

# **FAKE NEWS DETECTION USING NLP**



# Content

- news circulation process has changed
  - in the past it was edited by professionals
  - now it is edited by everyone
- “fake news” circulates largely due to social networks
  - 30-40%
- technically it is possible to interfere with circulation of “fake news”
  - but cannot be stopped entirely
- conclusion
  - yes we can but should we

# The news process



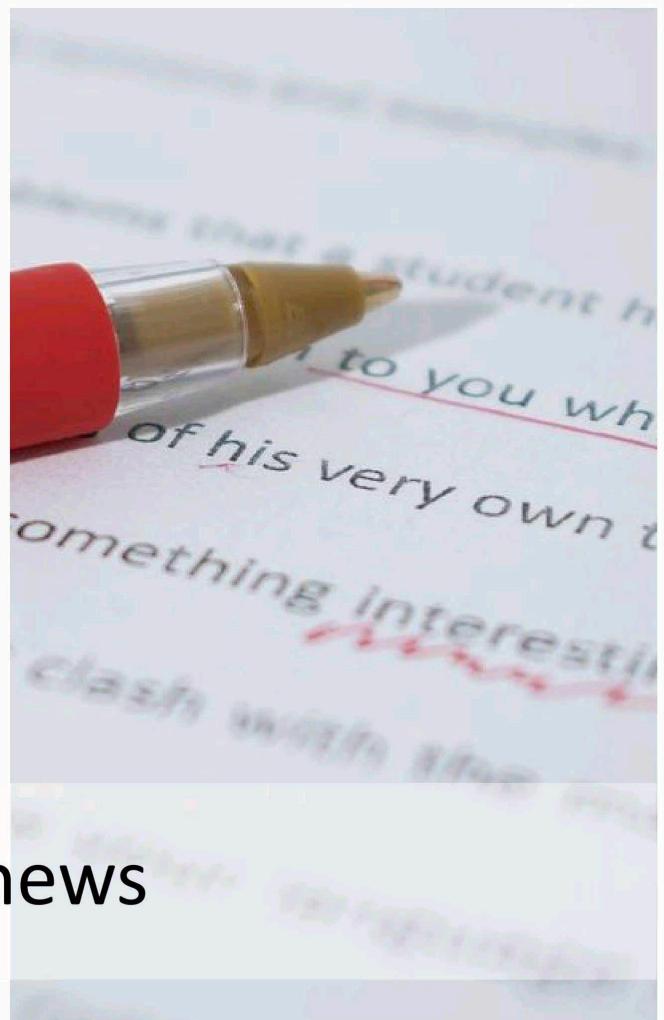
Travel is to make a journey or to have an adventure to somewhere by bicycle, train, airplane, car, motorcycle, or boat. It could be an exploration to somewhere new planned or unplanned to meet new people, new things and new places. There are different types of adventures waiting for you to explore.

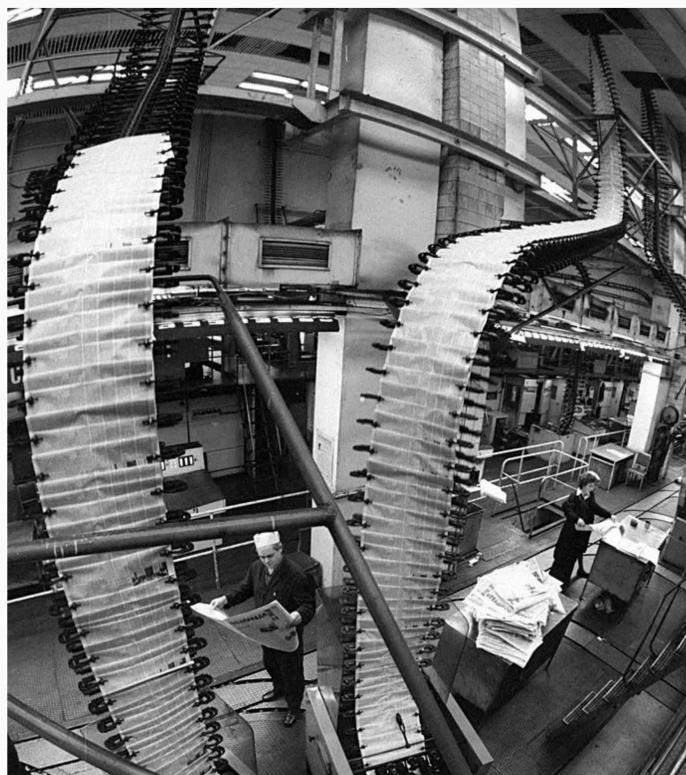
There are lots of places to explore. Places could be urban or suburban. Some people prefer to be with nature to free their minds and refresh their souls, but some like to be in the city. You will get lots of benefits such as exploring new culture.

