Disinformation Annotated Bibliography

By Gabrielle Lim

munk school
OF GLOBAL AFFAIRS & PUBLIC POLICY

UNIVERSITY OF TORONTO

THECITIZENLAB

Copyright

Electronic version first published in 2019 by the Citizen Lab. This work can be accessed through https://citizenlab.ca/2019/05/burned-after-reading-endless-mayflys-ephemeral-disinformation-campaign.

Document Version: 1.0

Suggested Citation

Gabrielle Lim. "Disinformation Annotated Bibliography." Citizen Lab, University of Toronto, May 2019.

Acknowledgements

Special thanks to Ron Deibert, John Scott-Railton, and Adam Senft. The design of this document is by Mari Zhou.

About the Citizen Lab, Munk School of Global Affairs & Public Policy, University of Toronto

The Citizen Lab is an interdisciplinary laboratory based at the Munk School of Global Affairs & Public Policy, University of Toronto, focusing on research, development, and high-level strategic policy and legal engagement at the intersection of information and communication technologies, human rights, and global security.

We use a "mixed methods" approach to research that combines methods from political science, law, computer science, and area studies. Our research includes investigating digital espionage against civil society, documenting Internet filtering and other technologies and practices that impact freedom of expression online, analyzing privacy, security, and information controls of popular applications, and examining transparency and accountability mechanisms relevant to the relationship between corporations and state agencies regarding personal data and other surveillance activities.

Contents

Introduction

This document serves as a reading list and primer on digital disinformation. While the proliferation of literature on the subject is a positive reaction to an otherwise vague yet troubling threat, it can be difficult to grasp how much has been accomplished and what questions remain unanswered. This document will therefore give readers a foundational understanding of the immense amount of work that has been done in the last few years on digital disinformation and where future research may be heading.

The sources are divided into nine categories of interest and include articles and reports from academic journals, research institutes, non-profit organizations, and news media, reflecting the multidisciplinary and sociotechnical nature of the subject. Although many of the sources can fit into more than one category, having a classification framework is useful for conceptualizing the kinds of research being done and provides direction for those new to the literature.

And finally, like a well-run state-sponsored troll farm, the scholarship of digital disinformation continues to produce new content every day. We would be remiss if we did not stress that this document is only a snapshot of a particular moment in this expanding field. As such, we've included a list of additional resources that are regularly updated with research and news on disinformation and media manipulation more broadly.

Note on definitions:

The study of disinformation covers a wide range of disciplines, geographies, and sociotechnical mechanisms and effects. Because of this, multiple definitions are often used to describe similar things and definitions are not always the same. Where possible, we choose to use the word "disinformation," which is most commonly understood as false information that is knowingly disseminated with malicious intent. Many of the sources below, however, use the term "fake news" instead, which is also used to describe verifiably false content.

Readers will notice that other terms for false and problematic information are also used, such as "propaganda" or "media manipulation." In these cases, please note that this is because the authors of the source have chosen to use those specific terms and their definitions. Instead of reinterpreting their choice of words, we have chosen to use them when describing their work.

For a more in-depth analysis of the words and definitions used in this field of study, please refer to Lexicon of Lies by Caroline Jack.

General

Recent years have seen an increase of reports from research organizations, think tanks, and government agencies on the role of disinformation and media manipulation more broadly. From Rand's *Truth Decay* report to Wardle and Derakhshan's article "Information Disorder," the following select articles and reports tend to be multidisciplinary and offer a broad view of digital disinformation.

The sources in this section are generally pessimistic about our ability to mitigate against the harms, noting that combatting the spread of disinformation is akin to playing "whack-a-mole" due to the relatively low barriers to conducting an information operation. While certainly not a new phenomenon, the authors below also find disinformation has been exacerbated by the Internet, specifically our media consumption patterns and the rise of social media platforms. Furthermore, disinformation is a global issue found in both established democracies and authoritarian or illiberal regimes.

Sources of note:

Primer on definitions: Lexicon of Lies
Global comparisons of social media manipulation: Challenging Truth and Trust
For communications professionals: Countering Information Influence Activities
On what the experts think: The Future of Truth and Misinformation Online
On state-use of digital disinformation: Digital Threats To Democratic Elections: How Foreign Actors Use Digital Techniques

The Future of Truth and Misinformation Online

Janna Anderson and Lee Rainie

Anderson, Janna and Lee Rainie. *The Future of Truth and Misinformation Online*. Pew Research Center, 2017. http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online.

Crux

This report summarizes the results from a survey of 1,116 experts, identified as scholars, practitioners, technologists, and strategic thinkers.

The respondents were initially asked whether they thought the information ecosystem over the next ten years would improve and whether trusted methods would emerge to block false and misleading narratives. The respondents were then asked follow-up questions based on their answer. There is an almost even split with 51% of respondents saying that "The information environment will NOT improve," and the remaining 49% saying that it will improve.

Those who think that the information ecosystem will not improve say that humans tend to shape technology to advance their "not-fully-noble purposes" and that there will always be bad actors that will foil any technological efforts to remedy the problem. Of those who are more optimistic, they believe that technological fixes can be implemented to "bring out the better angels guiding human nature."

Highlights

- A separate Pew Research Center study conducted after the 2016 US election found 64% of adults

believe fake news stories "cause a great deal of confusion" and 23% admitted they had shared fake political stories themselves, either by mistake or intentionally.

- Respondents in the pessimistic camp cited two reasons for why things WILL NOT improve: 1) the fake news ecosystem preys on some of our deepest human instincts and 2) our brains are not wired to contend with the pace of technological change.
- Respondents in the optimistic camp cited two reasons for why things WILL improve: 1) technology can fix these problems and 2) it is also human nature to come together and fix problems.
- Across the board, there was agreement that the issue of misinformation requires significant attention "urging a bolstering of the public-serving press and an expansive, comprehensive, ongoing information literacy education effort for people of all ages."

Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation

Samantha Bradshaw and Philip N. Howard

Bradshaw, Samantha and Philip N. Howard. "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation." Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda, 2017. https://comprop.oii.ox.ac.uk/research/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/.

Crux

This report offers a high-level overview of 28 countries' state-sponsored and political efforts in manipulating public opinion over social media. Through their analysis, the authors are able to create an inventory of the messages, strategies, organizational forms, capacities, and budgets behind social media manipulation.

Highlights

- Across the 28 countries, every authoritarian regime has social media campaigns targeting their own populations, while only a few of them target foreign audiences.
- Conversely, almost every democracy in this sample has organized social media campaigns that target foreign audiences.
- Increasingly, manipulating public opinion via social media networks has been contracted to private communication firms.
- Not all comments posted on behalf of a government or political party are positive or negative. Some may be neutral and designed to obfuscate data. An example of this is "hashtag poisoning," which Saudi Arabia routinely engages in to disrupt criticism.

Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation

Samantha Bradshaw and Philip N. Howard

Bradshaw, Samantha and Philip N. Howard. "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation." Working Paper 2018.1. Oxford, UK: Project on Computational Propaganda. http://comprop.oii.ox.ac.uk/research/cybertroops2018/.

Crux

A follow-up to the authors' 2017 report on state-sponsored efforts in manipulating public opinion over social media, this report expands their research to include 20 more countries. The authors identify the organizational form and prevalence of social media manipulation, messaging and valence, strategies, and capacity for each of the 48 states. In doing so, they note five trends: 1) an increase of computational propaganda during elections; 2) an increase of government agencies tasked with countering disinformation; 3) growing evidence of disinformation campaigns occurring on chat applications; 4) social media manipulation tactics continuing to evolve in order to keep up with regulation and counter-measures; and 5) a growing digital influence industry.

Highlights

- The report examines media manipulation in 48

countries: Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Brazil, Cambodia, China, Colombia, Cuba, Czech Republic, Ecuador, Egypt, Germany, Hungary, India, Iran, Israel, Italy, Kenya, Kyrgyzstan, Malaysia, Mexico, Myanmar, Netherlands, Nigeria, North Korea, Pakistan, Philippines, Poland, Russia, Saudi Arabia, Serbia, South Africa, South Korea, Syria, Taiwan, Thailand, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States, Venezuela, Vietnam, and Zimbabwe.

- Of the 48 countries examined, 30 had evidence of political parties using computational propaganda during elections or referenda.
- Since 2016, over 30 countries have introduced legislation designed to combat online "fake news."
- The range of platforms on which digital disinformation is carried out has grown to include chat applications (e.g., LINE, SnapChat, Telegram, Tinder, WeChat, WhatsApp). Evidence of this was found in 12 of 48 countries examined.

Lexicon of Lies: Terms for Problematic Information

Caroline Jack

Jack, Caroline. *Lexicon of Lies: Terms for Problematic Information*. New York: Data & Society Research Institute, 2017. https://datasociety.net/output/lexicon-of-lies/.

Crux

This report examines the various terms and concepts that have been used to describe problematic information, such as "fake news," "disinformation," "misinformation," or "propaganda." It underscores the difficulty in discerning between the terms due to the overlapping nature of some of the meanings.

Highlights

- "Nation-branding" entails hiring public relations and advertising firms to promote a country. It can be characterized as public affairs, publicity, or even as a form of information operations or propaganda.
- "White propaganda" uses accurate, albeit carefully presented, information from accurately identified sources, whereas "black propaganda" relies on inaccurate or deceptive information. In "black propaganda," the source of the information is obscured or misrepresented. "Gray propaganda" uses both tactics.

Media Manipulation and Disinformation Online

Alice Marwick and Rebecca Lewis

Marwick, Alice and Rebecca Lewis. *Media Manipulation and Disinformation Online.* New York: Data & Society Research Institute, 2017. https://datasociety.net/output/media-manipulation-and-disinfo-online/.

Crux

This report covers a broad range of topics regarding the use and manipulation of the online media ecosystem in propagating ideas and setting agendas. Though the report does not focus specifically on disinformation and fake news, the authors note that the rise in sensationalist, hyper-partisan, clickbait content may lead to a further distrust of mainstream media, increased misinformation, and further radicalization. Much of their research focuses on Internet subcultures that are known as the "alt-right."

The report is divided into six main chapters followed by a conclusion and case studies:

1) Who is manipulating the media (e.g., trolls, hate groups, politicians, hyper-partisan news outlets)
2) Where do these actors operate (blogs, websites, forums, message boards, social media)
3) What is their motivation (money, status and attention, recruitment and radicalization)
4) What techniques or tactics are used (participatory culture, networks, memes, bots, strategic framing)
5) Why is the media vulnerable (lack of trust in mainstream media, decline in local news, short attention spans)
6) What are the outcomes (misinformation,

growing distrust in mainstream media, further radicalization)

The authors include four case studies: the White Student Union; Trump and the Star of David image; Hillary Clinton's health; and Pizzagate.

Highlights

- The authors highlight 4Chan-style trolling, which is characterized by four properties:
  - Use of deliberately offensive speech
  - Antipathy toward sensationalism in the mainstream media
  - Desire to create emotional impact in targets
  - Preservation of ambiguity
- Far-right actors frequently game Twitter's trending topics feature to amplify certain stories or messages.
- Due to the declining profitability of local news, most local news outlets have been bought and amalgamated into larger corporations, which prioritize generic content that can appeal to multiple audiences and short-term profits.

Countering Information Influence Activities

James Pamment, Howard Nothhaft, Henrik Agardh-Twetman, and Alicia Fjällhed

Pament, James, Howard Nothhaft, Henrik Agardh-Twetman, and Alicia Fjällhed. *Countering Information Influence Activities, Version 1.4.* Department of Strategic Communication, Lund University, 2018. https://www.msb.se/RibData/Filer/pdf/28697.pdf

Crux

This report, commissioned to support the Swedish Civil Contingencies Agency, offers an in-depth overview of influence operations as well as recommendations for countering such operations from a strategic communications perspective. The report covers various influence strategies and the types of techniques and tactics commonly employed. Their recommendations follow the "communicator's mandate," which is divided into three interconnected steps that form a cycle: preparation, action, and learning. The authors note, however, that there are limits to countering information operations and that more importantly, we should act cautiously in our attempts to mitigate or counter their effects.

Highlights

- The authors note that there are three types of information operations: *positive or constructive* strategies; *negative or disruptive* strategies; and oblique strategies. Positive strategies try to establish a coherent narrative, whereas the negative strategies attempt to prevent the emergence of a coherent narrative. Oblique strategies try to draw attention away from key issues.
- The authors note that narratives tend to fall within persistent grand narratives, or *meta-narratives*, which people are socialized into. These tend to be a-factual (i.e., religion) and give identity to their adherents.
- Understanding meta-narratives is important as a single news item or story may only represent a "fractal" of the larger narrative. This requires understanding the "chain of event-perspective."

A Short Guide to the History of 'Fake News' and Disinformation

Julie Posetti and Alice Matthews

Posetti, Julie and Alice Matthews. *A Short Guide To The History Of 'Fake News' And Disinformation.* International Centre for Journalists, 2018. https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module.

Crux

Aimed at journalists or those working in journalism education, this report covers a broad history of disinformation beginning in 44 BC when Octavian targeted Mark Antony with a smear campaign. It highlights how the development of technology has aided in the dissemination of fake news as well as the risks to freedom of expression posed by certain counter-measures.

Highlights

- "The invention of the Gutenberg printing press in 1493 dramatically amplified the dissemination of disinformation and misinformation, and it ultimately delivered the first-large scale news hoax – 'The Great Moon Hoax' of 1835."
- "The disinformation contained within news stories in 1917 is said to have caused the accurate reports of Nazi atrocities to be doubted when they first appeared."
- The report discusses the impact satirical shows like *The Daily Show* and *Colbert Report* have had in blurring the lines between real and fake coverage.
- In 2017, the QNA news agency was hacked and a false story was published containing falsified quotes attributed to Qatar's emir, Tamim bin Hamad al-Thani. These false quotes criticized US president Donald Trump and praised Iran as an Islamic power. Computational propaganda then used to fuel the hashtag "قطع العلاقات مع قطر" — "#Cut relations with Qatar," which did in fact happen shortly after. Quartz called it "the first major geopolitical crisis to have been sparked by a computer hack."

Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making

Claire Wardle and Hossein Derakhshan

Wardle, Claire and Hossein Derakhshan. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Council of Europe, 2017. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

Crux

This report offers a broad overview of disinformation and its current challenges in relation to contemporary social technology. The authors provide a conceptual framework for examining what they call the "information disorder" by identifying three types of information: mis-, dis-, and mal-information. In addition, they break down the "elements" of information disorder into "the agent, message, and interpreter" and emphasize the three different "phases": creation, (re)production, and distribution. The report also draws from the work of scholar James Carey, stressing the need to "understand the ritualistic function of communication." The authors argue that rather than thinking about communication as simply the transfer of information from one person to another, we must also acknowledge that communication represents our shared beliefs. The report ends with 34 recommendations targeted at a variety of stakeholders.

Highlights

- The authors define misinformation as false information that is shared without the intent to do harm, whereas disinformation is false information with the intent to harm. They define malinformation as genuine information but with the intent to do harm (e.g., revenge porn)
- The report notes the use of visual-based disinformation, which can often be more powerful than textual information.
- The authors discuss the threat of declining local news media and what will take its place.
- The authors note that fact checking is popular, but its effectiveness is debatable. In some cases, there is no overlap between those who consume the false information and those who consume the debunking.
- Part 4 of the report examines future trends and challenges, such as encrypted chat apps, artificial intelligence, and augmented reality.
- The report includes a list of European fact-checking initiatives.

