

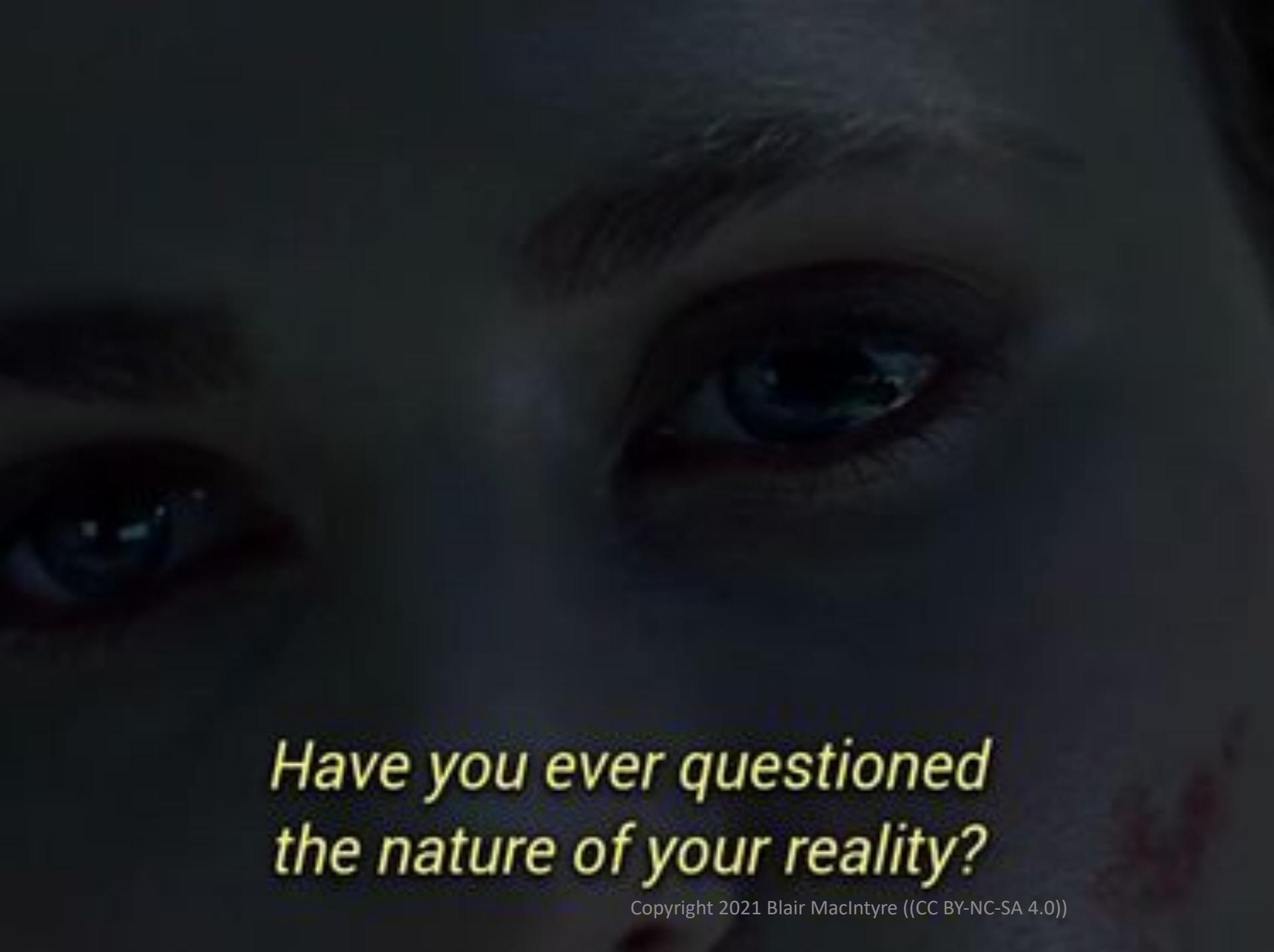
CS 4873: Computing, Society & Professionalism

Blair MacIntyre | Professor | School of Interactive Computing

Week 13: Fake News and Deep Fakes

April 14, 2021

Slides adapted from Sauvik Das, Munmun de Choudhury, and Amy Bruckman



*Have you ever questioned
the nature of your reality?*

Copyright 2021 Blair MacIntyre ((CC BY-NC-SA 4.0))

How do we develop our beliefs?