## Create news



Quality control of news





Publication of news

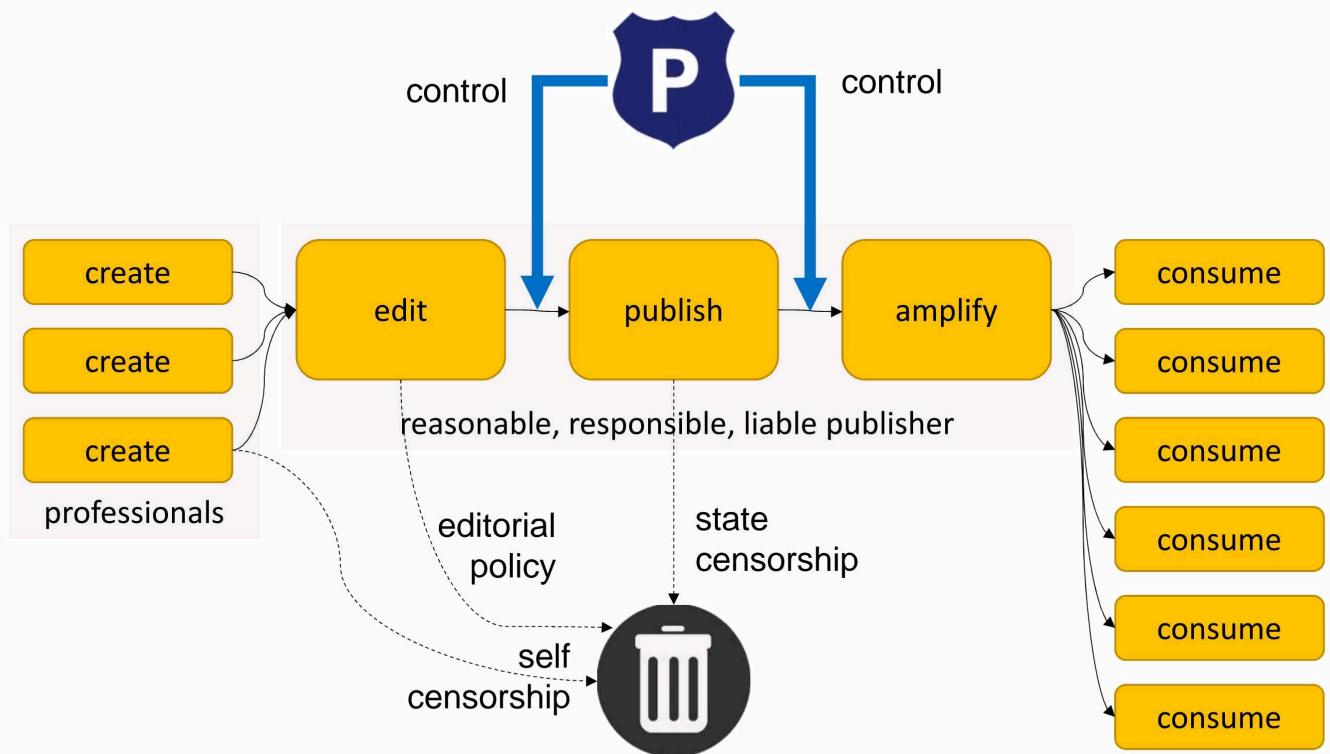




Consumption of news

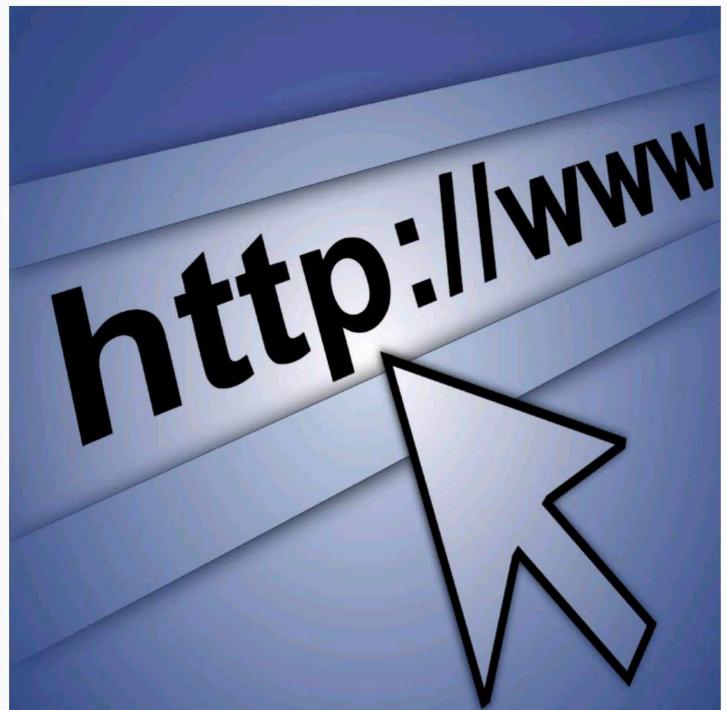
# The news process schematically

# Traditional news process



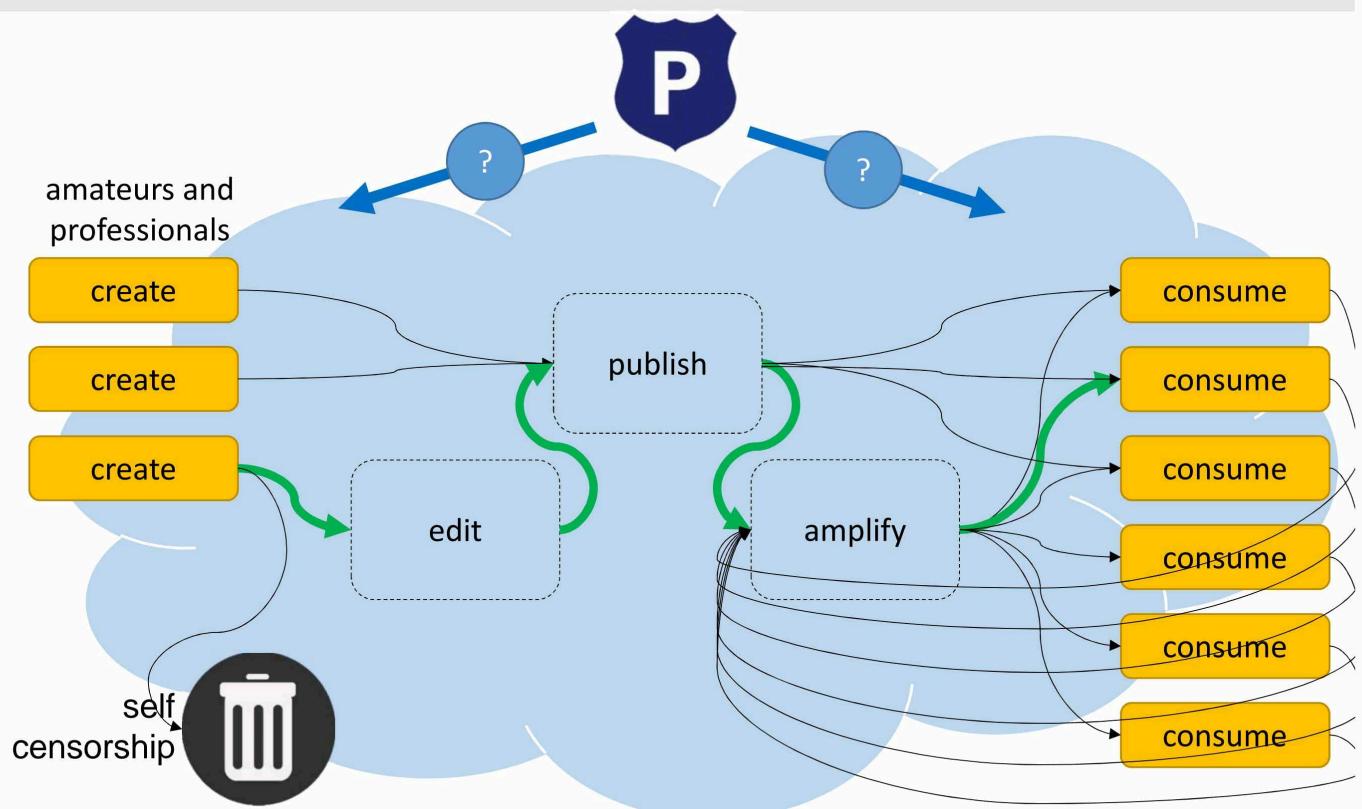
## Enters the Internet

- anyone can publish
- anyone can amplify
- anyone can edit
- editors are gone
- no trashcan



