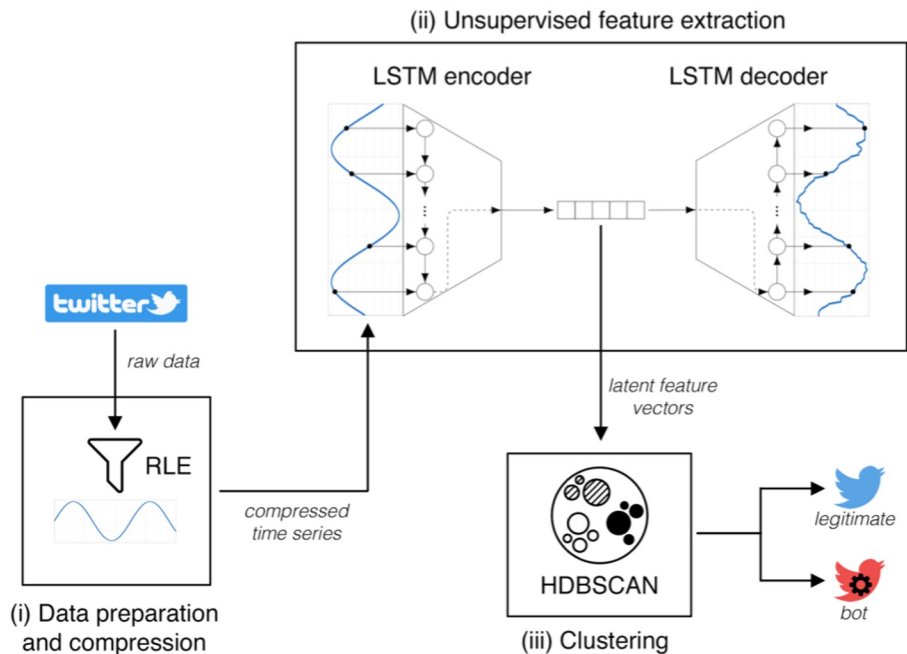
Deep Bot Detection

Devise a user representation based on behaviour (posting, retweeting) and sequence of tweets' content



Detecting Coordinated Behaviour: RTBust

- Detecting coordinated behaviour instead of fake accounts
 - Encode and cluster retweet patterns
 - Discriminate between normal vs. inauthentic behavior
 - Rationale: humans exhibit more behavioral heterogeneity than bots



Discussion

- Malicious users find new approach to tamper the online environment ----- such approach is identified ----- a countermeasure is devised
- **Challenging problem:** most machine learning techniques are designed for stationary and neutral environments
- **Coordinated behavior is not necessarily harmful**
 - An analysis of the content of the messages may show that (for example that it is a propaganda campaign)