Creation and Dissemination

The sources within this section deal primarily with the creation of disinformation and the means with which it is distributed. They cover a broad range of tools and tactics including the use of "troll farms," public relations companies, automated accounts (a.k.a. bots), deep fake technology, inauthentic personas, and artificial intelligence.

Broadly speaking, scholars tend to agree that the creation and dissemination of disinformation will continuously evolve in an attempt to sidestep any technological solutions aimed at curbing its distribution. This whack-a-mole interaction, along with advances in artificial intelligence and automation, will make it challenging for most humans to tell if what they're looking at or listening to is authentic.

This section is closely related to the Advertising and Marketing and Social Media sections below as the tools and tactics afforded by online marketing and social media platforms are widely used for content distribution and for monitoring and fine-tuning information operations.

Sources of note:

Ethnography of disinformation creators: Architects of Networked Disinformation
Role of journalism: Lies, Damn Lies, and Viral Content
On deep fake technology: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security
Comprehensive breakdown of tools and tactics: Digital Threats To Democratic Elections: How Foreign Actors Use Digital Techniques

The Agency

Adrian Chen

Chen, Adrian. "The Agency." *The New York Times*, June 2, 2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

Crux

Written by an investigative journalist, this piece from *The New York Times* delves into one of Russia's largest "troll farms" in St. Petersburg, known as the Internet Research Agency (IRA). Chen reveals what it's like to work there, how much one earns, the types of people who work there, and the people funding and running the place. The article also traces the evolution of trolling in Russia, noting that pro-Kremlin trolling first really took off after the anti-government protests of 2011. Chen closes the article by summarizing his own personal experience of being attacked by a pro-Kremlin disinformation campaign. In 2018, the IRA along with 13 Russian nationals and two other Russian entities were indicted by a federal grand jury for alleged illegal interference in the 2016 presidential elections.

Highlights

- At the time of writing, the Internet Research Agency was being sued by an ex-employee (and mole) for violating labour rights laws. The employee had enlisted the help of well-known human rights lawyer, Ivan Pavlov, who has spent years fighting for transparency in Russia.
- "Several Russian media outlets have claimed that the agency is funded by Evgeny Prigozhin, an oligarch restauranteur called "the Kremlin's

chef" in the independent press for his lucrative government contracts and his close relationship with Putin."

- "The point is to spoil it, to create the atmosphere of hate, to make it so stinky that normal people won't want to touch it," said Leonid Volkov, a liberal politician and campaign manager to anti-corruption crusader Alexei Navalny. "You have to remember the Internet population of Russia is just over 50 percent. The rest are yet to join, and when they join it's very important what is their first impression."

Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security

Bobby Chesney and Danielle Citron

Chesney, Robert and Danielle Citron. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107 (2019, Forthcoming). https://ssrn.com/abstract=3213954.

Crux

This forthcoming article offers an overview of how "deep-fake technology" will worsen the current information and media ecosystem, leading to a state of "truth decay." The authors discuss the technologies available now and the ones that are likely to come in the near future, the costs and benefits of such technologies, and an analysis of the technical, legal, and market solutions available to curb the creation and dissemination of deep fakes.

Highlights

- There are benefits to deep-fake technology when used in education, art, and automation.
- The harms, however, include exploitation and sabotage of individuals, distortion of democratic discourse, election manipulation, exacerbating social tensions, jeopardizing national security, and failure to prove something is real.
- Technological solutions are currently not scaleable or reliable enough to detect real from fake.
- Some American agencies that may have interest and jurisdiction over falsified media include the Federal Trade Commission (FTC), the Federal Communications Commission (FCC), and the Federal Election Commission (FEC).
- Lawsuits against platforms are unlikely to be successful due to Section 230 of the Communications Decency Act.
- There may be some avenues of recourse for victims of deep fakes through criminal law, such as the federal cyberstalking law (18 U.S.C. 2261A) or state anti-impersonation laws.

The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public

Lion Gu, Vladimir Kropotov, and Fyodor Yarochkin

Gu, Lion, Vladimir Kropotov and Fyodor Yarochkin. *The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public.* Trend Micro, 2017. https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf.

Crux

This paper from cybersecurity company Trend Micro explores the tactics and technologies used to propagate online propaganda and disinformation. In doing so, they demonstrate several techniques using social media data that allow one to trace campaigns back to their original perpetrators. The article also includes a few case studies that explore how much it would cost to create an online "celebrity"; take down a journalist; or instigate a street protest.

The authors define the core components of online disinformation campaigns as the "three legs of the fake news triangle," which are the social networks, motivation, and tools and services. Without one of these three, spreading fake news becomes virtually impossible.

Highlights

- In addition to creating and disseminating false news, there are also services in China that offer to take down false content. An example being 118t

Negative News (大良造负面信息理).

- Russian methods of manipulating the media including crowdsourcing by incentivizing users with points for completing certain tasks (i.e., liking a post or following a profile). These points can then be resold or used for self promotion. An example of this type of crowdsourcing platform is VTope, which supports VKontakte (VK), Ok.com, YouTube, Twitter, Ask.fm, Facebook, and Instagram.
- Russian company like4u takes crowdsourcing up another notch by touting its capability to control the speed of promotion and set up time limits for tasks, which helps avoid bans from media platforms.

Future Elections May Be Swayed by Intelligent, Weaponized Chatbots

Lisa-Marie Neudert

Neudert, Lisa-Marie. "Future Elections May Be Swayed by Intelligent, Weaponized Chatbots." *MIT Technology Review*, August 22, 2018. https://www.technologyreview.com/s/611832/future-elections-may-be-swayed-by-intelligent-weaponized-chatbots/.

Crux

This article warns of the future threats posed by advanced chatbots, which could target individuals, convince people to hand over personal information, or deliver customized propaganda.

Highlights

- Because Google and Amazon routinely make their research open source, anyone, including bad actors, have access to it.
- "Since 2010 political parties and governments have spent more than half a billion dollars on social-media manipulation, turning it into a highly professionalized and well-funded sector."

Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines

Jonathan Corpus Ong and Jason Vincent A. Cabañes

Ong, Jonathan Corpus and Jason Vincent A. Cabañes. *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines.* The Newton Tech 4 Dev Network, 2018. http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf.

Crux

This report maps out the disinformation ecosystem within the Philippines and includes a political analysis and ethnographic research through in-depth interviews with 20 individuals working as "architects of networked disinformation." The authors supplement these interviews with participant observations of several digital campaigns on social media and insider access to fake accounts shared with them by their informants.

Highlights

- The authors note that efforts to blacklist fake news websites, expose fake accounts, or vilify divisive digital influencers fail to address the institutions and systems that professionalize and incentivize disinformation production.
- Moral displacement by workers within the disinformation ecosystem is easier to achieve given the casual, short-term nature of the work.
- The top tier of networked disinformation campaigns are advertising and public relations executives who act as high-level political operators.
- With regards to the use of bots, one chief architect remarked, "Bots are like the white walkers in Game of Thrones. They're stupid and obvious and easily killed. They can't inspire engagement."
- There is a stark difference in output between paid digital workers and "real fans." One influencer noticed fans produced up to 30 to 50 posts a day, whereas non-supporters only did five posts.

- Politicians often hire in-house "community-level fake-account operators" who post content from generic greetings to political messages within Facebook community groups.
- High-level strategists lure digital influencers through symbolic and material means, promising them expensive gadgets and organizing photo-ops with their own celebrity clients to further enhance the influencers' fame.
- A disinformation campaign takes on three stages: 1) "Campaign plan design," where clients dictate their objectives; 2) "Click army mobilization," which identifies which social media accounts to use; and 3) "Creative executions," which takes on positive branding techniques, diversionary tactics, negative black ops campaigns against opponents, and trending and signal scrambling (i.e., gaming Twitter hashtags and trends).

The Oxygen of Amplification

Whitney Phillips

Phillips, Whitney. *The Oxygen of Amplification*. Data and Society Research Institute, 2018. https://datasociety.net/wp-content/uploads/2018/05/FULLREPORT_Oxygen_of_Amplification_DS.pdf

Crux

This practitioner-focused report incorporates interviews with over 50 mainstream journalists to provide an overview of an industry under pressure to deliver page views and cover "trolls" despite the "disgust" felt by accidentally propagating extremist ideology. This report is divided into three parts: a historical overview of the relationship between journalists and far-right manipulators during the 2016 US presidential election, the consequences of reporting on problematic information, and proposed editorial practices for addressing newsworthiness, false information, and harassment.

Highlights

- Phillips argues that amplification increases the likelihood that similar disinformation and harassment tactics will be used in the future and risks lending credence to false narratives.
- Phillips identifies four broad categories of structural challenges that induce journalists to cover problematic information:
  - Journalism is supported by advertising, which places intense pressure on management to increase page views and demonstrate return on their corporate backers' investments.
  - Journalism is guided by the basic tenet to publish, and therefore to spread, newsworthy information (a.k.a. the *information imperative*).
  - Labour issues, such as excessive word, story, and/or traffic quotas, contribute to the amplification.
  - The imagined homogeneity of audiences and hegemony of newsrooms. Publications tend to present content that aligns with white, middle-class to upper-middle-class sensibilities, which may impact editorial choices.

How Russia Targets the U.S. Military

Ben Schreckinger

Schreckinger, Ben. "How Russia Targets the U.S. Military." *Politico Magazine*, June 12, 2017. http://www.politico.com/magazine/story/2017/06/12/how-russia-targets-the-us-military-215247.

Crux

This article examines the tactics pro-Russian government organizations have used to gain influence over members of the U.S. military. In particular, it explores the relationship between the American website, *Veterans Today*, and their partnership with Russian media outlet *New Eastern Outlook*. The article also highlights the use of low-level tactics, such as individuals posing as beautiful women and trying to befriend U.S. soldiers on Facebook in order to then post content sympathetic to the Russian government.

Highlights

- Joel Harding, a former Army intelligence officer who now works as an independent researcher, describes *Veterans Today*, *Veterans News Now,* and

South Front as "Russian proxy sites."

- According to a report by cybersecurity firm SecureWorks, of the people targeted by Fancy Bear outside of the former Soviet Union, 41% were current or former members of the military; 22% were authors and journalists; NGOs, 10%; political activists, 4%; and government personnel, 8%.
- *Time* reported that American counterintelligence officials concluded in March 2017 that Russian hackers were targeting 10,000 Department of Defense employees.

The Spread of Low-Credibility Content by Social Bots

Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. "The Spread of Low-Credibility Content by Social Bots." *Nature Communications* 9 (2018). https://www.nature.com/articles/s41467-018-06930-7.

Crux

This paper examines how social bots were able to promote hundreds of thousands of false and misleading articles during and following the 2016 U.S. presidential campaign and election. The authors, by analyzing 14 million messages spreading 400 thousand articles on Twitter, found that social bots played a disproportionate role in spreading and repeating misinformation and are able to do so by targeting users with many followers through replies and mentions.

Highlights

- This study was performed using two tools developed in-house:
  - the *Hoaxy* platform to track the online spread of claims
  - the *Botometer* machine learning algorithm to detect social bots
- The authors find that human accounts retweet claims posted by bots as much as by other humans, indicating perhaps that humans can't tell the difference between bot accounts and human accounts.

Lies, Damn Lies, and Viral Content: How News Websites Spread (And Debunk) Online Rumors, Unverified Claims, and Misinformation

Craig Silverman

Silverman, Craig. *Lies, Damn Lies and Viral Content.* New York: Tow Center for Digital Journalism, 2015. https://doi.org/10.7916/D8Q81RHH.

Crux

In analyzing over 1,500 news articles about more than 100 online rumours that circulated in the online press between August and December of 2014, the author finds that online journalism standards have decreased significantly, giving attention to unverifiable rumours that traditionally would not have been worth the attention.

Highlights

- Tracking the source of an article's claims is increasingly difficult as sites prefer to link to other media reports, which simply link to more media reports.
- News organizations are inconsistent at best at following up on the rumours and claims they offer in their initial coverage.
- When reporting rumours and unverified claims, news organizations tend to bias the reader toward thinking the claim is true. Even though they hedge their language by using words like "reportedly" or "claims" to convey that information they are passing on is unverified, readers do not pick up on this and may be easily misinformed.
- *Hostile media effect:* "People who watched the exact same report came away with different perceptions of bias, based on which news

organization they thought produced it."

- Silverman notes the following trends in online journalism: pointing out something is interesting just because the Internet is talking about it; publishing content on unverified claims and not following up on it; unverified claims attract more interest than corrections or updates; and fake news articles generate far more shares and social interactions than debunking articles.

Digital Threats to Democratic Elections: How Foreign Actors Use Digital Techniques

By Chris Tenove, Jordan Buffie, Spencer McKay, David Moscrop, Mark Warren, Maxwell A. Cameron

Tenove, Chris, Jordan Buffie, Spencer McKay, David Moscrop, Mark Warren, Maxwell A. Cameron. *Digital Threats To Democratic Elections: How Foreign Actors Use Digital Techniques.* Vancouver, BC: Centre for the Study of Democratic Institutions, UBC, 2018. https://democracy.arts.ubc.ca/2018/01/18/digital-threats/.

Crux

This report addresses some of the broader questions facing digital threats from foreign actors and in doing so concludes with five thematic observations:

1) Foreign actors employ four key digital techniques: hacking attacks on systems and databases; mass misinformation and propaganda campaigns; micro-targeted manipulation; and trolling operations.
2) Digital interference is not limited to its impact on electoral outcomes. Other negative outcomes include decreased opportunities for citizen participation, vibrant public deliberation, and effective rules and institutions.
3) State and non-state actors use the aforementioned digital techniques, and often do so in 'partnership' with domestic actors.
4) There are five key sources of vulnerability to digital interference: deficits in digital literacy; shortcomings in the design and policies of social media platforms; high levels of political polarization; inadequate electoral regulations; and the lack of international laws and practices to address cyber-attacks and information operations.
5) There are many possible counter-measures to digital interference, but no proven solutions.

Highlights

- The report notes four types of political bots, proposed by Fenwick McKelvy and Elizabeth Dubois: *dampeners* suppress messages, *amplifiers* make messages appear more popular than they are, *transparency* bots share information relevant to informed citizenship, and *servant* bots are used by government and organizations to answer questions or provide other services.
- Bots and sockpuppets (human-operated fake accounts) can be purchased, though Russia, China, and other governments may task staff to act as sockpuppets.
- During the 2017 French elections, hundreds of bots published false and defamatory information against candidate Emmanuel Macron. However, the dissemination of leaked information was limited by its timing, by decisions of journalism outlets not to give the leaks extensive coverage (for legal and professional reasons), and by the electoral commission's prohibition on publishing hacked documents during the legal blackout period immediately preceding the election.

Mal-uses of AI-generated Synthetic Media and Deepfakes: Pragmatic Solutions Discovery Convening

Witness and First Draft

Witness and First Draft. *Mal-uses of AI-generated Synthetic Media and Deepfakes: Pragmatic Solutions Discovery Convening.* July, 2018. http://witness.mediafire.com/file/q5juw7dc3a2w8p7/Deepfakes_Final.pdf/file.

Crux

This report summarizes the discussions and recommendations from a convening of 30 independent and company-based technologists, machine learning

specialists, academic researchers in synthetic media, human rights researchers, and journalists on the threats of AI-generated disinformation. The report offers a broad overview of the negative impacts of artificial intelligence and disinformation; potential threat scenarios, and how civil society can counter such threats.