source: wikipedia

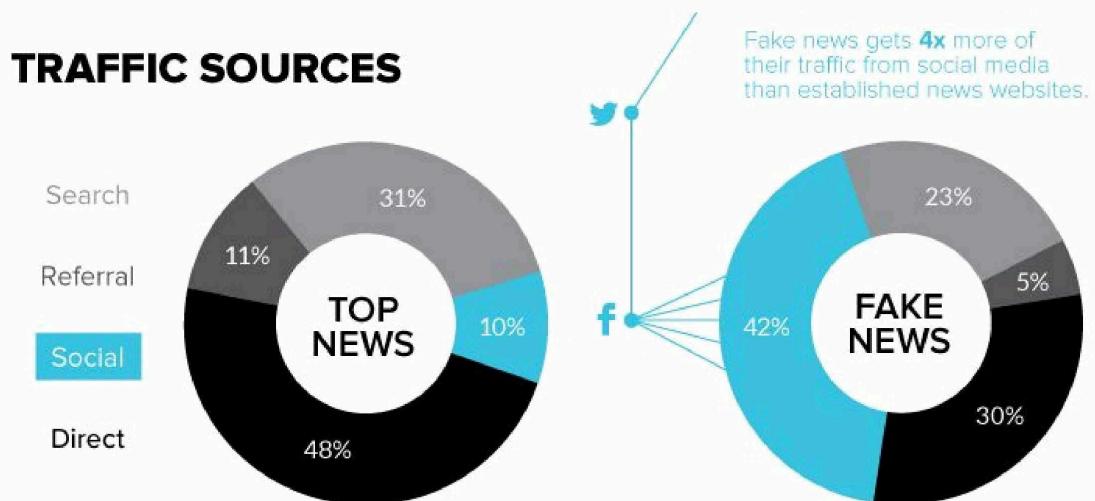
# Internet-age news process



Technology for “fake news” dissemination

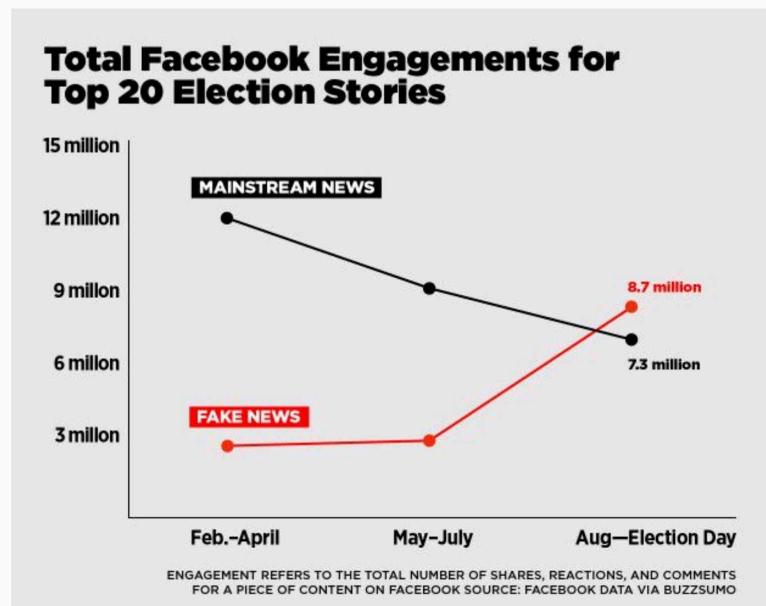
# Social media is main driver for “fake news”

- 10% of readers of top news come via social media
- 40% of readers of “fake news” come via social media



## Fake and real news have about the same engagement on social media

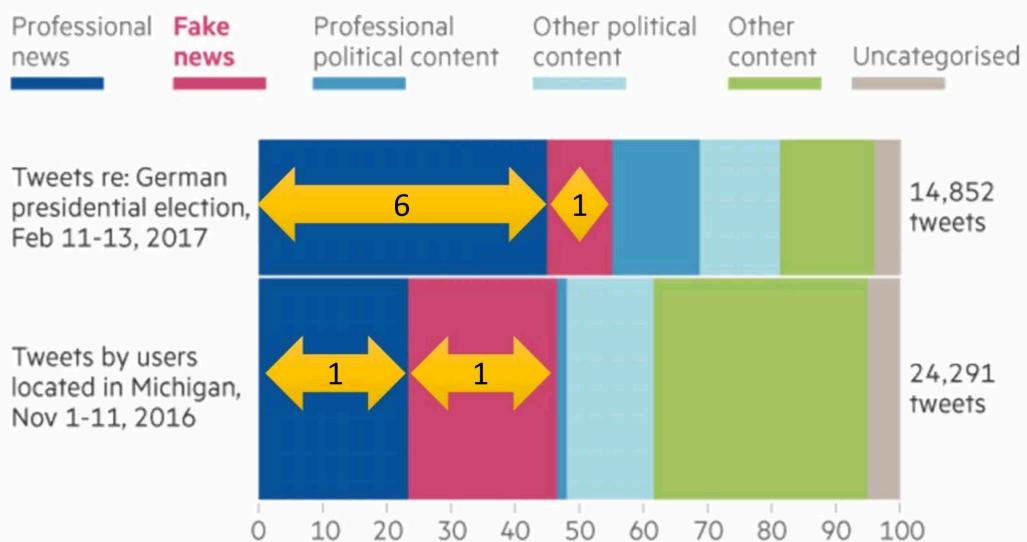
- this is 10% for real news
- but 40% for “fake news”



# In Europe “real” beats “fake” 6:1, in US it is 1:1

University of Oxford studies show differing rates at which “fake news” is shared on Twitter

Percentage of links shared



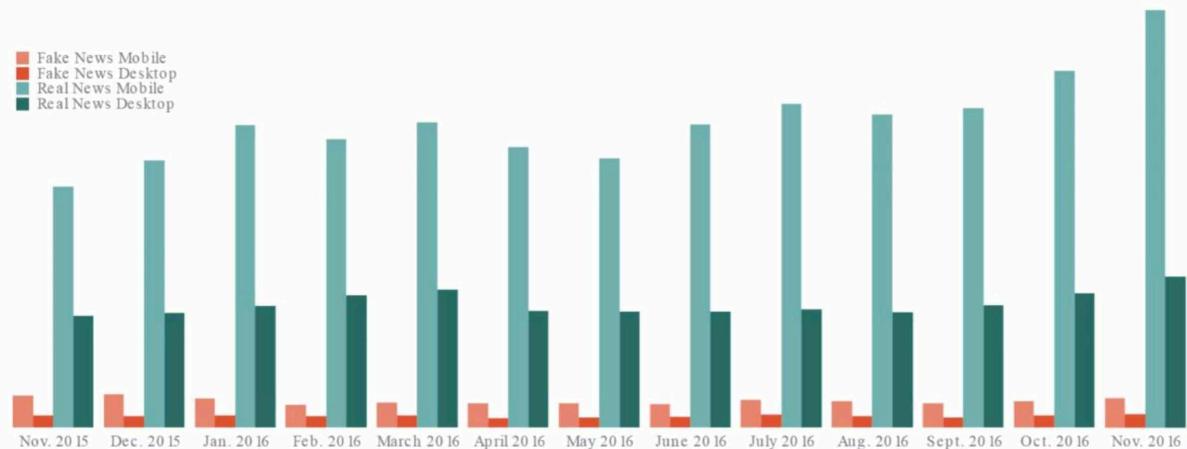
FT graphic: David Blood Source: Oxford Internet Institute