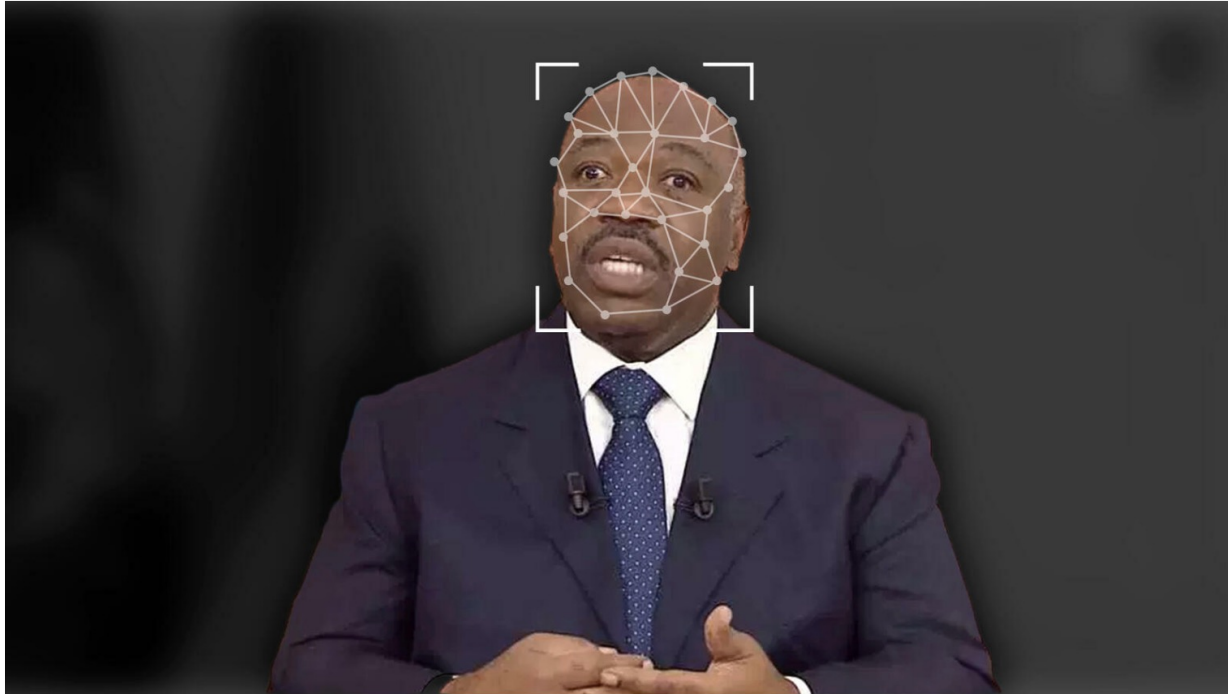
Neural Fake News

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

Fake News Generation Can Now Be Automated

GPT-2: <https://talktotransformer.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/>

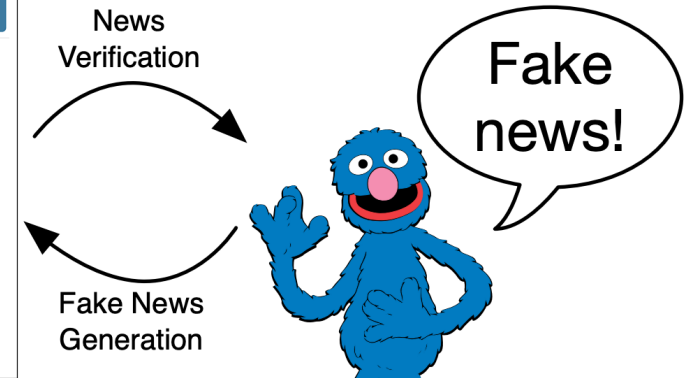
Grover: Defending Against Neural Fake News ...Using Fake News Generators...