- Sensory-perception: “I’ll believe it when I see it with my own eyes!”
- Social proof: “A lot of other smart people seem to believe this...”
- Independent research: Little “r”. Investigative journalism; finding evidence, reading primary resources. Your term paper.
- Which one of these do you rely on most heavily when assessing the truth of something?

How has computing changed this?

- Today, we'll cover two ways that are particularly concerning:
 - Deep Fakes: attacks on sensory-perception
 - Fake News: attacks on social proof and research

Deep Fakes

Attacks on sensory-perception

What are deepfakes?

[Tom \(@deeptomcruise\)](#)
[TikTok | Watch Tom's](#)
[Newest TikTok Videos](#)



“Deepfakes leverage powerful techniques from machine learning and artificial intelligence to manipulate or generate visual and audio content with a high potential to deceive” ([Kietzmann et al. 2020](#)).

Example: Rowan Atkinson (Mr Bean) unexpectedly stars in a perfume commercial (original recorded with Charlize Theron).

View the original advert here:
<https://youtu.be/VqSl5mSjXJs>

View the deepfake here:
<https://youtu.be/tDAToEnJEY8>



[Deepfakes: Trick or Treat?](#)

Implications stretch beyond Mr. Bean

HOME > TECH

A viral video that appeared to show Obama calling Trump a 'dips---' shows a disturbing new trend called 'deepfakes'

Kaylee Fagan Apr 17, 2018, 4:48 PM



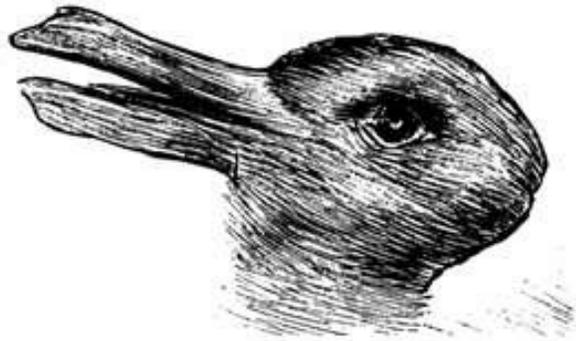
<https://www.youtube.com/watch?v=cQ54GDm1eL0>

- **BuzzFeed published a video that appeared to show former US President Barack Obama cursing and calling President Donald Trump names, but revealed the clip was actually fabricated using emerging video-editing technology.**

Deepfakes matter because:

Believability: If we see and hear something with our own eyes and ears, we believe it to exist or to be true, even if it is unlikely.

The brain's visual system can be targeted for misperception, in the same way optical illusions and bistable figures trick our brains.



Jastow rabbit-duck



Rubin vase-faces

Deepfakes matter because:

Accessibility: the technology of today and tomorrow, will allow all of us to create fakes that appear real, without a significant investment in training, data collection, hardware and software.

Zao, the popular Chinese app for mobile devices lets users place their faces into scenes from movies and TV shows, for free.



Deepfakes: Trick or Treat? | Kietzmann, J., Lee, L.W., McCarthy, I.P. and Kietzmann, T.C.

Advances driven by business interests

- DeepFakes aren't solely being developed by underground hacker groups for fake celebrity porn and disinformation campaigns
- They are being developed for business purposes as well
- Not just video or image

A typology of deepfakes and their applications

Type	Description	Current example	Business application
Photo deepfakes	<i>Face and body-swapping</i> Making changes to a face, replacing or blending the face (or body) with someone else's face (or body)	FaceApp's aging filter alters your photo to show how you might look decades from now (Kaushal, 2019).	Consumers can virtually try on cosmetics, eye glasses, hairstyles or clothes.
Audio deepfakes	<i>Voice-swapping</i> Changing a voice or imitating someone else's voice. <i>Text to Speech</i> Changing audio in a recording by typing in new text	Fraudsters used AI to mimic a CEO's voice and then tricked a manager into transferring \$243,000 (Supasorn Suwajanakorn, 2017). Users could make controversial Dr. Jordan Peterson (a famous professor of psychology and author) say anything they wanted, until his threat of legal action shut the site NotJordanPeterson down (Cole, 2019).	The voice of an audio book narration can sound younger, older, male, or female and with different dialects or accents to take on different characters. Misspoken words or a script change in a voiceover can be replaced without making a new recording.