FT

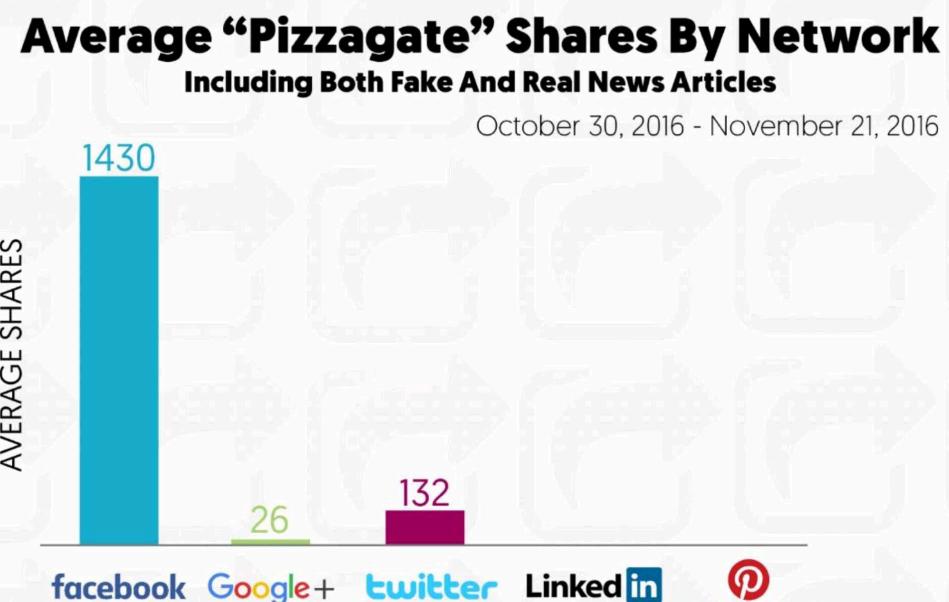
# “Fake news” (orange) audience is 10% of real news (green)

Columbia Journalism Review.

The real news audience even spiked in October and November 2016, while the fake news audience stayed the same.

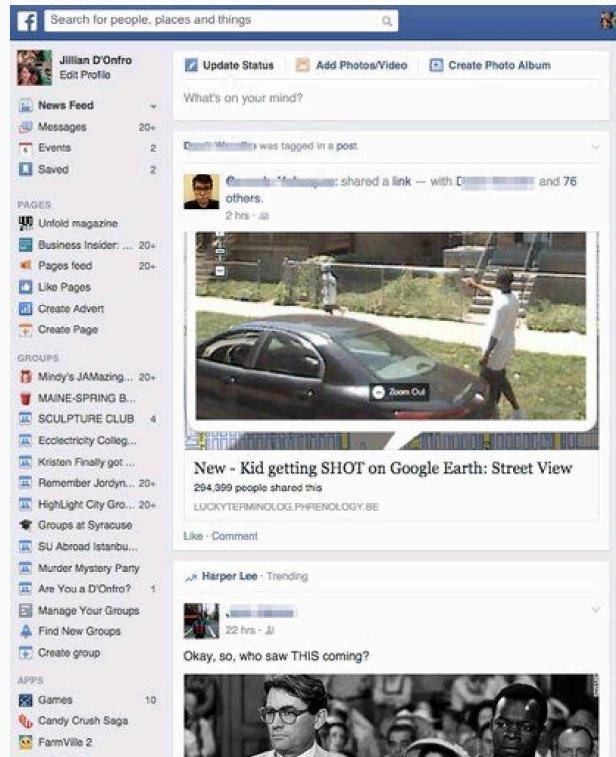


# For „fake news“ main source is Facebook



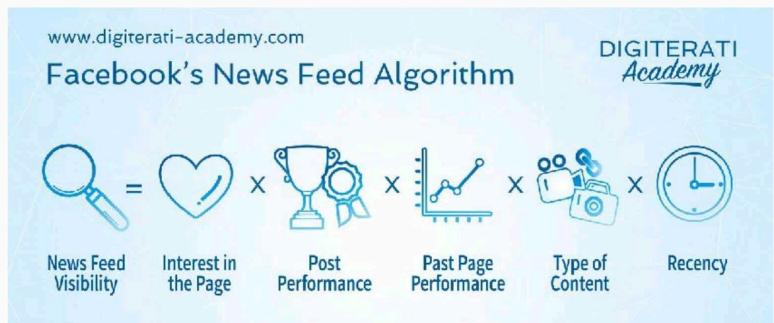
# On Facebook users see 10% of what they „subscribe” to

- Facebook decides which 10%
- Facebook chose to remove news from feed
  - hurting 10% of real news traffic
  - hurting 40% of fake news traffic
- Both unhappy!

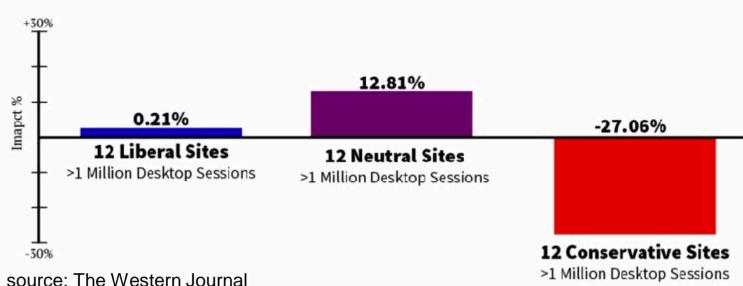


# How they choose the 10% is key issue for stopping “fake news”!

- Facebook's goal: „keep audience at site and engaged”
- Algorithm is opaque
- Algorithm is changing
- Algorithm is unfair?

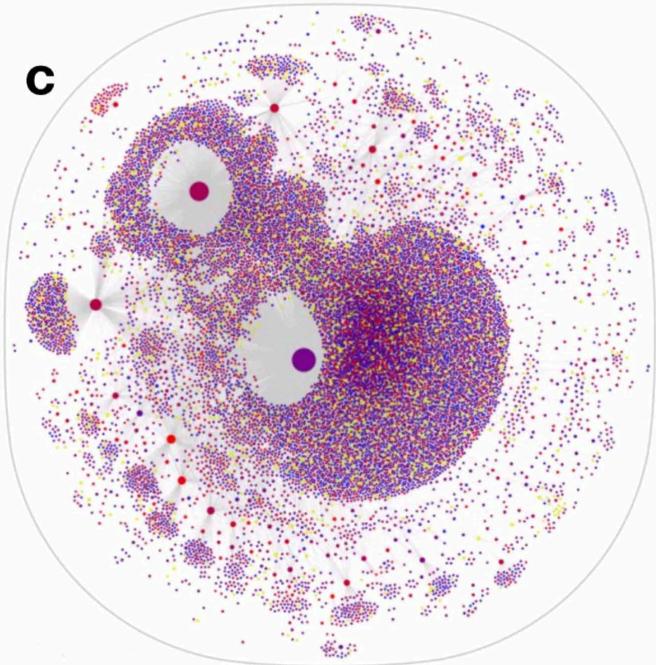


Average Facebook Algorithm Impact



# Bots spread fake news on Twitter. And real news.

- “We find evidence that social bots play a disproportionate role in spreading and repeating misinformation.” ([source](#)).
- “Contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it” ([source](#))



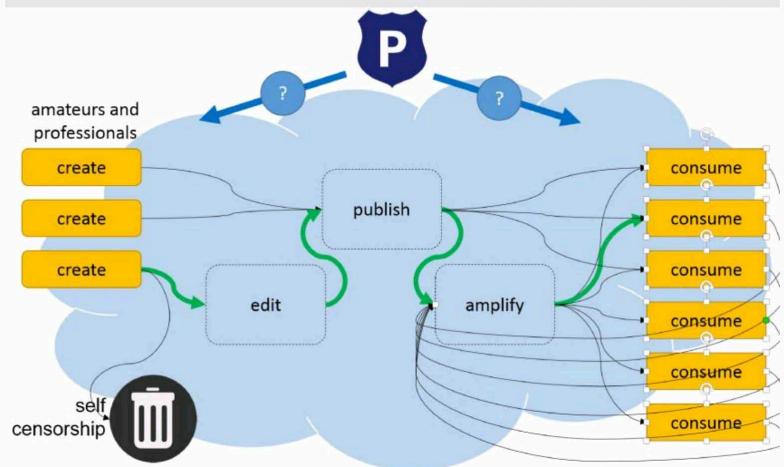
Propagation of fake news: Nodes and links represent Twitter accounts and retweets of the claim, respectively. Node size indicates account influence, measured by the number of times an account is retweeted. Node color represents bot score, from blue (likely human) to red (likely bot).