SCIENCE The New York Times SUBSCRIBE NOW LOG IN

Link Found Between Vaccines and Autism

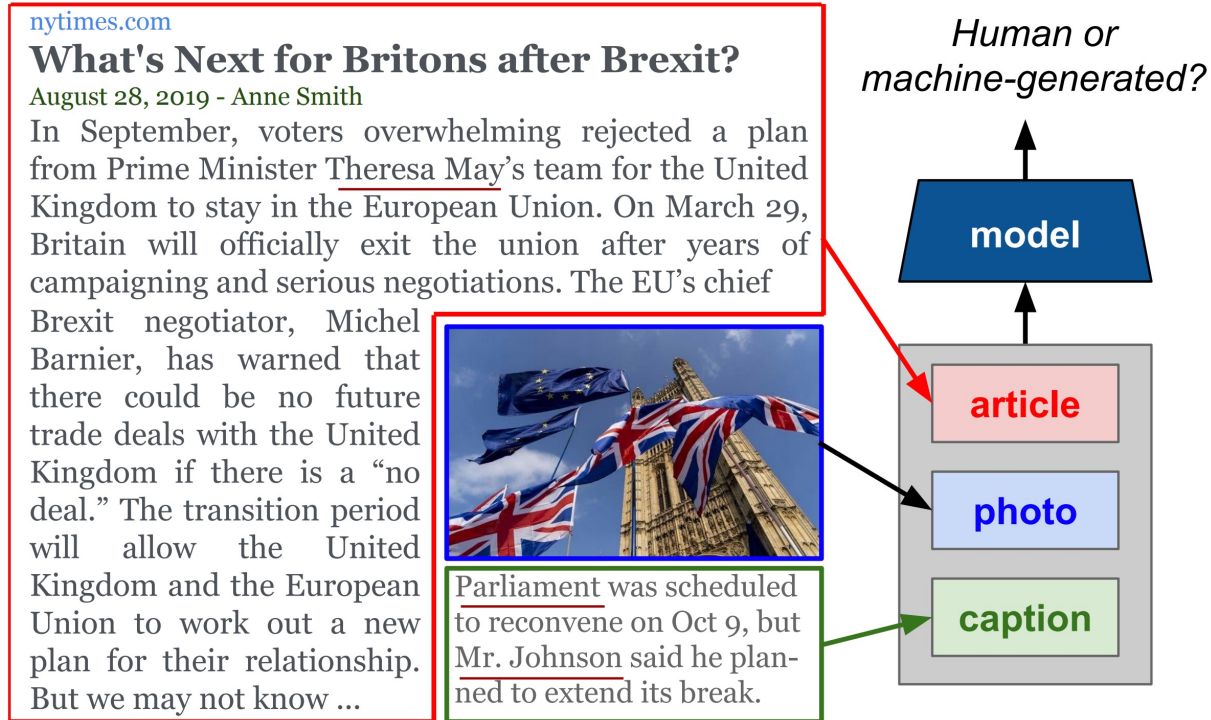
By Paul Waldman May 29, 2019

Those who have been vaccinated against measles have a more than 5-fold higher chance of developing autism, researchers at the University of California San Diego School of Medicine and the Centers for Disease Control and Prevention report today in the Journal of Epidemiology and Community Health. *(continued)*



Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, Yejin Choi:
"Defending against neural fake news." *NeuroIPS*, 32 (2019).

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



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

References (1)

- Qi, Peng, Juan Cao, Xirong Li, Huan Liu, Qiang Sheng, Xiaoyue Mi, Qin He, Yongbiao Lv, Chenyang Guo, and Yingchao Yu. "Improving Fake News Detection by Using an Entity-enhanced Framework to Fuse Diverse Multimodal Clues." In Proceedings of the 29th ACM International Conference on Multimedia, pp. 1212-1220. 2021.
- Dou, Yingtong, Kai Shu, Congying Xia, Philip S. Yu, and Lichao Sun. "User preference-aware fake news detection." In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2051-2055. 2021.
- Qian, Shengsheng, Jinguang Wang, Jun Hu, Quan Fang, and Changsheng Xu. "Hierarchical multi-modal contextual attention network for fake news detection." In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 153-162. 2021.
- Kou, Ziyi, Daniel Zhang, Lanyu Shang, and Dong Wang. "What and Why Towards Duo Explainable Fauxtography Detection under Constrained Supervision." IEEE Transactions on Big Data (2021).
- Sheng, Qiang, Juan Cao, Xueyao Zhang, Xirong Li, and Lei Zhong. "Article Reranking by Memory-Enhanced Key Sentence Matching for Detecting Previously Fact-Checked Claims." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 5468-5481. 2021.
- Kazemi, Ashkan, Kiran Garimella, Devin Gaffney, and Scott Hale. "Claim Matching Beyond English to Scale Global Fact-Checking." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 4504-4517. 2021.

References (2)

- Schlichtkrull, Michael Sejr, Vladimir Karpukhin, Barlas Oguz, Mike Lewis, Wen-tau Yih, and Sebastian Riedel. "Joint Verification and Reranking for Open Fact Checking Over Tables." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 6787-6799. 2021.
- Si, Jiasheng, Deyu Zhou, Tongzhe Li, Xingyu Shi, and Yulan He. "Topic-Aware Evidence Reasoning and Stance-Aware Aggregation for Fact Verification." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 1612-1622. 2021.
- Gupta, Ashim, and Vivek Srikumar. "X-Fact: A New Benchmark Dataset for Multilingual Fact Checking." In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pp. 675-682. 2021.
- Kotonya, Neema, and Francesca Toni. "Explainable Automated Fact-Checking for Public Health Claims." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 7740-7754. 2020.
- Shraman Pramanick, Dimitar Dimitrov, Rituparna Mukherjee, Shivam Sharma, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty: Detecting Harmful Memes and Their Targets. ACL/IJCNLP (Findings) 2021: 2783-2796
- Shraman Pramanick, Shivam Sharma, Dimitar Dimitrov, Md. Shad Akhtar, Preslav Nakov, Tanmoy Chakraborty: MOMENTA: A Multimodal Framework for Detecting Harmful Memes and Their Targets. EMNLP (Findings) 2021: 4439-4455
- Momchil Hardalov, Arnav Arora, Preslav Nakov, Isabelle Augenstein: Cross-Domain Label-Adaptive Stance Detection. EMNLP (1) 2021: 9011-9028

References (3)

- 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 Zaghoulani, 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: Propopy: A System to Unmask Propaganda in Online News. AAI 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 (4)

- 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 (5)

- Jeppe 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. AACL 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 (6)

- Kashyap Papat, Subhabrata Mukherjee, Jannik Strötgen, Gerhard Weikum:
Credibility Assessment of Textual Claims on the Web. CIKM 2016: 2173-2178
- Kashyap Papat, 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 Màrquez, 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)