Highlights

- The report stresses that humans are not good at discerning real from fake video content, but that machines are. The report also cites new developments in the emerging field of "automatic forensics."
- The report notes that video and images are a far bigger threat than text-based disinformation, especially in low-literacy countries.
- The report notes that increased authentication may lead to tradeoffs between security and privacy, which will likely create further risks for those already vulnerable.
- The report contends there are two classes of threats: "the inability to show that something real is real, and then the ability to fake something as if it was real." They note that the former is more worrisome.
- Some of their recommendations include watermarking by commercial software, supporting platform-based research and collaboration, investing in automated detection, and use of provenance data and chain-of-custody (e.g., blockchain).

Social Media

Although research into social media's role in the spread of disinformation can also be categorized under the Creation and Dissemination section, the following sources speak specifically to individual social media platforms like Facebook or Twitter. In the last few years, particular attention has been paid to how algorithms that promote trending or recommended content have led to conspiratorial, false, and extremist content, and how the most popular social media platforms have been used to manipulate public opinion.

Sources of note:

On Youtube's algorithms: 'Fiction Is Outperforming Reality': How Youtube's Algorithm Distorts Truth
On the dissemination patterns of true versus false news on Twitter: The Spread Of True And False News Online
On Facebook's impact in the Philippines: How Duterte Used Facebook To Fuel the Philippine Drug War
Examples of astroturfing: Cloaked Facebook pages: Exploring fake Islamist propaganda in social media

How Duterte Used Facebook to Fuel the Philippine Drug War

Davey Alba

Alba, Davey. "How Duterte Used Facebook To Fuel the Philippine Drug War." *Buzzfeed News*, September 4, 2018. https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war.

Crux

This article argues that Facebook has enabled President Duterte's authoritarian grip on the Philippines through the use of influencers, disinformation, and harassment. The author cites examples like the onslaught of disinformation targeting opposition Senator Leila De Lima, which culminated in her being jailed. Facebook's fact-checking initiatives, however, have been met with criticism as the bulk of the workload has fallen to those working in media and academia in the Philippines. Instead of writing investigative pieces, journalists are spending most of their time fact-checking Facebook.

Highlights

- Due to subsidies that keep Facebook free to use on mobile phones and the relative higher cost of accessing the Internet outside of Facebook, most Filipinos consider the platform itself as the Internet.
- While disinformation in the United States tends to drive clicks to third-party websites, disinformation lives mostly on Facebook in the Philippines.
- In addition to pushing disinformation, the state also engages in censorship, such as the revocation of news outlet *Rappler's* license to operate.
- "People went from no access to news to gaining access only through Facebook's algorithm-driven news feed." - Clarissa David (professor of political communication at the University of the Philippines)

Social Media and Fake News in the 2016 Election

Hunt Allcott and Matthew Gentzkow

Allcott, Hunt and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (2017): 211–36. https://doi.org/10.1257/jep.31.2.211.

Crux

Drawing on web browsing data, fact-checking websites, and a 1,200-person post-election online survey, the

authors find that although social media was important, it was not the dominant source of election news. Only 14% of Americans called social media their "most important" source.

Highlights

- The authors confirm that fake news was both widely shared and heavily tilted in favor of Donald Trump (115 pro-Trump fake stories were shared on Facebook a total of 30 million times versus 41 pro-Clinton fake stories that were shared a total of 7.6 million times).
- The authors mention two main motivations for providing "fake news":
  - Monetary: to gain advertising revenue when users click to the original site
  - Ideological: to advance candidates they favour
- When asked what their most important source of 2016 election news was:
  - 23.5% said cable TV
  - 19.2% said network TV
  - 14.8% said website
  - 14.5% said local TV
  - 13.8% said social media
  - 8% said print
  - 6.2% said radio

Cloaked Facebook Pages: Exploring Fake Islamist Propaganda in Social Media

Johan Farkas, Jannick Schou and Christina Neumayer

Farkas, Johan, Jannick Schou, and Christina Neumayer. "Cloaked Facebook pages: Exploring fake Islamist propaganda in social media." *New Media & Society* 20, no. 5 (May 2018): 1850–67. https://doi.org/10.1177/1461444817707759.

Crux

This paper analyzes the spread of inauthentic Islamist propaganda on Facebook in the Danish language by studying 11 cloaked Facebook pages, all of which claimed to be administered by radical Islamists living in Denmark, and the closed Facebook group named *Stop Fake Hate Profiles*. By "cloaked," the authors mean websites purporting to be radical Islamists living in Denmark but instead belong to individuals pushing anti-immigrant, anti-Muslim views. The paper breaks up their findings into five sections: 1) iterations of the cloaked Facebook pages; 2) how the cloaks were designed; 3) the reactions to the cloaked pages; 4) contesting the validity of the cloaks; and 5) what challenges lay ahead with regards to deceptive media.

Highlights

- Across the 11 Facebook pages, there was a general process of overlapping steps:
  1) the Facebook pages were created in a manner that disguised them as representing radical Islamist identities through symbolism, text, and imagery;
  2) the pages were disseminated through hateful and aggressive posts directed at the Danish people and state;
  3) users reacted to the posts with comments and shares without questioning the pages' authorship;
  4) the Facebook group *Stop Fake Hate Profiles* on Facebook acted by contesting the pages' authorship and reporting them to Facebook;
  5) Facebook deleted the pages due to violations of their content policies.
- Although the media coverage identified the Facebook pages as cloaked and producing racism, their reporting increased the pages' visibility.

Rumor Cascades

Adrien Friggeri, Lada A. Adamic, Dean Eckles, Justin Cheng

Friggeri, Adrien, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. "Rumor Cascades." In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8122/8110.

Crux

This paper examines the spread of rumours on Facebook by considering the uploading and re-sharing of photos as well as the copying-and-pasting of text posts. By tracking

the path of uploads and reshares, the researchers are able to determine the diffusion of such rumours, their longevity, and their references outside of Facebook. The rumours in question come from Snopes.com, a well-known site that collects and verifies claims. The study found a variety of trends: 1) political stories amount to 32% of rumour cascades, though they only account for 22% of Snopes stories; 2) food, crime, and medical photos also were uploaded more than expected; 3) 9/11 photos are underrepresented when compared to Snopes; 4) although false rumours are predominant, true rumours are more viral (i.e., they result in larger cascades); 5) reshares about false rumours are 4.4 times as likely to be deleted when snoped than when not; 6) rumours that are snoped and deleted still continue to propagate through their re-shares; 7) rumours can exhibit "bursty" behaviour where it dies down for a few weeks or months then suddenly revives and becomes popular again.

Highlights

- Comments containing Snopes links may retard the ability of the reshare to replicate, even if the original reshare itself is not deleted.
- 45% of rumours covered on Snopes are false, whereas 62% of rumour cascades on Facebook are tagged as false.
- Only 9% of rumour cascades are true on Facebook, whereas 26% of all Snopes stories were found to be true.

Parkland Conspiracies Overwhelm the Internet's Broken Trending Tools

Issie Lapowsky

Lapowsky, Issie. "Parkland Conspiracies Overwhelm The Internet's Broken Trending Tools." *Wired*, February 21, 2018. https://www.wired.com/story/youtube-facebook-trending-tools-parkland-conspiracy.

Crux

This article discusses the failings of Twitter and Facebook's trending algorithms in propagating a conspiracy video that alleges one of the survivors of the 2018 Parkland shooting was an actor. For a brief moment, this video received the top spot on Youtube under their Trending section. It appears this mistake was an accident on the part of Youtube's algorithm, which says its system "misclassified" the conspiracy video "because the video contained footage from an authoritative news source." The author also points out that Facebook had similar problems when articles containing the same accusations as the video surfaced in their Trending section, which received higher billing than more reputable sources like CBS Boston and the Toronto Star.

Highlights

- Content moderation by humans has been used in the past but has also received criticism that conservative news is being suppressed.
- Dipayan Ghosh, a fellow at the think tank New America and a former employee in Facebook's privacy and public policy team said, "Facebook has grown tremendously in its size and influence around the world, and part of that is because of the promotion of particularly engaging content that attracts eyeballs and keeps them on the screen for long periods of time."
- Facebook and Youtube pledge to hire 10,000 more human moderators but it is a losing battle as more than 400 hours of content gets uploaded to YouTube each minute.

'Fiction is Outperforming Reality': How Youtube's Algorithm Distorts Truth

Paul Lewis

Lewis, Paul. "'Fiction Is Outperforming Reality': How YouTube's Algorithm Distorts Truth." *The Guardian*, February 2, 2018. https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

Crux

This article examines the accusations that Youtube's recommendations algorithm tends to promote conspiratorial videos and fake news. Much of the article revolves around research conducted by Guillaume Chaslot, an ex-Google employee and computer programmer with a PhD in artificial intelligence,

who built a program to reverse-engineer YouTube's recommendation algorithm. His research suggests that YouTube systematically amplifies videos that are divisive, sensational, and conspiratorial.

Highlights

- When Chaslot's program found a seed video by searching the query "who is Michelle Obama?" the chain of "up next" suggestions mostly recommended videos that said she "is a man."
- More than 80% of the YouTube-recommended videos about the pope detected by Chaslot's program described the Catholic leader as "evil," "satanic," or "the anti-Christ."

Polarization, Partisanship and Junk News Consumption over Social Media in the US

Vidya Narayanan, Vlad Barash, Bence Kollanyi, Lisa-Maria Neudert, and Philip N Howard.

Narayanan, Vidya, Vlad Barash, Bence Kollanyi, Lisa-Maria Neudert, and Philip N. Howard. *Polarization, Partisanship and Junk News Consumption over Social Media in the US (Vol. 1)*. Oxford, UK: Project on Computational Propaganda, 2018. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/02/Polarization-Partisanship-JunkNews.pdf.

Crux

This article examines which groups on Facebook and Twitter are most likely to consume and share "junk news," which the authors define as misleading, deceptive, or incorrect information purporting to be real news about politics, economics, or culture. This includes various forms of extremist, sensationalist, conspiratorial, masked commentary, and fake news. They ultimately find that on Twitter, Trump supporters followed by conservatives are most likely to circulate junk news, while on Facebook, extreme hard right pages (different from Republican pages) share the widest range of junk news sources.

Highlights

- There is little overlap in sources of news consumption between supporters of the Democratic Party and the Republican Party. They find that Democrats show high levels of engagement with mainstream media sources compared to Republicans and conservatives.
- On Twitter, the "Trump Support Group" shares 95% of the junk news sites on the watch list, and accounted for 55% of junk news traffic in the sample.
- On Facebook, the "Hard Conservative Group" shares 91% of the junk news sites on the watch list, and accounted for 58% of junk news traffic in the sample.

The Spread of True and False News Online

Soroush Vosoughi, Deb Roy, and Sinan Aral

Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The Spread of True and False News Online." *Science* 359, iss. 6380 (2018): 1146–1151. https://doi.org/10.1126/science.aap9559.

Crux

In analyzing ~126,000 stories tweeted by ~3 million people more than 4.5 million times, the researchers found that fake news on Twitter spread significantly farther, faster, deeper, and more broadly than the truth in all categories of information. This effect was more pronounced for false political news than all other types. The study also found that bots spread true and false news at about the same rates, implying it is humans who are propagating fake news at faster rates. The researchers theorize that this may be because of the novelty of fake news, which humans are wired to value more.

Highlights

- There were clear increases in the total number of false political rumours during the 2012 and 2016 U.S. presidential elections and a spike in rumours that contained partially true and partially false information during the Russian annexation of Crimea in 2014.
- Politics was the largest rumour category, with ~45,000 cascades, followed by urban legends, business, terrorism, science, entertainment, and natural disasters.

- Whereas the truth rarely diffused to more than 1,000 people, the top 1% of false-news cascades routinely diffused to between 1,000 and 100,000 people.
- It took the truth about six times as long as falsehood to reach 1,500 people.
- Users who spread false news had significantly fewer followers, followed significantly fewer people, were significantly less active on Twitter, were verified significantly less often, and had been on Twitter for significantly less time.

Advertising and Marketing

The sources included in this section deal primarily with the ways in which advertising networks and marketing services aid in the dissemination and propagation of disinformation. While the sources below would also be appropriate in the Creation and Dissemination and the Social Media sections, there is enough literature specifically on the role of advertising platforms and marketing to warrant its own section. Researchers in recent years have focused on microtargeting, corporate tracking and retention of user data, and the enormous increases of political spend on digital advertising.

Sources of note:

Broad overview of advertising tactics: #DigitalDeceit: The Technologies Behind Precision Propaganda on the Internet
On microtargeting: Online Political Microtargeting: Promises and Threats for Democracy
On native advertising: Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell
On political advertising: Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech

Fake News and the Economy of Emotions: Problems, Causes, Solutions

Vian Bakir and Andrew McStay

Bakir, Vian and Andrew McStay. "Fake News and The Economy of Emotions: Problems, Causes, Solutions." *Digital Journalism* 6, no. 2 (2017): 154-175. https://doi.org/10.1080/21670811.2017.1345645.

Crux

The authors contend that the core of the fake news problem is the "economics of emotion," which they define as how emotions are leveraged to generate attention and viewing time, which then converts to advertising revenue. Based on this premise, they suggest that the potential to manipulate public sentiment via "empathically optimised automated fake news" is a problem that will happen very soon, and as such, more attention should be paid to the role of digital advertising, both in causing and combating contemporary and near-horizon fake news phenomena. The article includes an analysis of proposed solutions, weighing the opportunities and challenges of each, and concludes with a forward look at the next era of fake news, including algorithmically automated news that feeds on emotions.

Highlights

- The authors note that the contemporary fake news phenomenon is a logical outcome of five features of the digital media ecology: the financial decline of legacy news; the news cycle's increasing immediacy; the rapid circulation of misinformation and disinformation via user-generated content and propagandists; the increasingly emotionalised nature of online discourse; and the growing number of people financially capitalising on algorithms used by social media platforms and internet search engines.
- Automated journalism (or "algo-journalism") is already being increasingly used by legacy news agencies such as the *Associated Press* to provide detail-heavy news that does not require human interpretation or analysis.

When News Sites go Native: Redefining the Advertising-Editorial Divide in Response to Native Advertising

Matt Carlson

Carlson, Matt. "When News Sites go Native: Redefining the Advertising-Editorial Divide in Response to Native Advertising." *Journalism* 16, no. 7 (October 2015): 849–65. https://doi.org/10.1177/1464884914545441.

Crux

This article explores the problems and ethics associated with "native advertising," in which paid content promoting a product or idea is published in a way that mimics real editorial content. The case study the author chooses to highlight is a 2013 Church of Scientology piece of sponsored content that was published on the *Atlantic* website. In examining the various reactions to the incident, the author finds that this was not a clear-cut ethical violation, representing ongoing norm construction efforts concerning the relationship between advertising and online journalism. In this particular case, the *Atlantic* was criticized for taking money from the Church of Scientology and for blurring the lines between advertising and journalism.

Highlights

- The practice of native advertising has become increasingly common, with nearly three quarters of online publishers using some form of it in mid-2013.
- The author highlights how decisions over the provision of resources are all driven by revenue, affecting newspaper bureaus, sections/desks, new positions, terminations, and technological innovations.
- The Project for Excellence in Journalism (2013) estimates that for newspapers, every dollar earned online equates to 15 dollars lost from print.