Can we stop  
“fake news”?

# Yes we can!

- with author
- with publisher
- with platform
- with amplifier
- with reader

Internet-age news process



# Stopping authors to publish “fake news”

- demonetize
  - deny advertising at fake news sites
- threaten prison
  - e.g. Polish anti-holocaust law
- lock them up
  - e.g. Turkey

The Rubin Report (talk show) · Dave Rubin · +3

**What was YouTube's reason for demonetizing the Rubin Report conversation between Dave Rubin, Jordan Peterson and Ben Shapiro?**

Answer Request Follow 22 Comment Downvote Facebook Twitter Print ...

1 Answer ≡ FINANCIAL TIMES myFT

 Jeff Franz-Lien, Strategic Communications, Vodafone Group PLC + Add to myFT  
Answered Feb 2

Vodafone to block its ads appearing on offensive websites

**What was YouTube's reason for demonetizing the Rubin Report conversation between Dave Rubin, Jordan Peterson and Ben Shapiro?**

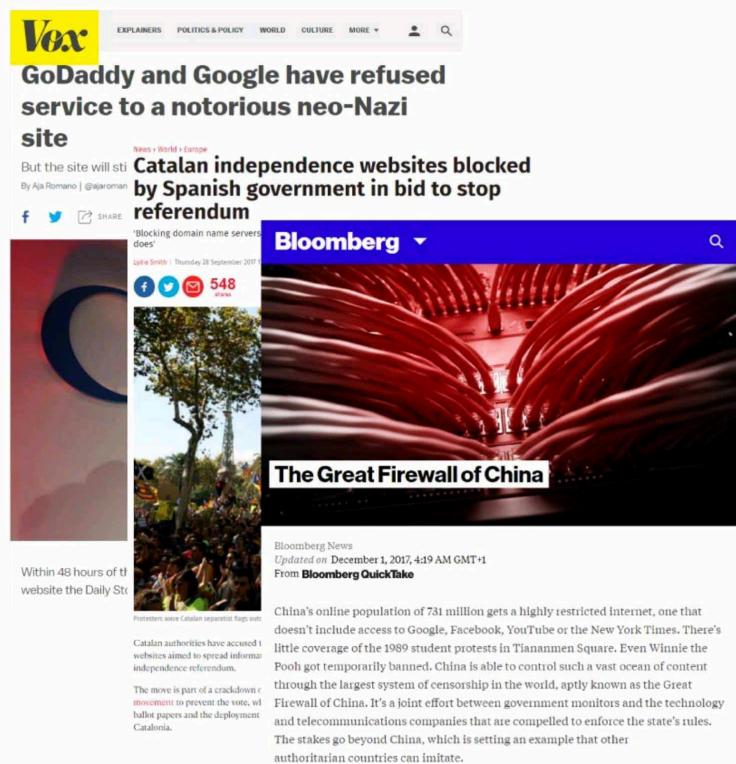
According to chatter on social media, Vodafone is one of several companies demonetizing “controversial” content on YouTube. Not just Rubin or other right-wing hosts, but others like Peterson were hit too, including Daily Beast host Ben Shapiro.

Rubin confirms this in the video above, noting that he’s been losing big bucks (Dave says \$100,000 per month) since his videos have been flagged for “offensive content. Who is responsible? It’s not YouTube and its adver-



# Stopping “fake news” when published

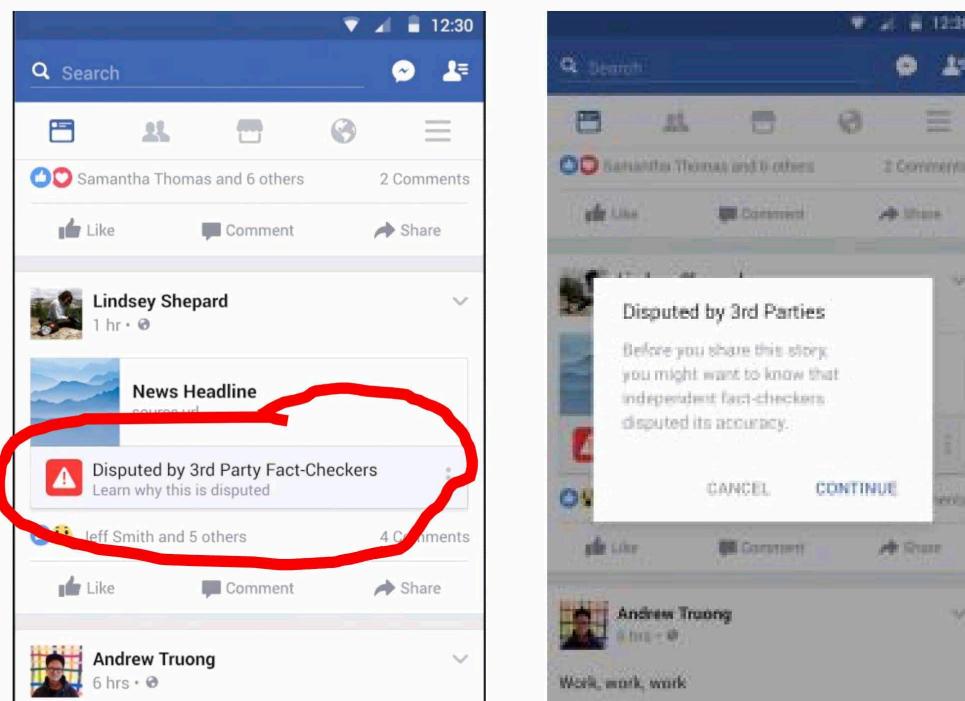
- deny hosting
  - deny listing in DNS directory
  - block internet traffic to “fake news” site