[Deepfakes: Trick or Treat](#)

A typology of deepfakes and their applications

Type	Description	Current example	Business application
Video deepfakes	Face-swapping Replacing the face of someone in a video with the face of someone else.	Jim Carrey's face replaces Alison Brie's in "Late Night with Seth Meyers" interview.	Face-swapped video can be used to put the leading actor's face onto the body of a stunt double for more realistic-looking action shots in movies.
	Face-morphing A face changes into another face through a seamless transition	Former "Saturday Night Live" star Bill Hader imperceptibly morphs in and out of Arnold Schwarzenegger in the talk show Conan.	Video game players can insert their faces onto that of their favorite characters.
	Full body puppetry Transposing the movement from one person's body to that of another.	"Everybody dance now" shows how anyone can look like a professional dancer.	Business leaders and athletes can hide physical ailments during a video presentation. Deepfakes: Trick or Treat?

Deepfakes: Trick or Treat? | Kietzmann, J., Lee, L.W., McCarthy, I.P. and Kietzmann, T.C.

How do DeepFakes & Fake News complicate the nature of our democracy?

- DeepFakes can be detected today—but it's only a matter of time before doing so is non-trivial
- What are the implications of that for the future of our democracy?



Tech policy / AI Ethics

Deepfake Putin is here to warn Americans about their self-inflicted doom

AI-generated synthetic media is being used in a political ad campaign—not to disrupt the election, but to save it.

Already being used for disinformation

Public examples of offensive use of deepfakes



Barack Obama



Mauricio Macri



Nancy Pelosi

[Presentation 19080701 \(blackhat.com\)](https://blackhat.com/presentation/19080701)

Fake News

Media's Next Challenge: Overcoming the Threat of Fake News



Jim Rutenberg

MEDIATOR NOV. 6, 2016



Defining “fake news”

- Lazer et al. defined fake news outlets as those that have the trappings of legitimately produced news but “lack the news media’s editorial norms and processes for ensuring the accuracy and credibility of information.”
- The attribution of “fakeness” is thus not at the level of the story but at that of the publisher.

Fake news spreads faster & farther

- Published in Science – study by MIT Researchers
- Falsehoods & Truths from 2006 – 2017
- Fact checking organizations had to agree 95-98% of the time
- Truth spread up to 1,000 times; top 1% of falsehoods 1,000 – 100,000
- Politics represented 45,000 of the 126,000 cascades
- 1 account started 4,700 false rumors
- False rumors start with young, unverified accounts with a small following
- Falsehoods contain more novelty than truth
- Bots accelerate true & false news at the same rate; designed to increase anarchy in online social systems
- Early studies have shown labeling news as false might increase its spread

Guarino, Ben. Fake news spreads 'farther, faster, deeper' than truth, new study finds. *Washington Post* picked up by *The Seattle Times*, March 8, 2018.

<https://www.seattletimes.com/seattle-news/science/fake-news-spreads-farther-faster-deeper-than-truth-new-study-finds/>. Accessed March 9, 2018.

How Does Fake News Spread?

Online, especially with Social Media - Sharing of lies, half-truths, omissions, and out of context information.

On Twitter, fact checks of misinformation get about four times fewer shares than the original falsehood. (PolitiScope)

Example:

- Eric Tucker took photos of large groups of buses in Austin, TX
- Tweeted buses were related to anti-Trump protesters, an unverified statement
- Shared thousands of times on Twitter and Facebook
- Maheshwari, S. (2016, Nov. 12). *How fake news goes viral*. The New York Times. Retrieved from https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html?_r=0



Why Does Fake News Spread?

Money

- Example: “BREAKING: ‘Tens of thousands’ of fraudulent Clinton votes found in Ohio warehouse.”
- Story shared online by 6 million, earned thousands of dollars in Web advertising revenue
- Scott, S. (2017, Jan. 1). *From headline to photograph, a fake news masterpiece*. NYT.com Retrieved from <https://www.nytimes.com/2017/01/18/us/fake-news-hillary-clinton-cameron-harris.html>



Opinion

- Example: Pizzagate
- Claimed that John Podesta's leaked emails contained hidden messages referring to human trafficking connected to multiple U.S. restaurants and members of the Democratic Party
- Man shows up at pizza restaurant with weapons to help save the children
- Robb, A. (2017, Nov. 16). *Anatomy of a fake news scandal*. RollingStone.com. Retrieved from <https://www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/>



Why Does Fake News Spread?