Awkward Conversation with Facebook

David Carroll

Carroll, David. "Awkward Conversation With Facebook: What Happened When I Caught Them Defaulting Us Back Into Behavorial Ad Tracking and Targeting." *Medium*, June 1, 2016. https://medium.com/@profcarroll/awkward-conversation-with-facebook-ef1734ecdc62.

Crux

This article reveals how Facebook defaults users into tracking, regardless of whether they have shown interest in opting out of interest-based advertising in one or more other contexts already. They do so by creating new settings and controls, which users are not notified of. The author speculates this is so Facebook can sidestep the FTC Consent Decree which prohibits them from changing existing settings without notifications related to privacy.

Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell

Mara Einstein

Einstein, Mara. *Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell.* New York: OR Books, 2016.

Crux

Einstein covers modern advertising tactics that blur the lines between unbiased journalism and advertising. It covers the rise of "sponsored content," "covert selling," content marketing, and how sharing is the ultimate form of the "subtle sell."

Highlights

- Rise in non-conventional advertising stems from ad blockers, "banner blindness," and declining ad revenues for print and TV.
- In order to boost ad revenues, Google changed

their algorithm to reward mobile-friendly sites, which buried content that may have been more relevant, but did not have the required responsive elements, such as big clickable links and larger fonts.

#DigitalDeceit: The Technologies Behind Precision Propaganda on the Internet

Dipayan Ghosh and Ben Scott

Ghosh, Dipayan and Ben Scott. *#DigitalDeceit:The Technologies Behind Precision Propaganda on the Internet.* New America, 2018. https://www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/.

Crux

This paper examines the role of Internet-based advertising and media platforms in the dissemination of political disinformation. The authors, in exploring the "entire toolbox of precision propaganda," organize the methods into five categories:

1) Behavioural data collection (web tracking; cookies; location tracking; cross-device tracking)
2) Digital advertising platforms (sponsored or promoted content; automation, profiled audience segments)
3) Search engine optimization
4) Social media management software
5) Algorithmic advertising technology (digital ad mediation; social media management software; lookalike audiences)

Their recommendations focus on transparency, public education, public service journalism, cybersecurity, corporate responsibility, consumer empowerment, and ad tech regulation.

Highlights

- The more successful the ad buy (including disinformation), the more effective the successive ad buys will be because the ad platform has learned more about the best targets.
- Google owns over 75% of the search market in the United States.
- The top five search results get around 75% of the traffic. The first page of links (top 10) get 95 percent.
- Most of search engine optimization is based on three things, which can all be gamed: content, links, and popularity.

The Man Who Made the Republican Internet — And Then Sold it to Far-Right Nationalists Overseas

Henry J. Gomez

Gomez, Henry J. "The Man Who Made The Republican Internet — And Then Sold It To Far-Right Nationalists Overseas." *Buzzfeed*, December 19, 2017. https://www.buzzfeed.com/henrygomez/the-man-who-made-the-republican-internet-and-then-sold-it.

Crux

This article highlights Harris Media founder, Vincent Harris's ascent in the world of campaign marketing for far-right and Republican candidates. It covers his work with Ted Cruz, Rand Paul, Mitch McConnell, and overseas for Marine LePen, Benjamin Netanyahu and the Likud Party, and Alternative für Deutschland. His tactics included creating "viral" videos, memes, divisive content, anti-immigrant material, and personal attacks against opponents. The author of the article points out that domestically Harris Media does not seem to be as active anymore relative to its overseas ventures.

Highlights

- Two former employees said the firm recently assisted Kenyan President Uhuru Kenyatta, who last month was sworn in for a second term after a disputed election and election do-over marred by violent clashes.
- There are noticeably fewer Senate campaigns paying Harris Media as the calendar flips toward the 2018 midterms, according to Federal Election Commission reports.

Psychological Targeting As an Effective Approach to Digital Mass Persuasion

S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell

Matz, S. C., M. Kosinski, G. Nave, and D. J. Stillwell. "Psychological Targeting as an Effective Approach to Digital Mass Persuasion." *PNAS* 114, no. 48 (2017): 12714–19. https://doi.org/10.1073/pnas.1710966114.

Crux

In three field experiments that reached over 3.5 million people, the authors find that targeting people with persuasive appeals tailored to their psychological profiles can be used to influence their behaviour as measured by clicks and conversions.

Highlights

- The experiments were run using Facebook advertising, which does not allow one to target psychological traits but what an individual "Likes." By extracting lists of Likes indicative of high and low levels of each of these traits from the myPersonality.org database, the authors were able to compute the average personality trait levels for each Like. They then selected 10 Likes characterized by the highest and lowest aggregate extraversion and openness scores.
- The effectiveness of large-scale psychological persuasion in the digital environment heavily depends on the accuracy of predicting psychological profiles from people's digital footprints.

Viral Content: The First Warning Sign of Fraud

Michael Misiewicz and Laura Yu

Misiewicz, Michael and Laura Yu. *Viral Content: The First Warning Sign of Fraud.* Appnexus, 2017. https://www.appnexus.com/sites/default/files/whitepapers/appnexus_viralcontentwhitepaper_december2017.pdf.

Crux

Written by the ad tech platform AppNexus, this white paper highlights some of their most salient observations that have come out of analyzing advertising fraud from their data. Most notably:

- Online advertising fraud poses an existential threat to programmatic advertising.
- Viral content goes hand-in-hand with online advertising fraud (publishers specializing in viral content are much more likely to receive non-human traffic).
- Traffic acquisition vendors deserve careful scrutiny.
- "Fake News" and hate speech are often just another form of viral content.

Highlights

- Popular ad fraud methods today include: browser or device hijacking programs; bot networks; and ad-stuffing (a tactic in which the publisher fills a web page with invisible ads the user can't see).
- The Interactive Advertising Bureau estimated that ad fraud cost the industry $8.2 billion in 2015.
- Viral content producers who rely heavily on trending topics for revenue generation are more susceptible to ad fraud. Some purchase massive audiences from vendors at below-market rates, without investigating how their partners deliver so much traffic for so little money. They then sell this traffic to advertisers at a higher rate.
- Traffic acquisition schemes in the gray area include content discovery networks and social media pay-per-clicks.
- Suspicious publishers typically use a handful of cheap tricks to pass themselves off as legitimate websites: "borrowed" content; awkward formatting; partially-scraped content; out of place links; no author pages; heavy ad loads; reliance on embedded video; and missing sections.

Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech

Anthony Nadler, Matthew Crain, and Joan Donovan

Nadler, Anthony, Matthew Crain, and Joan Donovan. *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech*. New York: Data & Society Research Institute, 2018. https://datasociety.net/output/weaponizing-the-digital-influence-machine/.

Crux

This report identifies the technologies and conditions that enable political actors to "weaponize" digital advertising infrastructure for the purpose of influencing audiences. The authors refer to this as the "Digital Influence Machine" (DIM), which has been developed by advertising platforms and other intermediaries. The DIM also includes consumer monitoring, audience-targeting, and automated tech. In analyzing these tactics and technologies, the report highlights three key conditions that have led to "weaponization of the DIM": the decline of professional journalism; the growth of financial resources devoted to political influence; and the expansion of targeted advertising in the absence of effective oversight.

Highlights

- The authors highlight the ways in which surveillance and profiling contribute to making targeting audiences easier. Data brokers like Axiom, Experian, and Oracle, for example, enable advertisers to combine profile data obtained from various online and offline contexts.
- The report notes that many ad platforms have dedicated teams tasked with providing technical assistance and other services to large political spenders.
- The report underscores the role of "dark money," which the authors define as "an overarching term for money spent on political influence campaigns where the identities of large donors are concealed from the public".
- Following the 2010 Supreme Court ruling in Citizens United v. Federal Elections Commission, some of the protections of the Bipartisan Campaign Reform Act of 2002 were rolled back. This allowed advocacy organizations to receive unlimited funds from donors while running ads directly calling for the election or rejection of a candidate, so long as these groups did not coordinate directly with a candidate's campaign.
- The authors make three suggestions for mitigating the harms from politically motivated manipulations of the Digital Influence Machine: ad tech companies should refuse to work with dark money groups; user consent prior to being shown political ads that are a part of a split-test; and further development of ethical guidelines on political advertising, with independent committees representing diverse community stakeholders.

Google Serves Fake News Ads in an Unlikely Place: Fact-Checking Sites

Daisuke Wakabayashi and Linda Qiu

Wakabayashi, Daisuke and Linda Qiu. "Google Serves Fake News Ads in an Unlikely Place: Fact-Checking Sites." *The New York Times*, October 17, 2017. https://www.nytimes.com/2017/10/17/technology/google-fake-ads-fact-check.html.

Crux

This article discusses how Google's Adsense ad network inadvertently puts ads on fact-checking websites that lead to fake news websites. Google refers to this as "tabloid cloaking" and in 2016 suspended over 1,300 websites for doing so. These websites mimic real websites in look and feel but after a few sentences transition into an ad for something completely different, like anti-aging cream. Due to the automation of the ad network, sites which run Adsense have little control over what appears. Likewise, companies advertising products have little control over where their ads end up.

Highlights

- Early 2017, Google said it would crack down on

misinformation sites by kicking 340 websites and 200 publishers off its AdSense platform.

Online Political Microtargeting: Promises and Threats for Democracy

Frederik J. Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes de Vreese.

Zuiderveen Borgesius, Frederik J., Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balázs Bodó, and Claes de Vreese. "Online Political Microtargeting: Promises and Threats for Democracy." *Utrecht Law Review* 14, no. 1 (2018): 82-96. http://doi.org/10.18352/ulr.420.

Crux

This paper defines political microtargeting and offers an analysis of its benefits and harms. Combining insights from a legal and social science perspective, the authors focus primarily on European states and the United States. The paper also discusses why the threats of political microtargeting, while serious, should not be overstated and what solutions policymakers could explore.

Highlights

- Online political microtargeting is defined as finely honed messages targeted at narrow categories of voters based on demographic characteristics and consumer and lifestyle habits.
- One of the benefits of microtargeting is communicating with audiences that traditionally would not have been reached via mass broadcast methods like television or radio.
- The authors outline three types of threats to citizens: invasion of privacy; manipulation; being purposely ignored.
- The authors discuss the differences between the United States and Europe, such as the use of data brokers in the U.S., data protection laws in the EU, and electoral systems.
- The article notes that any attempts to restrict political communication must comply with Article 10 of the European Convention on Human Rights, which guarantees the right to freedom of expression.

Political Science and International Relations

This section looks at disinformation from a political science perspective and encompasses a wide range of research questions and methodologies. The sources here are less about tools and tactics and more about the role of disinformation in geopolitics, governance types, and information operations. Broadly speaking, they try to answer questions about what types of states use disinformation, why, and what their real world impacts on democracy and state power are. The research on all these questions, however, is still nascent with little consensus on how effective disinformation is in achieving state goals.

Sources of note:

On why disinformation produces stability versus chaos: Common-Knowledge Attacks on Democracy
On state-sponsored harassment of civil society: State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns
Effects of Russian interference: Russia Has Been Meddling in Foreign Elections for Decades. Has It Made a Difference?
Overview of threats to the democratic process and national security: WHO SAID WHAT? The Security Challenges of Modern Disinformation
Framework for analyzing state responses: How Can European States Respond To Russian Information Warfare? An Analytical Framework

Russian Information Warfare as Domestic Counterinsurgency

Stephen Blank

Blank, Stephen. "Russian Information Warfare As Domestic Counterinsurgency." *American Foreign Policy Interests* 35, no. 1 (2013): 31–44. https://doi.org/10.1080/10803920.2013.757946.

Crux

This article examines the use of information warfare by the Kremlin as a tool of counterinsurgency on its own domestic population. He characterizes it as distinct from Western forms of information warfare in that it targets both external and internal threats and uses the cognitive, social, and economic aspects of cyberspace along with the technical to achieve its goals. As such, "the mental sphere, a people's identity, and its national and cultural identity have already become battlegrounds." In his example of the Chechen war, the author argues that "information warfare was both a surrogate for missing combat power and a strategic weapon in its own right that was designed for a, if not the, critical front in the conflict."

Highlights

- The author argues Russia has felt under threat ever since the breakup of the Soviet Union and the wave of "colour" revolutions in Georgia, Ukraine, and Kyrgyzstan. Because of this, they feel that they are under a constant state of conflict.
- In analyzing public statements and official documents and positions taken by members of the Russian military, the author emphasizes how information warfare will continue to rise. According to Russian Colonel S. G. Chekinov,

"electronic warfare will become an independent operation in its own right, not just a support operation."

- In discussing the role of information warfare in Chechnya, the author highlights the Russians' systematic campaign to capture Russian hearts and minds (as opposed to Chechnyan hearts and minds) through media campaigns that raised public support for the armed forces, isolated insurgents from domestic and foreign support, and framed the war as a terrorist campaign.
- In advance of the large opposition demonstrations in Moscow in September 2012, the government launched a major disinformation campaign comprising a rapid flood of spam tweets that resembled authentic tweets. The messages used inflammatory language in support of the demonstrations to discourage moderates from attending and associated opposition leaders like Alexei Navalny and Sergei Udaltsov with the faked messages to discredit them.

WHO SAID WHAT? The Security Challenges of Modern Disinformation

Canadian Security Intelligence Service (CSIS)

Canadian Security Intelligence Service. *WHO SAID WHAT? The Security Challenges of Modern Disinformation.* Ottawa: CSIS, 2018. https://www.canada.ca/content/dam/csis-scrs/documents/publications/disinformation_post-report_eng.pdf.

Crux

This report highlights the papers that were presented during a disinformation workshop conducted by CSIS through their Academic Outreach program in November 2017. While it acknowledges the role of anti-government ideologues who push conspiracy theories and disinformation in the Philippines, Syria, and China, the report places heavy emphasis on Russia (mentioned 224 times). The first 50 pages summarizes Russia's various disinformation campaigns, their methods and alliances, and how they've evolved. The report then continues to cover the role of Twitter during Brexit, fake news on Syria, China's approach to influence operations, the impact of social networks in the Philippines, disinformation in Ukraine, and the profit motive behind fake news. The report repeatedly stressed the threat disinformation poses to democracy and truth.

Highlights

- "Virtually every type of action [Russia] has undertaken against the West was first implemented in Russia, against the Russian people, and against Russia's many ethnic, national and religious minorities."
- "Russia, China and the Philippines use disinformation techniques to control their internal populations."
- The report notes that in the medium-term, better media literacy is needed to combat disinformation campaigns.
- Fake-news "start-ups" are on the rise with at least 200 or 300 people engaged in similar enterprises across Kosovo, Macedonia, and Albania.
- The report is alarmist, perhaps justifiably so. For example, this line in the executive summary: "Disinformation poisons public debate and is a threat to democracy… There are many ways for governments and organisations to counter the threat, but there is no guarantee that even effective counter-campaigns can defeat the high volume flow of malicious communications."
- After Ukraine and the US, the report suspects the Baltic states will be Russia's next targets, if not already. Russian propaganda in the Baltics draws upon "'widespread poverty', 'depopulation', raging far-right ideology and the 'semi-colonial status' of these countries."

International Relations in the Age of 'Post-Truth' Politics.

Rhys Crilley

Crilley, Rhys. "International Relations in the Age of 'Post-Truth' Politics." *International Affairs* 94, no. 2 (2018): 417–425. https://doi.org/10.1093/ia/iiy038.