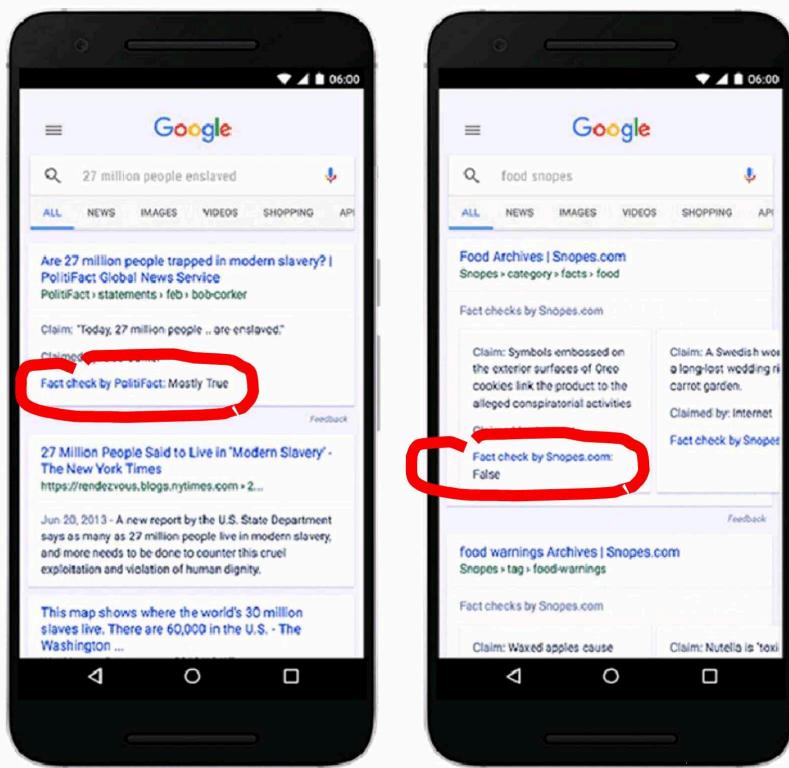
## Stopping fake news at platforms

- public platforms
  - Facebook, Google, Twitter, Baidu ...
  - label, augment, hide or remove “fake news”
- private messaging platforms
  - Viber, WeChat, Snapchat
  - listen to private conversations?!
- needs collaboration of platforms

# Label „fake news“ on Facebook



# Label “fake news” on Google



# Label “fake news” on Twitter browser

**Donald J. Trump** @realDonaldTrump

Are we talking about the same cyberattack where it was revealed that head of the DNC illegally gave Hillary the questions to the debate?

RETWEETS 20,819 LIKES 64,122

6:09 AM - 16 Dec 2016

This is incorrect or false.

Documents published after Clinton campaign chairman John Podesta's email was hacked indicate that then-CNN-contributor Donna Brazile sent some questions from a CNN Democratic primary town hall event to the campaign in advance. When the chair of the Democratic National Committee resigned after documents stolen from the DNC were leaked, Brazile stepped in as acting chair -- after the leaks above. Both hacks are believed by intelligence agencies to have originated in Russia. What Brazile did, by the way, is in no way illegal. [LEARN MORE]

Brought to you by The Washington Post.

# Hide “fake news” by downranking – pushing down on search results



**Facebook now flags and down-ranks fake news with help from outside fact checkers**

Josh Constine @joshconstine / Dec 15, 2016

Comment

Facebook adds warnings to fake news



Snopes, FactCheck.org, Politifact, ABC News, and AP will help Facebook make good on four of the six promises Mark Zuckerberg made about fighting fake news without it becoming “the arbiter of truth.” It will make fake news posts less visible, append warnings from fact checkers to fake news in the feed, make reporting hoaxes easier and disrupt the financial incentives of fake news spammers.



REUTERS

World Business Markets Politics TV

Myanmar Energy & Environment Brexit North Korea Earnings Futures

TECHNOLOGY NEWS NOVEMBER 22, 2017 / 9:15 PM / 4 MONTHS AGO

**Russia says Google down-ranking Sputnik, RT would be censorship**

Reuters Staff

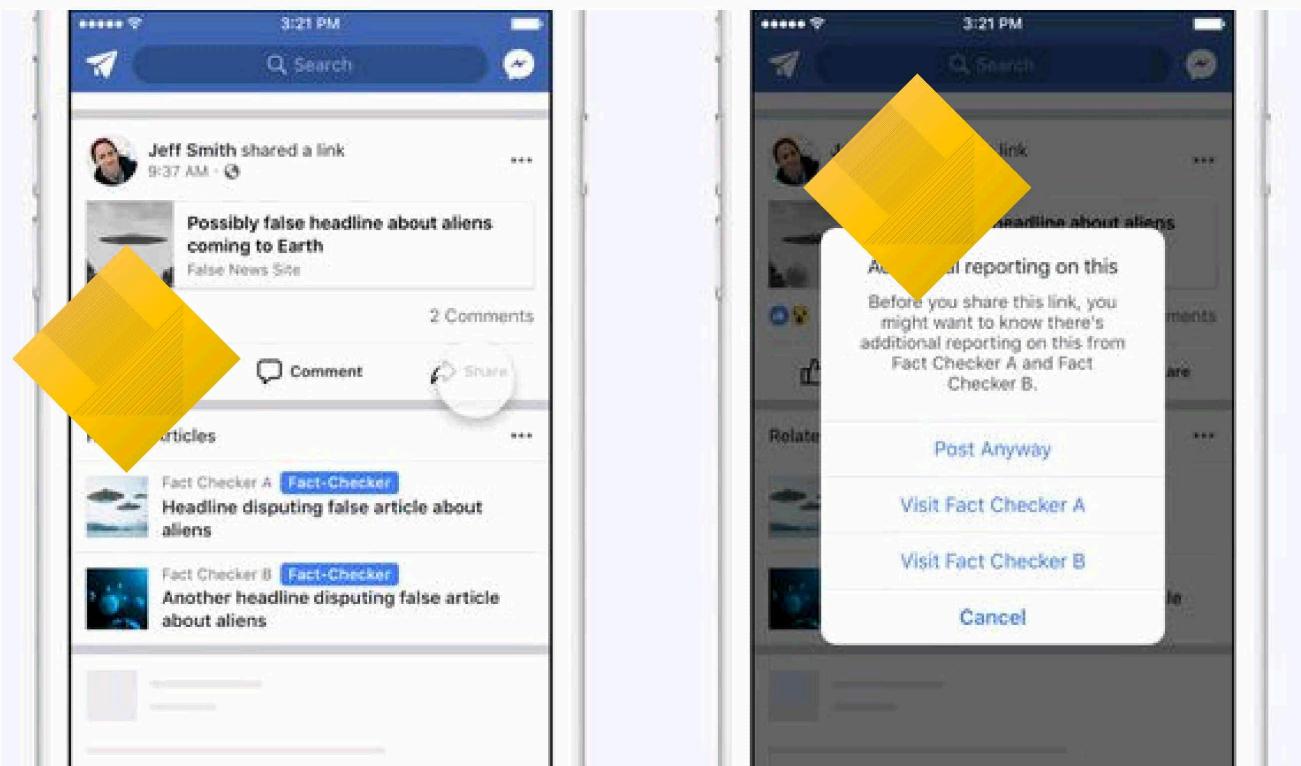
1 MIN READ



MOSCOW (Reuters) - Russia's foreign ministry said on Thursday that moves by Alphabet Inc's (GOOGL.O) Google to place articles from Russian news outlets Sputnik and Russia Today lower in search results would amount to censorship.