Emotion

- Example: GoFundMe scam
- Homeless military veteran's random act of kindness and a New Jersey couple intent on helping him get back on his feet during the holidays
- Inspired people to donate more than \$400,000 in an online fundraiser that went viral
- Campaign was found to be a lie
- The three were each charged with second-degree conspiracy and theft by deception
- Stableford, D. (2018, Nov. 15). *New Jersey couple and homeless man whose feel-good story went viral charged with GoFundMe scam*. Yahoo.com. Retrieved from <https://www.yahoo.com/news/new-jersey-couple-homeless-man-whose-feel-good-story-went-viral-charged-gofundme-scam-193432689.html>

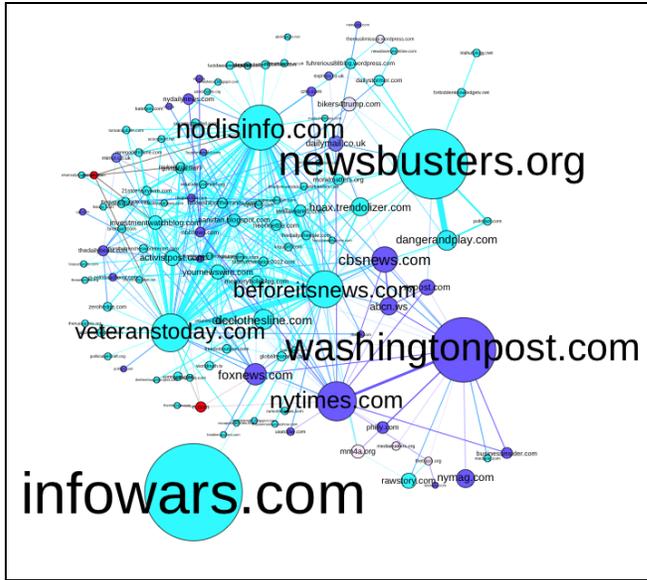


Examining Alternative Media Ecosystems

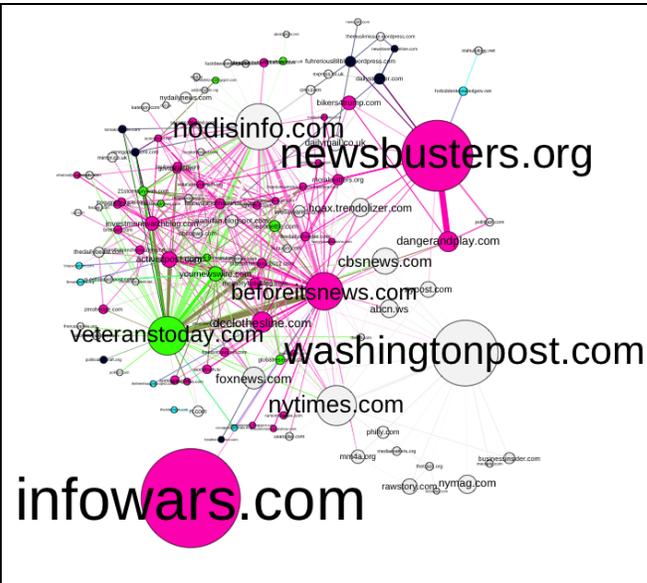
- The paper presents the first study of “fake news”
 - The context: in recent years, alternative media outlets have appropriated social media platforms for their perceived economic and political reach and for hosting inaccurate or under-sourced content
- Goals:
 - Provide a systematic lens for exploring the production of a certain type of “fake news” — alternative narratives of man-made crisis events
 - Examine the production of alternate narratives (rumors, conspiracy theories) through Twitter and across the external websites that Twitter users reference as they engage in these narratives

Summary (2)

Purple = mainstream media; Aqua = alternative media; Red = government controlled media



Pink = U.S. Alt-Right; Green = Intl. Anti-Globalist; Black = White Nationalist/Anti-Semitic; White = other