Crux

This book review essay explores how the assumed shift from an age of reason and facts to an age of emotion and lies might impact the discipline of international relations. The author does this through the review of four books: *Post-Truth: The New War On Truth And How To Fight Back* by Matthew d'Ancona; *Post-Truth: How Bullshit Conquered The World* by James Ball; *Post-Truth: Why We Have Reached Peak Bullshit And What We Can Do About It* by Evan Davis; and *The Invention Of Russia: The Rise Of Putin And The Age Of Fake News* by Arkady Ostrovsky. The author finds that D'Ancona, Ball, and Davis are not entirely convincing when describing what 'post-truth' politics is, what caused it, and what should be done about it, and that they place too much emphasis on psychology. Ultimately, the author concludes that "IR as a discipline must place greater focus on understanding the everyday lived experiences of people, and take these as a serious source of inquiry." A greater understanding and discussion of the role misogyny, racism, and xenophobia in disinformation needs to be had.

Highlights

- The author on the alleged sudden shift from reason to emotion in decision-making: "Emotions have always been important in politics, economics and society. Feminists, critical theorists and others outside the mainstream of academic inquiry have argued so for decades. What is new is the recognition, both within the study of these respective fields and within wider public discourse, that emotions matter."
- "Ball's work is at its strongest when discussing how the current media ecology has led to the rise of Trump and Britain's vote to leave the European Union. He deftly highlights how traditional media business models are failing, and draws attention to how PR companies, tech companies and major news sites all play a role in creating a broken media environment whereby almost everyone profits from 'outright fake news or articles of very dubious quality'."
- The author criticizes the "post-truth" books for placing an unconvincing amount of blame on the collective "we," which lets "those who strategically stretched the truth off the hook."
- The author finds that the "starkest omission from these books on 'post-truth' politics is any engagement with gender."

Common-Knowledge Attacks on Democracy

Henry Farrell and Bruce Schneier

Farrell, Henry and Bruce Schneier. "Common-Knowledge Attacks on Democracy." Berkman Klein Center Research Publication No. 2018-7 (October 2018). http://dx.doi.org/10.2139/ssrn.3273111.

Crux

This article develops a theory for how autocracies and democracies operate as information systems, and why democracies are more vulnerable to "measures that "flood" public debate and disrupt shared decentralized understandings of actors and coalitions." The authors argue that while stable democracies may have contested knowledge over who is in charge, there is common public knowledge over who the political actors are and how they may legitimately gain support. In other words, there is a common understanding of how the system works, who won, who lost, and why. Attacks that undermine these collective expectations will thus undermine the elections process, leading to an unstable democracy. Autocracies, on the other hand, have common knowledge over who is in charge and what their ideological and policy goals are, but instead contested knowledge is generated over who the political actors are and how they may gain public legitimacy. This type of information system therefore makes it more difficult for opposition coalitions to gain public support.

Highlights

- The authors define common political knowledge as "the consensus beliefs that hold political systems such as democracies together."
- The authors identify two different kinds of common political knowledge among collective actors and the general public, which are key to the proper functioning of democracy: 1) political institutions and 2) the range of actors, beliefs, and opinions in the society.
- The report acknowledges that although autocracies may benefit from contested political knowledge about non-governmental groups and actors in society and elections processes, they may still want accurate information for themselves about the population, so that they can keep track of their legitimacy. An example the authors give is in the former Soviet Union, which kept tight control of public information through censorship and surveillance while conducting extensive survey polling. The results were available only to elite members of the party.

State, Media and Civil Society in the Information Warfare Over Ukraine: Citizen Curators of Digital Disinformation

Yevgeniy Golovchenko, Mareike Hartmann, and Rebecca Adler-Nissen

Golovchenko, Yevgeniy, Mareike Hartmann, and Rebecca Adler-Nissen. "State, Media and Civil Society in the Information Warfare Over Ukraine: Citizen Curators of Digital Disinformation." *International Affairs* 94, no. 5 (September 2018): 975–994. https://doi.org/10.1093/ia/iiy148.

Crux

In analyzing 950,000 tweets surrounding the downing of Malaysian Airlines Flight 17, the authors find that individual citizens are not passive recipients of state-sponsored messages, but also active disseminators of both disinformation and any attempts to counter disinformation.

Highlights

- Content-analysis was completed only in English.
- The authors coded the Twitter accounts based on the profiles' self-description. Though they account for bots (~2%), they do not account for inauthentic or fraudulent accounts.
- "The most important profile in the counter-disinformation network is the journalistic civil society group 'Ukraine Reporter'. This is followed by the individual account of Eliot Higgins, the founder of the citizen journalist group Bellingcat, together with that group's own account, @bellingcat."

How Can European States Respond to Russian Information Warfare? An Analytical Framework

Maria Hellman and Charlotte Wagnsson

Hellman, Maria and Charlotte Wagnsson. "How Can European States Respond to Russian Information Warfare? An Analytical Framework." *European Security* 26, no. 2 (2017): 153–170. https://doi.org/10.1080/09662839.2017.1294162.

Crux

This article offers an analytical framework that can be used to distinguish between and analyse different governmental strategies for countering strategic narratives. By comparing European states' strategies as illustrative examples, the authors propose four ideal-type models representing different strategies for engagement in information warfare: confronting, blocking, naturalizing, and ignoring. Confronting and blocking are classified as outward projection strategies (target is foreign audience), whereas naturalizing and ignoring are classified as internal projection strategies (target is domestic audience).

1) *Confronting:* Involves an outward-looking strategy that entails actively producing and projecting counter-narratives, often in direct response to a particular narrative.
2) *Naturalizing:* Similar to confronting but far less engaging, and concerns itself with projecting its own narrative without directly contrasting it with the narrative projected by the "other."
3) *Blocking:* Inward-looking and protective. Rather than producing and actively transmitting a narrative, this strategy protects its own narrative by blocking that of the opponent.
4) *Ignoring:* The strategy is one of not responding, but ignoring what is seen as false and manipulated narratives. Essentially a no-narrative strategy.

Highlights

- In 2015, in order to combat Russian's information campaigns, Estonia launched a national public broadcasting network in Russian, the Eesti Televisoon+ (ETV+). This is an example of strategic confrontation.
- An example of blocking is when Latvian authorities temporarily barred the Russian state-owned channel Rossiya RTR from broadcasting in the country for three months in 2014 due to pro-Russian framing of Russia's military intervention in Eastern Ukraine.
- The blocking strategy can be criticised as clashing with values of a free and open society.
- States that depend on ignoring as a strategy have a deep trust in democratic institutions and their ability to defend an "honest, open and just society."
- Sweden tends towards the "ignoring" strategy. Few efforts have been made to build institutions especially set up to coordinate national strategic narratives as counter forces to the Russian propaganda.

On Cyber-Enabled Information/Influence Warfare and Manipulation

Herbert Lin and Jackie Kerr

Lin, Herbert and Jaclyn Kerr. "On Cyber-Enabled Information/Influence Warfare and Manipulation." In *the Oxford Handbook of Cybersecurity*. Oxford University Press: 2018 forthcoming. https://ssrn.com/abstract=3015680.

Crux

This chapter in the forthcoming Oxford Handbook of Cybersecurity on "cyber-enabled information/influence warfare and manipulation (IIWAM)", which the authors argue is a particular vulnerability for the United States and other liberal democracies. The authors begin by defining IIWAM, how to identify a victory versus a loss when under IIWAM, the types of targets and tactics used, and the psychological and emotional underpinnings of IIWAM operations. The authors use Russia as a case study, providing examples from the Russian annexation of Crimea. The chapter closes with recommendations on how to combat IIWAM.

Highlights

- The authors place IIWAM's "battlespace" in the information environment, which they describe as "the aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information." They further define the information environment as having three interrelated dimensions: physical, informational, and cognitive/emotional.
- IIWAM takes place below legal thresholds of "use of force" or "armed attack" to avoid triggering the use of conventional military force in response.
- The authors identify three types of IIWAM: propaganda operations, leak operations, and chaos-producing operations.

Computational Propaganda in Taiwan: Where Digital Democracy Meets Automated Autocracy

Nicholas J. Monaco

Monaco, Nicholas J. "Computational Propaganda in Taiwan: Where Digital Democracy Meets Automated Autocracy No. 2017.2." Project on Computational Propaganda Working Paper No. 2017.2, Oxford, UK, 2017. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Taiwan-2.pdf.

Crux

This working paper discusses the use of "computational propaganda" in Taiwan, which they define as "the assemblage of social media platforms, autonomous agents, and big data tasked with the manipulation of public opinion." In their analysis, the author explores three main questions:

1) Is computational propaganda present in Taiwanese society?
2) What is the composition of computational propaganda in Taiwan (manual vs automated)?
3) Where are campaigns most likely to come from?

They find that domestically, computational propaganda campaigns take the form of manual propaganda through the use of cyber army tactics and automated intelligence-gathering techniques. With regards to cross-Strait propaganda, China far outweighs Taiwan in the use of computational propaganda and again relies heavily on manual tactics as opposed to bots.

Highlights

- All subjects interviewed unequivocally agreed that manual propaganda is alive and thriving in Taiwan. The term "cyber army" is often used to describe this phenomenon in Taiwan.
- There has been a recent proliferation of fake news on LINE, one of the most popular social media platforms in Taiwan.
- Media literacy programs are being pushed throughout Taiwan.
- Johnson Liang, an engineer working in Taiwan, decided in late 2016 to build a LINE bot to combat fake news. His bot, when friended over LINE, will let a user know whether their submitted story is fake or not. In the first three months of the bot's activity, over 5,000 stories were reported to the bot.

Protecting Democracy in an Era of Cyber Information War

Joseph Nye

Nye, Joseph. "Protecting Democracy in an Era of Cyber Information War." *Governance in an Emerging World*, iss. 318 (November 2018). https://www.hoover.org/research/protecting-democracy-era-cyber-information-war.

Crux

This article by Joseph Nye outlines the threats of Chinese and Russian soft and sharp power, their ability to protect themselves by controlling information flows, and the evolution of information warfare. He cautions against outright banning of Russian and Chinese soft power efforts and engaging in offensive information operations. Nye concludes with three interconnected strategies that are needed to effectively defend democracy from cyber information war: domestic resilience and defense, deterrence, and diplomacy.

Highlights

- The author notes that cyber deterrence need not be limited to cyber responses, but can cross domains. He also identifies four major mechanisms that can reduce and prevent adverse actions in cyberspace: threat of punishment, denial by defense, entanglement, and normative taboos.
- Nye acknowledges that a traditional arms control treaties may be too difficult, but setting limits on

certain types of civilian targets and behaviour may be feasible. He cites as an example the Incidents at Sea Agreement in 1972 between the United States and Soviet Union to limit naval behavior that might lead to escalation.

- Nye acknowledges the role social media platforms must play in mitigating against the harms of information warfare, but cautions against regulation. He suggests instead formal means of information sharing between government and the private sector and making the algorithms open to public scrutiny.

State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns

Carly Nyst and Nicholas Monaco. Editor in Chief, Samuel C. Woolley.

Nyst, Carly and Nicholas Monaco. Edited by Samuel C. Wooley. *State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns.* Palo Alto: Institute for the Future, 2018. http://www.iftf.org/statesponsoredtrolling.

Crux

For 18 months, the authors in partnership with NGOs and researchers, examined the role of state-sponsored trolling by conducting a thorough literature review, interviews with targets of state-sponsored trolling, and quantitative analyses of attacks where feasible. The report, which summarizes their findings is split into four sections: 1) latest research; 2) framework for attributing attacks in cyberspace; 3) in-depth case studies pertaining to seven countries; and 4) a series of recommendations. The authors conclude that changes in law will be ineffective in preventing the practice of state-sponsored trolling and that technology companies are the only actors with the ability to curb the practice and effects of state-sponsored harassment campaigns.

Highlights

- The authors' policy recommendations are:
  - Require social media platforms to detect and, in some cases, remove hate speech, harassment, and disinformation. They must do so, however, in a transparent and accountable manner that respects due process and reinforces human rights.
  - Adapt the First Amendment by building upon existing hate speech prohibitions that are permitted by the First Amendment.
  - Create exceptions and add possible new regulations to Section 230 of the Communications Decency Act of 1996.
  - Within technology companies, develop business practices to detect and identify state-linked accounts.
- The authors use the term "Black PR firms," which refer to public relations firms that directly engage in disinformation and/or harassment campaigns against perceived regime opponents.
- Commonalities between disinformation campaigns include:
  - Targeting critics, such as journalists and activists.
  - Narratives that include accusations of colluding with foreign agencies, treason or using violent hate speech to overwhelm targets.
  - Use of bots
- The report identifies four overlapping ways states are responsible for online harassment campaigns:
  - State-executed
  - State-directed or -coordinated
  - State-incited or -fueled
  - State-leveraged or endorsed

When Media Worlds Collide: Using Media Model Theory to Understand How Russia Spreads Disinformation in the United States

Sarah Oates

Oates, Sarah. 2018. "When Media Worlds Collide: Using Media Model Theory to Understand How Russia Spreads Disinformation in the United States." Paper presented at the 2018 *American Political Science Association Annual Meeting*, Boston, MA, September 2018. http://dx.doi.org/10.2139/ssrn.3238247.

Crux

This paper offers an explanation as to why foreign disinformation is so easily disseminated in the United States compared to Russia. The author bases her theory on differences of national media systems, which reflect national political cultures and systems. In the American model, which can be described as Libertarian, the media is expected to work in service of the citizen. The Russian model, which Oates describes as neo-Soviet, works in service of the state. The American Libertarian model, however, is at a disadvantage when it comes to foreign interference due to its openness and the burden it places on citizens to discern fact from fiction. The paper concludes with recommendations to curb disinformation that address the supply-side of the problem.

Highlights

- The author notes that although there is a deep partisan divide on trust in the media within the United States, there appears to be little evidence that Americans feel that the media should not be working to inform citizens.
- "Glasnost was broadly misinterpreted as freedom of the press by Western observers, although later studies provided compelling evidence that the media never shifted from a role as political cheerleaders rather than a democratic institution working to inform and empower citizens."
- In a 2001 survey, the majority of Russian respondents said they preferred the media during Soviet times over the media during glasnost or under President Boris Yeltsin. This was due partly to the feeling of being overwhelmed and disempowered by the negative news under glasnost.

Commanding the Trend: Social Media as Information Warfare

Jarred Prier

Prier, Jarred. "Commanding the Trend: Social Media as Information Warfare." *Strategic Studies Quarterly* 11(Winter 2017): 50–85. http://www.airuniversity.af.mil/Portals/10/SSQ/documents/Volume-11_Issue-4/Prier.pdf.

Crux

This article discusses the new wave of cyberattacks against the United States, which differ from conventional cyberattacks on military infrastructure and communications. The author refers to this as "commanding the trend," in which US adversaries seek to control and exploit social media to discredit public and private institutions and sow domestic strife. He uses two case studies as examples, both looking at Twitter: ISIS during 2014 - 2016 and Russian interference in the 2016 US elections. He concludes with recommendations on how the US can respond to such attacks.

Highlights

- Controlling a trend on Twitter requires four factors:
    1) a message that fits an existing, even if obscure, narrative
    2) a group of true believers predisposed to the message
    3) a relatively small team of agents or cyber warriors
    4) a network of automated "bot" accounts
- There are three methods to control a trend on Twitter:
    - Trend distribution - applying a message to every trending topic

  - Trend hijacking - requires more resources in the form of either more followers spreading the message or a network of "bots" designed to spread the message automatically.
  - Trend creation - requires the most effort, either money to promote a trend or knowledge of the social media environment around the topic, and most likely, a network of automated bot accounts

- 72% of Americans get digital news primarily from a mobile device, and people now prefer online news sources to print sources by a two-to-one ratio.
- The story that the pope endorsed Donald Trump for president received over one million shares on Facebook alone.
- According to an Indiana University School of Journalism study, journalists have an over-reliance on social media for breaking news, which can perpetuate the spread of fake news or unverifiable claims.
- ISIS's mobile app "Dawn of Glad Tidings" provides updates on IS activities and spiritual guidance to the user, but when users download the app, they create an account that links to their Twitter account, which then gives the app permission to tweet using that user's account.