# Augment “fake news” with links to real news, warn before sharing



# Remove “fake news” - deplatform



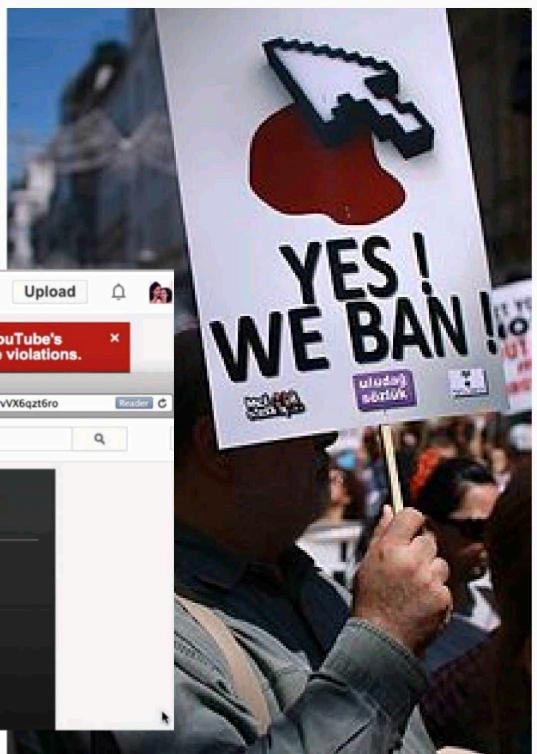
Science and nature  
Richard Dawkins event cancelled over his 'abusive speech against Islam'

Berkeley's KPFA after learning of author contests



A screenshot of a YouTube video player window. A red banner at the top reads: "This account has been suspended due to multiple or severe violations of YouTube's policy against spam, gaming, misleading content, or other Terms of Service violations." Below the banner, the video player displays a message: "This video is no longer available because the YouTube account associated with this video has been terminated." A red sad face icon is centered on the screen.

▲ "Why is it fine to criticise Christianity but not Islam? ... Richard Dawkins. Photograph: Murdo MacLeod for the Guardian



# Stop bots!



The image shows a screenshot of the Twitter 'Verified Account' beta testing page. At the top, the Twitter logo is visible. Below it, a blue header bar contains the text 'Verified Account' with a small 'BETA' tag next to it, and a blue checkmark icon. The main content area has a light blue background. It starts with the question 'What is a verified account?'. Below this, a paragraph explains the purpose of the feature: 'To prevent identity confusion, Twitter is experimenting (beta testing) with a 'Verified Account' feature. We're working to establish authenticity with people who deal with impersonation or identity confusion on a regular basis. Accounts with a  are the real thing!'.



# Stopping “fake news” in private chat, email – the next frontier

- What's Up
- Facebook Chat
- Snapchat
- Gmail
- Signal
- Telegram
- Viber
  
- can be done, but!?



Censorship in China  
How China's highly censored WeChat and Weibo fight fake news ... and other controversial content

As debate rages over Facebook's role in spreading false information on US elections, we look at how Chinese social media sites control such content shared online

PUBLISHED : Friday, 16 December, 2016, 4:01pm  
UPDATED : Monday, 12 June, 2017, 11:15pm

COMMENTS: 7



Key element: how to identify “fake news”?

## Fake news can be identified by humans, machines or combination

- Human
  - human work by fact checkers
  - fact-checking industry appearing, looking for business model
- Human-machine teamwork
  - machine identifies potentially fake stories
  - double-checked by human
- Machine only
  - artificial intelligence

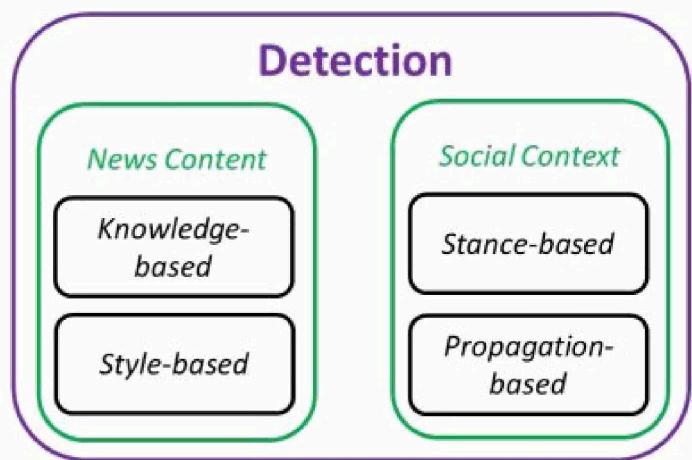
# Two strategies for automatic identification of “fake news”

- From content

- facts
- style

- From social

- stance
- propagation



Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu.  
"Fake news detection on social media: A data mining perspective." *ACM SIGKDD Explorations Newsletter* 19, no. 1 (2017): 22-36.

## Content analysis is hard because ...

- ... needs to understand content
  - few news pieces are blatantly false
- style?
  - authors will adapt

**POLITICAL LANGUAGE**  
is designed to make lies sound truthful and murder respectable, and to give an appearance of solidity to pure wind.

George Orwell

# Artificial Intelligence can help. But “Computer did it” is a fallacy

[SUBSCRIBE](#)

SCIENTIFIC  
AMERICAN

English ▾ Cart  Sign In | Register 

THE SCIENCES MIND HEALTH TECH SUSTAINABILITY EDUCATION VIDEO PODCASTS BLOGS STORE

COMPUTING

## Could AI Be the Future of Fake News and Product Reviews?

An experiment using a computer algorithm to create deceptive Yelp reviews was disturbingly successful, and could point to bigger problems as artificial intelligence matures

By Lacy Greenlee on October 30, 2017

SHARE

 SHARE 423

 TWEET

 COMMENT

 EMAIL

TO THE BUSINESS OF IT BY JODA  
**HUMANS CAN'T EXPECT AI TO JUST FIGHT FAKE NEWS FOR THEM**



 YOSHITOSU NEMOTO/GETTY IMAGES



Most Popular



 MORE STORIES

FAKE NEWS DETECTOR AI

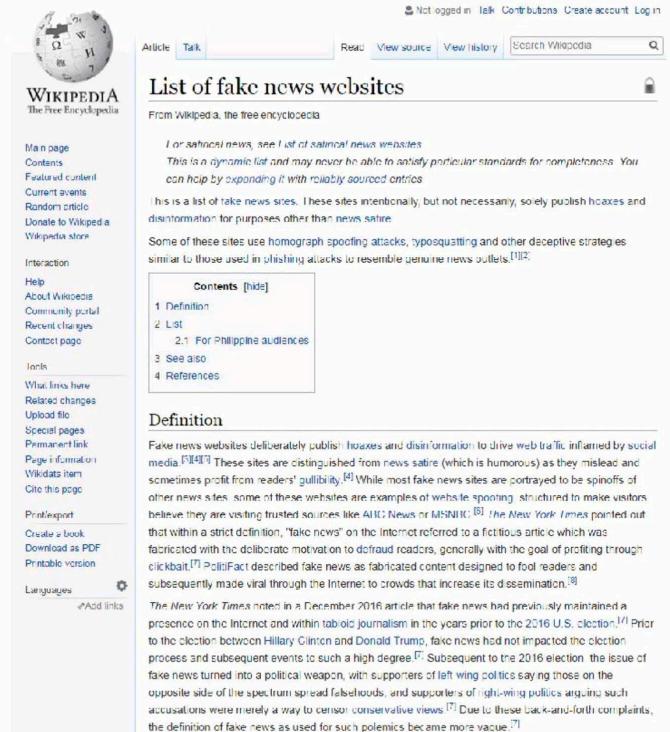
Ex: nytimes.com

Detect fake news sites using the power of artificial intelligence!

We analyze websites to see if they are similar to known fake news sites using a neural network. The same technology is used to power other artificial intelligence applications, like Siri and self-driving cars!