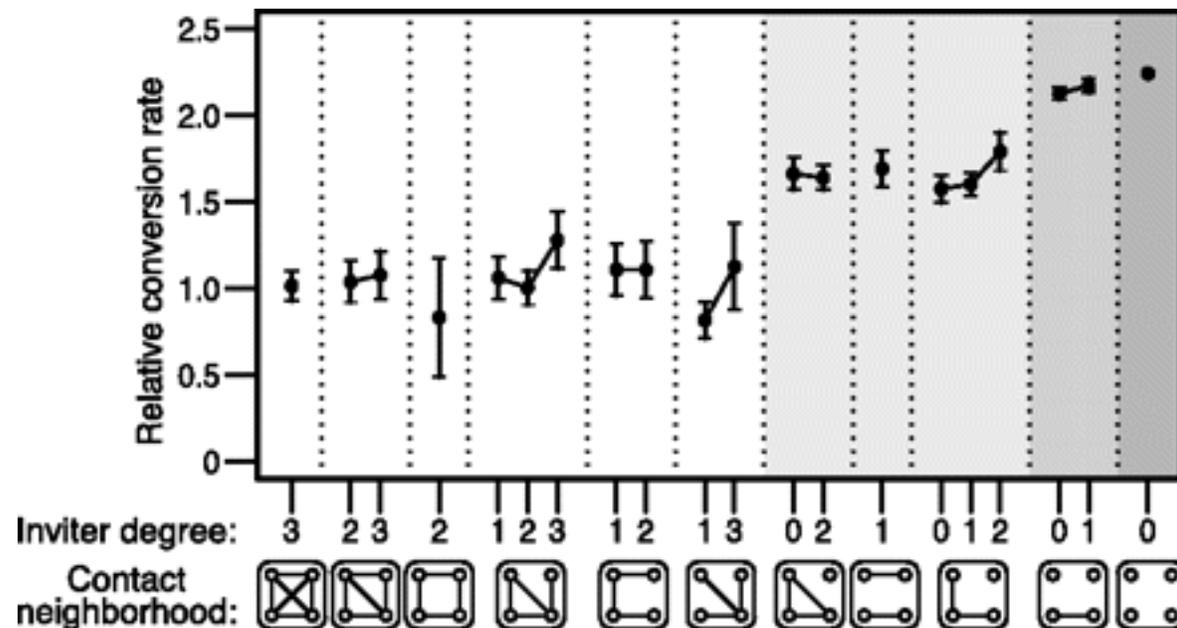


Leaning	Description
U.S. Alt Right	U.S. focused, anti-mainstream media, pro-Christian, anti-LGBT, anti-feminist, anti-globalist, climate change denying
U.S. Alt Left	U.S. focused, anti-mainstream media, anti-corporatist, critical of police, pro-prison reform, pro-BlackLivesMatter
International Anti-Globalist	Internationally focused, anti-globalist or anti-New World Order/Cabal, anti-corporatist, conspiracy-focused
White Nationalist and/or Anti-Semitic	primarily white-nationalist or anti-Semitic positions
Muslim Defense	primarily challenges mainstream narratives of terrorist attacks by Muslims
Russian Propaganda	primarily supports Russian interests, anti-globalist

- Strong political agendas underlying many of alternative narratives and the domains that hosted them
- More than half of the alternative media sites were coded as primarily motivated by a political agenda— with the conspiracy theories serving a secondary purpose of attracting an audience and reflecting or forwarding that agenda

Structural diversity in social contagion

- “Many” alternative media sources, but most are controlled by the same agencies and groups.
- “False perception of information diversity”
 - Why? Seeing multiple sources say the same things makes people more likely to believe in something



POLITICAL SCIENCE

Fake news on Twitter during the 2016 U.S. presidential election

Nir Grinberg^{1,2*}, Kenneth Joseph^{3*}, Lisa Friedland^{1*},
Briony Swire-Thompson^{1,2}, David Lazer^{1,2†}

The spread of fake news on social media became a public concern in the United States after the 2016 presidential election. We examined exposure to and sharing of fake news by registered voters on Twitter and found that engagement with fake news sources was extremely concentrated. Only 1% of individuals accounted for 80% of fake news source exposures, and 0.1% accounted for nearly 80% of fake news sources shared. Individuals most likely to engage with fake news sources were conservative leaning, older, and highly engaged with political news. A cluster of fake news sources shared overlapping audiences on the extreme right, but for people across the political spectrum, most political news exposure still came from mainstream media outlets.

In 1925, *Harper's Magazine* published an article titled "Fake news and the public," decrying the ways in which emerging technologies had made it increasingly difficult to separate rumor from fact (1). Nearly a century later, fake news has again found its way

social media have described its spread within platforms (5, 6) and highlighted the disproportionate role played by automated accounts (7), but they have been unable to make inferences about the experiences of ordinary citizens.