Mapping and Quantifying Political Information Warfare

Share Lab

Share Lab. "Mapping and Quantifying Political Information Warfare." *Share Lab*. October 26, 2016. https://labs.rs/en/mapping-and-quantifying-political-information-warfare/.

Crux

This article summarizes the results of Share Lab's monitoring of over 300 different cases of breaches of online rights and freedoms in the span of almost 3 years. The cases that they have collected and monitored include content blocking, cyber attacks on journalists and independent media, arrests of social media users and bloggers, disinformation campaigns, and electronic surveillance. Though attribution is often difficult, the main goal of their analysis is to explore the various methods of interventions that different political actors or power structures can use to control the online sphere. Their research and analysis focuses primarily on Serbia.

Highlights

- According to Share Lab's research, the average lifespan of the news in Serbian online media is between one and two hours.
- According to a series of leaks published by the web portal Teleprompter.rs in 2014 and 2015, the ruling party SNS (in Serbia) has used different types of software that could be used for astroturfing and other means of the public opinion manipulation. Share Lab has identified three—Valter, Skynet, and Fortress—that allow for comments and voting manipulation online.
- Comment and voting manipulation is also gamified, in which each user is awarded with points for each comment on a news portal. For example, commenting on media outlets that are concordant with official government politics will result in fewer points than commenting on more 'hostile' websites, where there are other commenters with potentially opposing viewpoints.
- Share Lab identifies two forms of astroturfing: Russian "internal enemy" strategy versus China's "cheerleading" strategy. The Russian approach is characterized by personalised content, active political discussions, and attacks on the "internal enemy," whereas the Chinese approach is marked by the strength of sheer numbers and mostly pro-commentary (a.k.a. cheerleading).
- While DDoS attacks are frequent, Share Lab is unable to attribute any of them to any government body or political party.

Russia Has Been Meddling in Foreign Elections for Decades. Has It Made a Difference?

Lucan Ahmad Way and Adam Casey

Way, Lucan Ahmad and Adam Casey. "Russia Has Been Meddling in Foreign Elections for Decades. Has It Made a Difference?" *The Washington Post*, January 8, 2018. https://www.washingtonpost.com/news/monkey-cage/wp/2018/01/05/russia-has-been-meddling-in-foreign-elections-for-decades-has-it-made-a-difference/.

Crux

To assess the impact of Russian meddling in elections, the authors compiled and examined a data set of all 27 Russian electoral interventions since 1991. They find that Russian efforts have made little difference in the outcomes of these elections.

Highlights

- The authors note two waves of Russian meddling with the first wave occurring in the former Soviet states between 1991 to 2014 and the second wave occurring in Western democracies from 2014 onward.
- The authors note that while nine out of 16 elections swung in Russia's favor since 2015, it's unclear whether that was due to Russian interference or other factors like increased immigration.

Cognitive Science

This section includes sources primarily from psychologists, behavioural scientists, and other cognitive science scholars. The research here tends to focus on the individual and seeks to explain why we believe the things we do, how we process information, and what leads us to discern true from false.

There tends to be two schools of thought in this area of research: 1) that cognitive ability is positively correlated with the ability to discern true from fake information or 2) that having higher cognitive abilities leads to "motivated reasoning," preventing individuals from accepting true but politically dissonant information. There is no conclusive agreement between the two lines of thinking as of yet with studies providing evidence for both hypotheses.

Sources of note:

Why we fall for false information: Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning
On the "illusory truth effect": Prior Exposure Increases Perceived Accuracy of Fake News
On cognitive ability and corrections: 'Fake News': Incorrect, But Hard to Correct. The Role of Cognitive Ability on the Impact of False Information on Social Impressions

The Search Engine Manipulation Effect (SEME) and its Possible Impact on the Outcomes Of Elections

Robert Epstein and Ronald E. Robertson

Epstein, Robert and Ronald E. Robertson. "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 33 (2015): E4512-21. https://doi.org/10.1073/pnas.1419828112.

Crux

Because of the importance and trust humans put on items placed at the top of a list, search engine rankings have considerable sway when it comes to voter preferences and elections. This is compounded by the belief held by most users that search engines will return results best suited to their needs. The authors hypothesize that search engine rankings will therefore influence voter preferences. They test this hypothesis through five double-blind, randomized controlled experiments and find that biased search rankings do have the ability to shift the voting preferences of undecided voters. In their fifth experiment, conducted in India during an election, the authors find that biased search rankings can shift the voting preferences of undecided voters by 20% or more if unaware of the manipulation. The authors term this phenomenon "the search engine manipulation effect" and argue it is more influential than traditional media sources, especially if a region has only one available search engine.

Highlights

- The article cites a study which suggests that flashing "VOTE" ads to 61 million Facebook users caused more than 340,000 people to vote that day who otherwise would not have done so.

- In experiment 3, using only a simple mask, none of the subjects appeared to be aware that they were seeing biased rankings. In their fifth study, only 0.5% of their subjects appeared to notice the bias.
- Consistent with other studies, even those subjects who showed awareness of the biased rankings were still impacted by them in the predicted directions.
- Because search engine rankings are impacted by the popularity of a website, there may be a synergistic effect.
- North American companies now spend more than 20 billion US dollars annually on efforts to place results at the top of rankings

The Cognitive and Emotional Sources of Trump Support: The Case of Low-Information Voters

Richard C. Fording and Sanford F. Schram

Fording, Richard C. and Sanford F. Schram. "The Cognitive and Emotional Sources of Trump Support: The Case of Low-Information Voters." *New Political Science* 39, no. 4 (2017): 670-686. https://doi.org/10.1080/07393148.2017.1378295.

Crux

The authors of this report find that respondents with a lower "Need for Cognition," as measured by whether they agree with the statement "Thinking is not my idea of fun" or "I would rather do something that requires little thought than something that is sure to challenge my thinking abilities" have relatively "warmer" feelings toward Trump. Similarly, by using the respondents' answers on governance as a proxy for their level of political knowledge, the authors find a negative correlation between level of political knowledge and preference for Trump. Note that these answers were taken from the 2016 American National Election Studies Pilot Study (ANES), which sampled around 1,200 individuals chosen to be representative of the United States population. Psychologists and behavioural scientists, on the other hand, may use the CRT (Cognitive Reflection Test), which is a test designed specifically to gauge a person's cognitive ability.

Highlights

- The authors conclude by noting that Trump's campaign exploited the void of reasoning among low-information voters, which made them more vulnerable to relying on emotions, such as fear, anxiety, hate, and rage about Mexican immigrants, Muslim refugees, and Barack Obama.

Rumor Has It: The Adoption of Unverified Information in Conflict Zones

Kelly M. Greenhill and Ben Oppenheim

Greenhill, Kelly M. and Ben Oppenheim. "Rumor Has It: The Adoption of Unverified Information in Conflict Zones." *International Studies Quarterly* 61, no. 3 (2017): 660–76. https://doi.org/10.1093/isq/sqx015.

Crux

This paper presents an analysis of rumour adoption in conflict-affected areas using insights from cognitive science, psychology, and political science. In doing so, the authors offer a theoretical framework for understanding rumour adoption, which identifies three factors that drive an individual to embrace rumours: worldview, threat perception, and repetition. Their findings counter previous beliefs that rumours stem from poverty and lack of education, and instead reinforce more recent research that suggests rumour adoption is driven by security-related anxiety, the rumour's congruence with individuals' pre-existing worldviews, and repeated exposure to a rumour's content.

Highlights

- The spread of rumours is nearly universal before the outbreak of riots and other forms of political violence.
- Rumours enable individuals and groups to function in times of acute stress, reinforce group solidarity, and provide guidance when verifiable facts are hard to come by.

- Rumours can also lead to the dissemination of faulty intelligence to military forces, as happened during the 2015 MSF hospital bombing in Kunduz, Afghanistan.
- Rumours can be categorized as "wish" rumours (hopeful or pleasant rumours) or "dread" rumours (feared or calamitous consequences).
- Dread rumours are more prevalent than wish rumours. The authors find no agreement on why this is the case.

The Collaborative Construction and Evolution of Pseudo-knowledge in Online Conversations

Joshua Introne, Luca Iandoli, Julia DeCook, Irem Gokce Yildirim, Shaima Elzeini

Introne, Joshua, Luca Iandoli, Julia Decook, Irem Gokce Yildirim, and Shaima Elzeini. "The Collaborative Construction and Evolution of Pseudo-Knowledge in Online Conversations." In the *8th International Conference on Social Media & Society*, Toronto, 2017. https://doi.org/10.1145/3097286.3097297.

Crux

In this paper the authors examine how people decide to incorporate or exclude contributions to a body of pseudo-knowledge (PK). In doing so, they examine a popular discussion thread about the existence of alien "stargates" on Earth. They focus on the interplay of narrative and argumentation in this conversation, examining the changes to the PK. They find that a diverse set of users engage in an activity akin to participatory storytelling. Those attempting to debunk their beliefs do little damage but instead help its supporters identify and eliminate less tenable components of the PK. Simultaneously, other contributors enrich the story adding evidence and extending the PK in new ways. The authors suggest that understanding PK evolution is important as it may provide an explanation for why the Internet is able to amplify the growth and spread of conspiracy theories and pseudoscience. Furthermore, they note that because PK responds well to diverse audiences and can rapidly assimilate new information, it can increase the size of its engaged audience.

Highlights

- According to some studies, conspiracy theories proliferate in the face of distressing social events or situations and many researchers have argued that people generate PK to make an inexplicable world understandable.
- Other studies have also demonstrated that conspiracy theories help people come to terms with their own powerlessness.
- In analyzing the type of content in the forum, the authors find there are four main categories:
  - Argumentative - includes evidence, defence of the PK, trolling, or supporting/attacking.
  - Narrative - reflection of PK or established story, piecing different pieces of evidence or PK together, modifications to the story.
  - Contextual - other PK, relation to current events, explaining PK.
  - Discursive - cheering, conversation management, requesting elaboration or evidence.

Third Person Effects of Fake News: Fake News Regulation and Media Literacy Interventions

S. Mo Jang and Joon K. Kim

Jang, S. Mo, and Joon K. Kim. "Third Person Effects of Fake News: Fake News Regulation and Media Literacy Interventions." *Computers in Human Behavior* 80 (2018): 295-302. https://doi.org/10.1016/j.chb.2017.11.034.

Crux

This study explores the beliefs surrounding the effects of fake news, using the theoretical framework of the third-person perception (TPP), which examines whether perceived effects of fake news are greater among other political groups than among themselves or their

supporting political groups. Through an analysis of survey responses (n = 1,299), they find the following:

1) Individuals perceive other political groups as being more influenced by fake news than they are or their own political groups.
2) As perceived undesirability of fake news increases, so does TPP.
3) As partisan identity increases, TPP also increases.
4) As external political efficacy increases, TPP also increases.
5) Contrary to other studies, they found that TPP was positively correlated with support for media literacy and less likely to support media regulation.

Highlights

- The authors note that a potential consequence of TPP and fake news is that if partisans do not believe they are influenced by fake news stories, they may develop a false impression that information shared among them, regardless of their actual accuracy, is perceived to be true.
- The findings also resonated with the *social group identity approach* and the notion of social distance corollary, as the comparisons were made not only between self and other but also between in-group and out-group members.
- Authors note that there may be psychological dispositions other than partisan identity that could influence individuals' levels of third-person perception.
- Authors hypothesize that the preference for media literacy may be based on the idea that if individuals perceive fake news to have effects on others, educating others is more reasonable than regulating everyone's freedom of speech.

Ideology, Motivated Reasoning, and Cognitive Reflection

Dan M. Kahan

Kahan, Dan M. "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making* 8, no. 4 (2013): 407–424. http://dx.doi.org/10.2139/ssrn.2182588.

Crux

This paper finds through observational and experimental data that conservatives did no better or worse than liberals on the Cognitive Reflection Test (CRT) and that ideologically motivated reasoning is not a result of an over-reliance on heuristic or intuitive forms of reasoning. Instead, the author finds that subjects who scored highest on the CRT were most likely to exhibit ideologically motivated cognition.

Highlights

- Scholars have identified three sources of ideological polarization: heuristic-driven information processing, motivated reasoning, and the association between ideological or cultural values and cognitive-reasoning styles.

The Role of Cognitive Ability on the Impact of False Information on Social Impressions.

Jonas De Keersmaecker and Arne Roets

De Keersmaecker, Jonas and Arne Roets. "Fake news": Incorrect But Hard to Correct. The Role of Cognitive Ability on the Impact of False Information on Social Impressions." *Intelligence* 65 (November 2017): 107–110. https://doi.org/10.1016/j.intell.2017.10.005.

Crux

Based on previous studies by other researchers in the field of cognitive science, the authors hypothesize that individuals with lower cognitive abilities would be less likely to change their attitudes when presented with facts that prove their previous beliefs to be wrong. In

other words, simply providing the correct information afterwards is not enough to change someone's incorrect beliefs about an issue or person. In order to test this hypothesis, the authors conducted an experiment in which participants made an initial judgement of an unknown person based on available information, that afterwards were proved to be unequivocally incorrect. They then compared their revised judgements with the judgements of those who did not receive the incorrect information (control group). In line with their hypothesis, they find that individuals with lower cognitive abilities adjusted their judgements to a lesser degree than those with higher cognitive abilities.

Highlights

- Individuals with higher levels of cognitive ability made more appropriate attitude adjustments. After learning that the negative information was false, they adopted attitudes that were similar to those who had not received false information.
- The effects of cognitive ability on attitude adjustment were obtained regardless of whether or not we controlled for open mindedness (i.e., need for closure) and authoritarianism as potential confounding variables.

When Corrections Fail: The Persistence of Political Misperceptions

Brendan Nyhan and Jason Reifler

Nyhan, Brendan and Jason Reifler. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32, no. 2 (2010): 303–30. https://doi.org/10.1007/s11109-010-9112-2.

Crux

This paper reports on the results of two rounds of experiments investigating whether corrective information embedded in realistic news reports reduce prominent misperceptions about contemporary politics. In each of the four experiments, ideological subgroups failed to update their beliefs when presented with corrective information that runs counter to their predispositions. In several cases, the authors find that corrections actually strengthened misperceptions among the most strongly committed subjects.

Highlights

- In the first study, for very liberal subjects, the correction succeeded, making them more likely to disagree with the statement that Iraq had WMD compared with controls.
- For individuals who labeled themselves right of center, the correction backfired and after the correction were more likely to believe that Iraq had WMD than those in the control condition

Misinformation, Disinformation, and Violent Conflict: From Iraq and the "War on Terror" to Future Threats to Peace.

Stephan Lewandowsky, Werner G. K. Stritzke, Alexandra M. Freund, Klaus Oberauer, and Joachim I. Krueger

Lewandowsky, Stephan, Werner G. K. Stritzke, Alexandra M. Freund, Klaus Oberauer, and Joachim I. Krueger. "Misinformation, Disinformation, and Violent Conflict: From Iraq and the "War on Terror" to Future Threats to Peace." *American Psychologist* 68, no. 7 (2013): 487–501. https://doi.org/10.1037/a0034515.