# New source is an easy identifier of “fake news”

- what site is the origin of news
- shortlist of “fake news” sites



The screenshot shows a Wikipedia article titled "List of fake news websites". The page header includes links for Article, Talk, Read, View source, View history, and Search Wikipedia. The main content area is titled "List of fake news websites" and includes a note about satire. Below this, there is a sidebar with a "Contents [hide]" section containing links for "Definition", "List", "For Philippine audiences", "See also", and "References". The main text discusses fake news websites that publish hoaxes and disinformation to drive web traffic. It notes that these sites are often spinoffs of real news outlets and may use homograph spoofing or typosquatting. The text also mentions the 2016 U.S. election and the role of fake news in spreading political misinformation.

Not logged in | Talk | Contributions | Create account | Log in

Article | Talk | Read | View source | View history | Search Wikipedia

## List of fake news websites

From Wikipedia, the free encyclopedia

For satirical news, see [List of satirical news websites](#)

This is a dynamic list and may never be able to satisfy particular standards for completeness. You can help by expanding it with reliably sourced entries.

Some of these sites use homograph spoofing attacks, typosquatting and other deceptive strategies similar to those used in phishing attacks to resemble genuine news outlets.<sup>[1][2]</sup>

**Contents [hide]**

- 1 Definition
- 2 List
  - 2.1 For Philippine audiences
- 3 See also
- 4 References

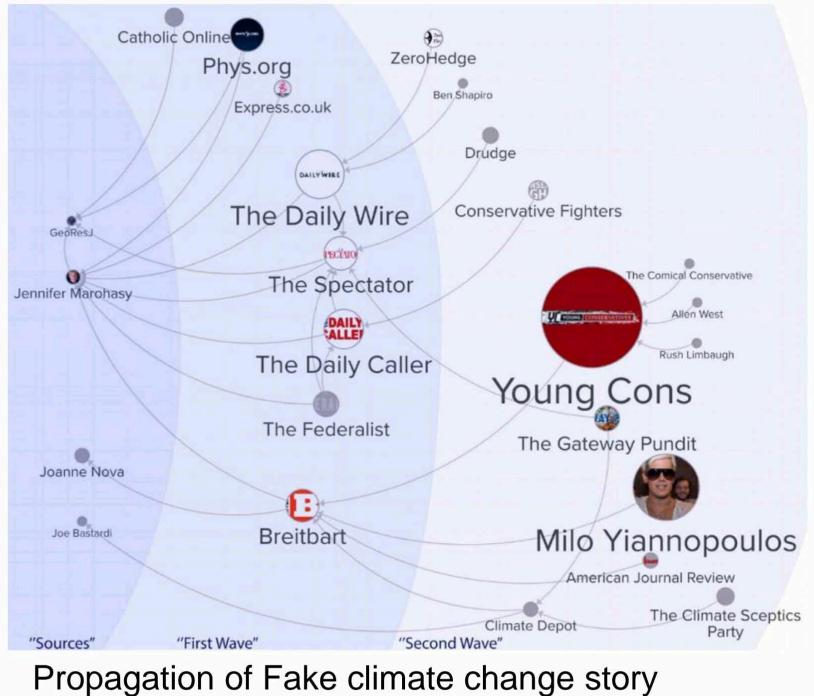
**Definition**

Fake news websites deliberately publish hoaxes and disinformation to drive web traffic, informed by social media.<sup>[3][4]</sup> These sites are distinguished from news satire (which is humorous) as they mislead and sometimes profit from readers' gullibility.<sup>[4]</sup> While most fake news sites are portrayed to be spinoffs of other news sites, some of these websites are examples of website spoofing, structured to make visitors believe they are visiting trusted sources like ABC News or MSNBC.<sup>[5]</sup> The New York Times pointed out that within a strict definition, "fake news" on the Internet referred to a fictitious article which was fabricated with the deliberate motivation to distract readers, generally with the goal of profiting through clickbait.<sup>[6]</sup> Politifact described fake news as fabricated content designed to fool readers and subsequently made viral through the Internet to crowds that increase its dissemination.<sup>[8]</sup>

The New York Times noted in a December 2016 article that fake news had previously maintained a presence on the Internet and within tabloid journalism in the years prior to the 2016 U.S. election.<sup>[1]</sup> Prior to the election between Hillary Clinton and Donald Trump, fake news had not impacted the election process and subsequent events to such a high degree.<sup>[7]</sup> Subsequent to the 2016 election, the issue of fake news turned into a political weapon, with supporters of left wing politics saying those on the opposite side of the spectrum spread falsehoods, and supporters of right-wing politics arguing such accusations were merely a way to censor conservative views.<sup>[7]</sup> Due to these back-and-forth complaints, the definition of fake news as used for such polemics became more vague.<sup>[7]</sup>

# Social context tells a lot about trustworthiness of news

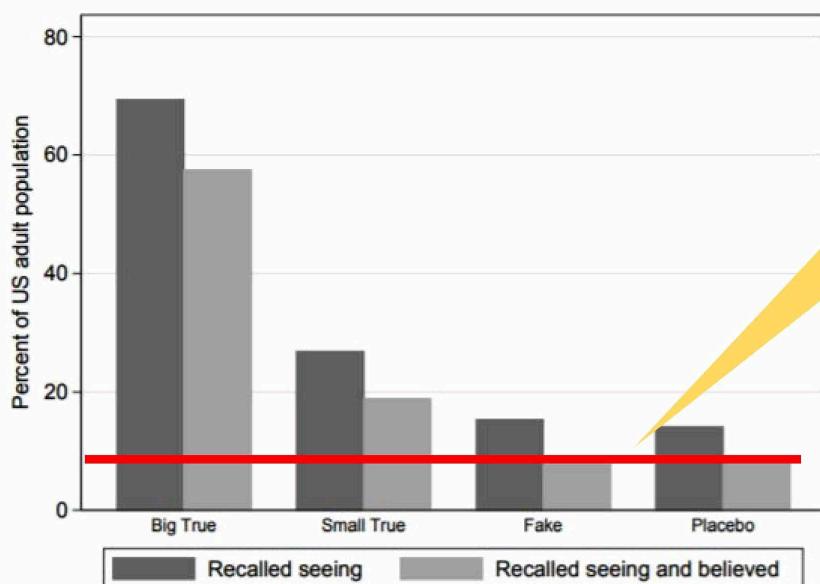
- “influencers”
- propagation patterns
- deployment of bots
- need for speed



**Stop at consumption**

# Media literate user is the last and best line of defense

Figure 5: Percent of US Adult Population that Recall Seeing or Believed Election News



people remember and believe “fake news” about as much as placebo news (non existent news)

“Available evidence suggests that for now the influence of fake news is limited”.

Hunt Allcott & Matthew Gentzkow, 2017. "[Social Media and Fake News in the 2016 Election](#)," Journal of Economic Perspectives, vol 31(2), pages 211-236.

# Conclusions

## Yes we can. But may we?

- gatekeepers are gone
  - only guerrilla war on fake news is possible
- technology to disturb circulation of fake news exists
  - most efficient if platforms cooperate
  - some approaches are clearly interfering with free speech
- users seem to be quite capable at identifying fake news
  - too early to blame election results on fake news

# The End

For more information:

[ziga.turk@gmail.com](mailto:ziga.turk@gmail.com)

@zigaTurkEU