Outside of social media, fake news has been

Objective: sow discord

- Bots and trolls promote radicalizing / polarizing opinions, backed by fake or unverified sources, to undermine people's trust in information online.
- Inspired nihilism, learned helplessness
- Combined with confirmation bias, allows people to believe and cherry pick their sources and discredit any contrary narratives
- Traditional, mainstream media not immune
 - But barrier to entry much higher (e.g., the NYT has developed a reputation over decades)
 - Regulations and norms in industry to verify facts

Deep Fakes X Fake News

- Fake News attacks social proof and undermines our trust in journalism
- Deep Fakes attacks sensory perception
- Fake News spread faster than real news — fewer barriers to entry, fewer regulations and norms, collusion networks that bootstrap “virality”

The scope is global

THE TECH & DESIGN ISSUE

The Agency

From a nondescript office building in St. Petersburg, Russia, an army of well-paid “trolls” has tried to wreak havoc all around the Internet — and in real-life American communities.

Copyright 2021 Blair MacIntyre ((CC BY-NC-SA 4.0))

Winston Churchill



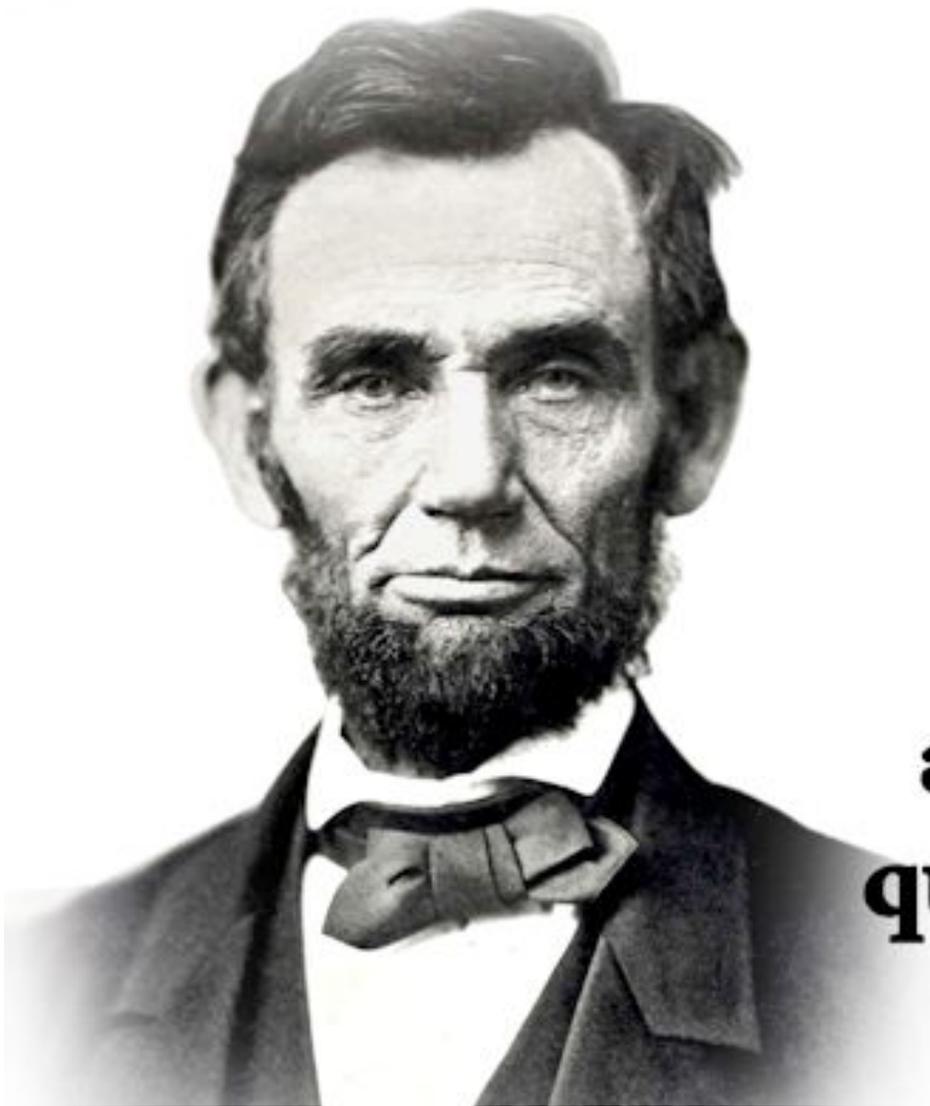
**“A Lie Can Travel Halfway
Around the World While the
Truth Is Putting On Its Shoes”**

Winston Churchill



**“A Lie Can Travel Halfway
Around the World While the
Truth Is Putting On Its Shoes”**

Not a real quote



**"Don't believe
everything you
read on the
internet just
because there's
a picture with a
quote next to it."**

-Abraham Lincoln