Crux

Using cognitive science and psychology, the authors of this paper explain why misinformation and disinformation have such considerable sway in facilitating the rise of violent conflict, and conversely, how refuting such false claims can contribute to peace. They use two case studies to illustrate their arguments: 1) the 2003 Iraq War and "War on Terror" and 2) climate change and conflict. With regards to the Iraq War, they discuss the importance of narratives, framing, moral disengagement, repetition of frame-consonant information, pluralistic ignorance (the divergence between the prevalence of actual beliefs and what people think others believe), and the "worldview backfire effect," whereby factual but worldview-opposing information can sometimes reinforce beliefs based on initial misinformation. With regards to combating misinformation, the authors discuss three psychological

processes: 1) the role of an individual's skepticism in the discounting of misinformation; 2) the potential role of the media in debiasing; and 3) ways in which the role of worldview as an obstacle to debiasing can be mitigated. Using a similar framework of analysis, the researchers then explore the narratives and framing effects related to climate change and their potential for escalating conflict.

Highlights

- "Skeptics were more likely to endorse true information than people who accepted the WMD-related casus belli (Lewandowsky et al., 2009)."
- The authors place heavy emphasis on framing and the importance of narratives. They argue that understanding narratives is what allows us to understand the information landscape surrounding any conflict. Beyond "spin" or "propaganda," narratives are necessary cognitive tools, which facilitate communication even if they emphasize facts unevenly.

Prior Exposure Increases Perceived Accuracy of Fake News

Gordon Pennycook, Tyrone D. Cannon, and David G. Rand

Pennycook, Gordon, Tyrone Cannon, and David G. Rand. "Prior Exposure Increases Perceived Accuracy of Fake News." *Journal of Experimental Psychology: General* 147, no. 12 (2018): 1865–1880. http://dx.doi.org/10.1037/xge0000465.

Crux

Basing their hypothesis on cognitive science, this paper examines how prior exposure to fake news impacts subsequent perceptions of accuracy. Citing previous studies, it has been shown that exposing an individual to a statement repeatedly increases the likelihood that person will judge it to be accurate. The researchers then test whether this phenomenon, termed the "illusory truth effect," is also applicable to false news on social media. They find that while the effect does exist, it is tempered by the plausibility of the statements being tested and that there is a U-shaped relationship between the plausibility of the statement and the magnitude of the illusory truth effect.

The first study found that there was no significant effect of familiarity on the accuracy of judgements which are obviously false (e.g., Earth is a perfect square). The second study found that exposing participants to fake news headlines increased accuracy ratings, even when the stories were tagged with a warning indicating that they had been disputed by third-party fact checkers. The third study reiterated the findings of the second study even after a week after exposure to the stories.

Highlights

- In Studies 2 and 3, the researchers found that the illusory truth effect was evident for fake news headlines that were both politically discordant and concordant.
- In Study 3, the effect of repetition on perceived accuracy persisted after a week and increased with an additional repetition.

Lazy, Not Biased: Susceptibility to Partisan Fake News is Better Explained by Lack of Reasoning than by Motivated Reasoning

Gordon Pennycook and David G. Rand

Pennycook, Gordon and David G. Rand. "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition* (2018). https://doi.org/10.1016/j.cognition.2018.06.011.

Crux

Using the Cognitive Reflection Test (CRT) as a measure of an individual's propensity to engage in analytical reasoning, the authors find that CRT performance is negatively correlated with the perceived accuracy of fake news, while being positively correlated with the ability to discern fake from real news. These findings hold true even when the headlines did not align with

the individual's political ideology. The authors conclude that analytic thinking is used to assess the plausibility of headlines, regardless of one's political leanings.

Highlights

- Authors found no evidence that analytic thinking exacerbates motivated reasoning.
- In one study, regardless of a participant's political leanings, more analytic individuals were better able to differentiate fake from real news.
- Both Trump and Clinton supporters were more accurate when judging the accuracy of politically-consistent news headlines.
- Participants were recruited through Amazon's Mechanical Turk.

Why do We Fall for Fake News?

S. Shyam Sundar

Sundar, S. Shyam. "Why do We Fall for Fake News?" *The Conversation*, December 7, 2016. http://theconversation.com/why-do-we-fall-for-fake-news-69829.

Crux

This article argues that one of the core reasons fake news spreads so easily is because online news consumption has replaced traditional "gatekeepers" (such as news editors) with our peers via social media. Eventually, the author argues, our friends and peers become the sources of the news itself. Furthermore, how news is presented now is often personalized (i.e., your Facebook news feed), which lowers our guard, allowing false or misleading news to appear more credible.

Highlights

- The author's research shows that Internet users are less skeptical of information that appears in customized environments. Participants who customized their own online news portal tended to agree with statements like "I think the interface is a true representation of who I am" and "I feel the website represents my core personal values."
- Participants in a study conducted by the author who had customized their news portal were less likely to scrutinize the fake news and more likely to believe it.

Mitigation and Solutions

The following sources include proposed solutions to prevent the spread of disinformation or to mitigate the harms associated with disinformation, and their criticisms. The range of solutions includes content moderation, fact-checking, algorithm-based detection mechanisms, legislation, and media literacy. As of yet, there is no clear consensus on what combination of solutions are most effective in curbing the spread of disinformation.

Sources Of note:

The pitfalls of media literacy: You Think You Want Media Literacy… Do You?
A list of state-sponsored solutions: A Guide to Anti-Misinformation Actions Around the World
Possible policy solutions from a Canadian perspective: Democracy Divided: Countering Disinformation and Hate in the Digital Public Sphere
Why warning tags may backfire: The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings

Keeping Up with the Tweet-dashians: The Impact of 'Official' Accounts on Online Rumoring

Cynthia Andrews, Elodie Fichet, Yuwei Ding, Emma S. Spiro, and Kate Starbird

Andrews, Cynthia A, Elodie S Fichet, Yuwei Ding, Emma S Spiro, and Kate Starbird. 2016. "Keeping Up with the Tweet-Dashians: The Impact of ` Official- Accounts on Online Rumoring." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 451–64. https://doi.org/10.1145/2818048.2819986.

Crux

This paper examines the role that mainstream media, new media, and other "official" channels play in the propagation and correction of crisis-related rumours on Twitter. They find that an official source can revitalize conversation and correct misinformation after rumouring has slowed and that it can influence rumouring as it is occurring.

Highlights

- From the two events studied in this paper, official corrections encourage some of those who were involved in rumouring to correct themselves. However, for accounts participating in both affirming and denying behavior, denials received about half the retweets that their affirms had.
- Their findings support previous arguments that a dependable source for information during a crisis is important in shaping the flow of information.
- In one case, 47% of all tweets affirming a rumour were retweets of posts from local affiliates of media conglomerates that may have been considered trusted accounts. Only one corrected itself afterwards. However, in their second case, mainstream media were more influential in denying the rumours.

You Think You Want Media Literacy... Do You?

danah boyd

boyd, danah. "You Think You Want Media Literacy… Do You?" *Points by Data & Society*, March 9, 2018. https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2.

Crux

This article cautions against media literacy as the ultimate solution to disinformation. boyd stresses that critical thinking, which is often cited as a pillar of media literacy, may be "weaponized" and that the spread of disinformation, which is often rooted in trust of news media, cannot be corrected by media literacy education. She also stresses the need to understand epistemological differences between groups of people.

Highlights

- "If we're not careful, "media literacy" and "critical thinking" will simply be deployed as an assertion of authority over epistemology."
- "What does it mean to encourage people to be critical of the media's narratives when they are already predisposed against the news media?"
- "While we have many problems in our media landscape, the most dangerous is how it is being weaponized to gaslight people."

The Promises, Challenges, and Futures of Media Literacy

Monica Bulger and Patrick Davison

Bulger, Monica and Patrick Davison. *The Promises, Challenges, and Futures of Media Literacy.* New York: Data & Society Research Institute, 2018. https://datasociety.net/output/the-promises-challenges-and-futures-of-media-literacy/.

Crux

The authors of this report offer a critique of contemporary media literacy education and programming, arguing that while there have been some positive outcomes, they should not be treated as a "panacea" for fake news. The five recommendations they put forward are: 1) developing a coherent understanding of the current media environment; 2) improve cross-disciplinary collaboration; 3) leverage the current media crisis to consolidate stakeholders; 4) prioritize the creation of a national media literacy evidence base; and 5) develop curricula for addressing user behavior in addition to interpretation.

Highlights

- Studies of media literacy education have shown improvements in critical thinking skills and, in some cases, behaviour change.
- Media literacy suffers from the same issues that plague education generally, such as incoherent expectations and outcomes and incomparable measurements.
- Current studies on media literacy fail to account for differences in socio-economic classes.
- Media literacy places the burden of discerning truth from fiction on the individual, which is becoming increasingly more difficult as media platforms like Facebook and Twitter increase the personalization of information.

Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches

Robyn Caplan

Caplan, Robyn. *Content or Context Moderation? Artisanal, Community-Reliant, and Industrial Approaches.* New York: Data and Society, 2018. https://datasociety.net/output/content-or-context-moderation/.

Crux

Drawing from interviews with representatives at ten major digital platforms, the author identifies three content moderation strategies: artisanal, community-reliant, and industrial. In doing so, the report provides insight into how these content moderation policies are influenced by the companies' missions, business models, and team sizes. She finds that each approach has different trade-offs between moderating content in context and maintaining consistency across a large scale.

Highlights

- The three types of content moderation strategies as defined by the author include:

  - *Artisinal* - Typically conducted on a smaller scale, often manually in-house, and with limited automated filtering. Examples of this type of content moderation include Vimeo and Patreon.
  - *Community-Reliant* - Tends to rely on a large volunteer base to enforce the content moderation policies set up by a small policy team employed by the larger organization. Wikipedia and Reddit employ this type of content moderation strategy.
  - *Industrial* - Global, large-scale operations that primarily depend on automation, operationalization of the rules, and where there is a separation between policy development and enforcement teams. Facebook and Google employ this type of moderation strategy.
- The author highlights the difficulties in maintaining context, which is required to assess whether content is hateful, and as a company grows, "the review of content often happens far outside the context where it is produced."

Dead Reckoning: Navigating Content Moderation After "Fake News"

Robyn Caplan, Lauren Hanson, and Joan Donovan

Caplan, Robyn, Lauren Hanson, and Joan Donovan. *Dead Reckoning: Navigating Content Moderation After "Fake News."* New York: Data and Society, 2018. https://datasociety.net/pubs/oh/DataAndSociety_Dead_Reckoning_2018.pdf.

Crux

Based on a year of field-based research, the focus of this report is on mitigation and accountability of content moderation. The report is broken up into two parts. Part 1 discusses how fake news is used as a political tool and Part 2 covers strategies of intervention. Some of the key findings of the report include:

1) The highly politicized nature of the term "fake news" and its dual use in criticizing mainstream media as well as identifying propaganda
2) Strategies for limiting the spread of "fake news" include trust and verification, disrupting economic incentives, de-prioritizing content and banning accounts, and limited regulatory approaches.
3) Content producers can bypass regulations by claiming satire or parody.
4) Identifying fake news requires knowledge of the source and context, which AI is not advanced enough to handle yet.
5) Third-party fact-checking and media literacy organizations are expected to close the gap between public interest and the platforms but are under resourced.

Highlights

- Because fake news has no concrete universal definition, attempts to define it include three approaches: identifying the intent; classifying "fake news" by type of content; and gathering large lists of identifiable features that can then be used by machine learning and content moderators.
- Warning tags on potential fake news have been tested by both Facebook and Google News. However, Facebook stopped doing so in December 2017.
- Warning tags are costly, resource-heavy, and open to the same critique by hyper-partisan personalities.
- In June 2017, the German parliament approved a bill (the NetzDG law) to limit the spread of hate speech and criminal material over social media, requiring social media platforms to remove these types of posts within 24 hours after receiving a notification or complaint or to block the offending content within seven days. Failure to do so may result in a 50 million Euro fine.

Protecting Democracy from Online Disinformation Requires Better Algorithms, Not Censorship

Eileen Donahoe

Donahoe, Eileen. "Protecting Democracy from Online Disinformation Requires Better Algorithms, Not Censorship." *Council on Foreign Relations*, August 21, 2017. https://www.cfr.org/blog/protecting-democracy-online-disinformation-requires-better-algorithms-not-censorship.

Crux

Donahoe argues that laws like Germany's new NetzDG law, also known as the Network Enforcement Act or social media law, may pave the way for autocratic countries to legitimate censorship and crackdowns on the internet. The NetzDG law, which aims to eradicate hate speech and propaganda on digital platforms, also imposes steep fines (up to €50 million) for failure to take down "evidently criminal" content within twenty-four hours. However, Donahoe points out that this essentially "handed over judicial authority for determining criminality to the private sector." She also argues that the new German law has incentivized platforms to err on the side of taking down flagged content even if not criminal and eroded the core concept of limited platform liability for third-party speech. Shortly after the adoption of the German law, the Russian Duma proposed a copy-cat bill, with multiple explicit references to the German law as its model.

Highlights

- Donahoe claims that the German law hit the wrong target: platforms should not be liable for user-generated content, but rather be accountable for their own algorithms when they push information to users to monetize attention.

A Guide to Anti-Misinformation Actions Around the World

Daniel Funke

Funke, Daniel. "A Guide to Anti-Misinformation Actions Around the World." *Poynter*, May 22, 2018. https://www.poynter.org/news/guide-anti-misinformation-actions-around-world.

Crux

This article serves as a frequently updated guide on global attempts to legislate against what can broadly be referred to as online misinformation, "fake news," and disinformation. Though not every law or regulation listed in this article may relate to false news specifically, they're related to the broader discussion of disinformation.

Highlights

- The authors note the risks to freedom of expression and the lack of effective government proposals.
- Tanzania is proposing to charge bloggers about $920 a year for the privilege of publishing online. The nominal per capita income in Tanzania is less than $900.
- Ugandan President Yoweri Museveni has proposed a "gossip tax," which would charge mobile phone users 200 Ugandan shillings ($0.05) for using platforms like WhatsApp, Viber, Twitter, and Skype. It has been tabled until the government figures out how to implement it.
- Building on an 1881 law criminalizing "false news," French President Macron has proposed allowing the Superior Audiovisual Council, France's media regulator, to fight against "any attempt at destabilization" from TV stations controlled by foreign states — an indirect reference to Russian outlets such as RT.

Democracy Divided: Countering Disinformation and Hate in the Digital Public Sphere

Edward Greenspon and Taylor Owen

Greenspon, Edward and Taylor Owen. *Democracy Divided: Countering Disinformation and Hate in the Digital Public Sphere.* Ottawa: Public Policy Forum, 2018. https://www.ppforum.ca/publications/social-marketing-hate-speech-disinformation-democracy/.

Crux

This report offers an array of policy options to address increasing disinformation and threats to the digital public space. Divided into three parts, it offers a discussion of the forces at play, assumptions and principles underlying any actions, and a catalogue of potential policy options.

Highlights

- The report outlines six assumptions that underlie their policy recommendations:
  - Assumption 1 - Access to trustworthy information is foundational to a well-functioning democracy; direct attacks on it cannot be tolerated.
  - Assumption 2 - Elected representatives have a responsibility to protect the public sphere from disinformation by setting rules, not by serving as regulators.
  - Assumption 3 - Policy is needed; self-regulation is not enough.
  - Assumption 4 - Public policy needs to address this as a supply-side problem, while ensuring the citizenry is better educated on civics and digital risks.
  - Assumption 5 - Governance of technology must be designed to avoid disrupting innovation and restricting freedom of expression.
  - Assumption 6 - Canada should be a leader in ensuring an open and trustworthy Internet.
- The policy options are put into four categories:
  - Rebuilding informational trust and integrity
  - Shoring up Canada's civic infrastructure
  - Keeping information markets open, competitive, and clean
  - Modernizing governance of data rights and opportunities

Could Europe's New Data Protection Regulation Curb Online Disinformation?

Karen Kornbluh

Kornbluh, Karen. "Could Europe's New Data Protection Regulation Curb Online Disinformation?" *Council on Foreign Relations*, February 20, 2018. https://www.cfr.org/blog/could-europes-new-data-protection-regulation-curb-online-disinformation.

Crux

This article discusses how the EU's General Data Protection Regulation (GDPR) could potentially curb online disinformation by sharply limiting the ability to micro-target ad campaigns and content based on an audience's online political views. GDPR would require companies using political or philosophical views to obtain explicit user consent separately for each use. In addition, companies cannot make the use of their service contingent on the user opting in.

Highlights

- Users will be provided clear notice of how their data is being used and can revoke their opt-in consent at any time.
- GDPR also creates a right of "data portability," allowing users to take their data with them to another provider.
- Potential fines for violations may be up to 4% of global revenue or $20 million, whichever is greater, for a serious offence.

Winning the Information Wars: Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe

Edward Lucas and Peter Pomeranzev

Lucas, Edward and Peter Pomeranzev. *Winning the Information Wars.* Washington, DC: Center for European Policy Analysis, 2016. https://cepa.ecms.pl/files/?id_plik=2715.

Crux

This report begins with an overview of information warfare, its historical background, and its current iteration. It then goes over seven case studies looking at Ukraine, Estonia, Latvia, Lithuania, Poland, Czech Republic and Slovakia, and Sputnik International. It concludes on a set of recommendations for how to fight Russia's influence operations in Central and Eastern Europe. The authors break up their recommendations into tactical (short-term, reactive), strategic (medium term, pro-active), and long-term sections. Their tactical level suggestions focus on international coordination and research, empowering civil society, more media regulations, and creating culturally and historically sensitive counter-messaging content addressing Russian speakers. The strategic level suggestions emphasize strengthening public broadcasters, allowing for activist journalism, creating an ethics charter for those engaged in information and activism, and encouraging local media ecosystems to create more Russian-speaking content that can counter Russian disinformation and compete with the "glossier" content coming out of Russia. At the strategic level, the authors recommend improving media literacy and preventing the monetization of media outlets which produce and share hate speech.

Highlights

- Pro-Kremlin content is highly customized, taking advantage of local events and historical grievances.
- Overall, Russia's message is that the US is engaged in a bid for world domination, NATO is hypocritical and unjust, and that Central and Eastern Europe are run by Western puppets who are Russophobic and persecute ethnic Russians.
- Information warfare expert and former NATO spokesman Ben Nimmo characterizes Russia's tactics as "dismissing the critic, distorting the facts, distracting from the main issue and dismaying the audience."
- Russia's information warfare strategy does not seek to directly attack the enemy but allow it to self-destruct through "self-disorganization" and "self-disorientation."
- The Kremlin's "active measures" consist of Influencing the policies of another government; undermining confidence in its leaders and institutions; disrupting its relations with other nations; and discrediting and weakening governmental and nongovernmental opponents.
- Russia's greatest accomplishment is mixing emotion, engagement, and entertainment with their propaganda, making it easily consumable by various niche audiences.

Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society

Paul Mihailidis and Samantha Viotty

Mihailidis, Paul and Samantha Viotty. "Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in "Post-Fact" Society." *American Behavioral Scientist* 61, no. 4 (April 2017): 441–454. https://doi.org/10.1177/0002764217701217.

Crux

Drawing upon the works of critical theorist Guy Debord and critical media scholar Douglas Kellner, the authors explore the phenomenon of spectacle in a ubiquitous digital media culture. They then apply the frame of "spreadable media" to explore how online

citizen expression initiated, sustained, and expanded the media spectacle that pervaded the 2016 U.S. presidential election. In doing so, they ask how spectacle is perpetuated and sustained by spreadable media and how appropriation of content affects civic expression and dialogue online. In answering these questions, the authors articulate the emergence of "spreadable spectacle," which then leads to a critique of proposed media literacy models.

Highlights

- The authors argue that current media literacy programs need to be repositioned in an era of "polarization, distrust, and self-segregation." As such, they suggest media literacy programs include a focus on "being in the world with others towards a common good."
- Debord (1967) explains that in society "the spectacle is not a collection of images; rather, it is a social relationship between people that is mediated by images." In digital culture, networked social relationships emerge not only mediated by images but defined by mediated texts and networked publics: "the imagined communit[ies] that emerges as a result of the intersection of people, technology, and practice," and which benefit from social networks that are designed to support persistence, visibility, spreadability, and searchability (Boyd, 2014).
- In making their recommendations for media literacy programs, the authors make four suggestions:
  - Repositioning media literacies for spreadable connectivity (e.g., focus on connecting humans, embracing differences, etc.)
  - Repositioning media literacies as mechanisms for caring
  - Repositioning media literacies as facilitators of "everyday" engagement
  - Reimagining media literacies as intentionally civic

The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings

Gordon Pennycook and David G. Rand

Pennycook, Gordon and Rand, David G. "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings." Working paper, December 10, 2017. http://dx.doi.org/10.2139/ssrn.3035384.

Crux

This working paper reports on the findings of five experiments conducted on the efficacy of warnings on Facebook news stories as a means to mitigate against political misinformation. In other words, would tagging fake stories with warnings prevent users from labeling them as true? In these tests, warnings were attached to news stories that had been disputed by third-party fact-checkers, in the hopes that readers would read it with a more critical viewpoint. Ultimately, the researchers find that although there is a small reduction in the perceived accuracy of fake news relative to a control, there is an "implied truth" effect, whereby stories with no warnings are assumed to be true. The authors also note that the modest reduction may be limited to these tests only as previous experiments have yielded different results. For example, other researchers have found that warnings may be rendered ineffective due to politically motivated reasoning, whereby people are biased against believing information which challenges their political ideology. Overall, the results of their experiments along with previous work done on the issue pose great challenges for the use of warnings in the fight against misinformation.

Highlights

- Their findings indicate that warnings are more effective for headlines that individuals have a political identity-based motivation to believe, which is inconsistent with accounts of motivated reasoning of fake news under which it is predicted that people should discount information that

contradicts their political ideology.

- A meta-analysis found that while participants 26 years and older showed no significant implied truth effect for fake news (d = .03, z = .84, p = .402) those ages 18-25 showed an implied truth effect for fake news that was much larger than in the overall sample (d = .26, z = 3.58, p < .001).

The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence

Thomas Wood and Ethan Porter

Wood, Thomas and Ethan Porter. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41, no. 1 (2019): 135-163. https://doi.org/10.1007/s11109-018-9443-y.

Crux

This paper addresses the "backfire effect," which occurs when an individual, instead of ignoring or accepting factual information, doubles down on their beliefs. However, the authors find through their own experiments that this did not occur and that people overall heed factual information, even if it challenges their ideological beliefs. They performed five separate studies and enrolled more than 10,100 subjects.

Highlights

- The "backfire effect" has been demonstrated in past studies and appears to have a larger effect on conservatives. However, other studies have found that people can absorb and accept factual information that is contrary to their ideological beliefs.
- "Americans have a well documented facility for avoiding cognitive effort when deducing political attitudes (Sniderman, Brody & Tetlock 1992, Mondak 1993, Mondak 1994)."
- In their first study, respondents were given statements from politicians, then a correction, and then asked to rate the veracity of a statement on the issue. The authors find that "No ideological group exposed to the correction moved in the opposite direction; that is, no group demonstrated backfire."
- One theory as to why the backfire effect is so difficult to induce may be because people tend to shy away from cognitive effort, which counterarguing would require.

Detection

This section highlights recent work on developing detection technology to track and identify disinformation campaigns, accounts, and content. They span a variety of techniques from natural language processing to combined manual and automated means to traits-based metrics. As deep fake technology becomes more prevalent, it is expected that technological solutions for identifying and removing such fakes will become more commonplace.

Sources of note:

Detecting botnets with machine learning: Don't @ Me: Hunting Twitter Bots at Scale
Video verification techniques: Advanced Guide on Verifying Video Content
Detecting automated comments with natural language processing: More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked

Building the 'Truthmeter': Training Algorithms to Help Journalists Assess the Credibility of Social Media Sources

Richard Fletcher, Steve Schifferes, and Neil Thurman

Fletcher, Richard, Steve Schifferes, and Neil Thurman. "Building the 'Truthmeter' Training Algorithms to Help Journalists Assess the Credibility of Social Media Sources." *Convergence* (2017). https://doi.org/10.1177/1354856517714955.

Crux

Due to the proliferation of doctored images, fake news, and misleading information, the authors of this study have created a "Truthmeter" to help journalists assess the credibility of a social media contributor. The Truthmeter is built on a number of open-source Twitter metrics (e.g., number of tweets, number of retweets, verified status, popularity, ratio of followers to followings) that are weighted. After creating the framework and calibrating it, the authors find that it is able to produce scores that were within 1 point of the mean journalist scores when they conducted their benchmarking and evaluation process. In other words, the Truthmeter was able to produce credibility scores for contributors that aligned well with the credibility scores assigned by journalists.

Highlights

- A study conducted by Brandtzaeg et al. in 2016 found that journalists from across Europe deemed elite individuals and institutions (celebrities, politicians, news organizations) as credible, due to the fact that trust had been built up over many years.
- The European Union SocialSensor project, which aims to develop automated software to help journalists determine the credibility of a news item, is based on the 3 Cs:
    1) Contributor: who the information came from;
    2) Content: what is contained within the information;
    3) Context: why the information was provided.

Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures

Jim Maddock, Kate Starbird, Haneen Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason

Maddock, Jim, Kate Starbird, Haneen J. Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. "Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures." *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15* (2015): 228–41. https://doi.org/10.1145/2675133.2675280.

Crux

In this study, the authors examine four rumours that spread via Twitter after the 2013 Boston Marathon Bombings. They establish connections between quantitative measures and the qualitative "story" of rumours, revealing differences among rumour types. In doing so, they construct multi-dimensional signatures (patterns of information flow over time and across other features) that describe a rumour's temporal progression of rumour spreading behavior (within individual tweets), URL domain diversity, domains over time, lexical diversity, and geolocation information.

1) Temporal graphs of tweet volumes over time clearly demonstrate (visually) the concept of a signature
2) Domain volume and diversity are markers of the influence of outside sources on rumour propagation
3) Lexical diversity has been identified as a marker of truthfulness but the authors suggests that it may also provide a view of how a rumour has changed over time, possibly reflecting the number of unique messages and voices that took part in its spread.

Highlights

- Each tweet within each rumour subset was coded into one of seven distinct categories related to the rumour behavior type:
  - Misinformation (relaying rumour as fact)
  - Speculation (develop or support growing rumour)
  - Correction (clearly engate rumour)
  - Question (actively challenges rumour)
  - Hedge (passes along rumour but with doubt)
  - Unrelated, or neutral/other (position unclear or neutral to researcher)
- The four methods used in their analysis include:
  - Qualitative and visual analysis to understand the origins and evolution of each rumour
  - Calculation of lexical diversity (the number of different words that tweets in this corpus use)
  - Analysis of URL propagation and domain diversity
  - Geolocation analysis to draw parallels between event proximity and rumour propagation
- From their study, lexical diversity appears to correlate with different kinds of rumour spreading behavior. Speculation, for example, has higher lexical diversity than misinformation. They find that rumours with low lexical diversity spread without much substantial content variation, while rumours with high lexical diversity tend to be more "conversational" and appear to engage users in "collaborative sensemaking."

More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked

Jeff Kao

Kao, Jeff. "More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked." *Hackernoon*, November 23, 2017. https://hackernoon.com/more-than-a-million-pro-repeal-net-neutrality-comments-were-likely-faked-e9f0e3ed36a6.

Crux

Using natural language processing techniques to analyze net neutrality comments submitted to the FCC from April - October 2017, Kao found that there were at least 1.3 million fake pro-repeal comments. Including the suspicious comments, they estimate that the sum of fake pro-repeal comments in the proceeding may actually number in the millions.
Kao's process involved first identifying unique comments.

They then mapped each comment into semantic space vectors and ran some clustering algorithms on the meaning of the comments. This method identified nearly 150 clusters of comment submission texts of various sizes. Through this process, Kao found that less than 800,000 of the 22M+ comments submitted to the FCC (3-4%) could be considered truly unique. The difficulty then comes from differentiating which of these are legitimate public mailing campaigns, and which of these were bots. Fortunately, the first and largest cluster of pro-repeal documents was especially notable. Using mail-merge tactics, the campaign swapped in a synonym for each term to generate unique-sounding comments. These comments numbered 1.3 million.

Highlights

- One pro-repeal spam campaign used mail-merge to disguise 1.3 million comments as unique grassroots submissions.
- Kao estimates that it is highly likely that more than 99% of the truly unique comments were in favor of keeping net neutrality.
- After taking a 1,000 comment random sample of the 800,000 organic (unique) comments and scanning through them, Kao was only able to find three comments that were clearly pro-repeal.

Disinformation on the Web: Impact, Characteristics, and Detection of Wikipedia Hoaxes

Srijan Kumar, Robert West, and Jure Leskovec

Kumar, Srijan, Robert West, and Jure Leskovec. "Disinformation on the Web." In *Proceedings of the 25th International Conference on World Wide Web - WWW '16*, 591–602. New York: ACM Press, 2016. https://doi.org/10.1145/2872427.2883085.

Crux

By analyzing over 20,000 hoax articles from Wikipedia, the researchers explore the impact, characteristics, and detection of Wikipedia hoaxes. They split their research and resulting paper into four sections: 1) how impactful hoax articles are on Wikipedia; 2) what the typical characteristics of hoaxes are and how they compare to legitimate articles; 3) if it is possible to build a machine-learning program to detect hoax articles; and 4) how good humans are at telling apart hoaxes from legitimate articles in a typical reading situation. Their findings suggest that hoax articles generally have negligible impact and that on average successful hoaxes are nearly twice as long as legitimate articles, but that they look less like typical Wikipedia articles in terms of the templates, infoboxes, and inter-article links they contain. Their classifier program was successful in determining whether an article was a hoax or not 91% of the time and that humans were only able to determine accurately 66% of the time.

Highlights

- 1% of hoaxes are viewed over 100 times per day on average before being uncovered.
- It takes a day to catch 92% of eventually detected hoaxes, a week to catch 94%, a month to catch 96%, and one in a hundred survives for more than a year.

Hoaxy: A Platform for Tracking Online Misinformation

Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer

Shao, Chengcheng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. "Hoaxy: A Platform for Tracking Online Misinformation." In *WWW '16 Companion*, 745–50, 2016. https://arxiv.org/abs/1603.01511.

Crux

The authors of this paper present Hoaxy, a platform for the collection, detection, and analysis of online misinformation and its relation to fact-checking efforts. The paper details the design of the platform and offers a preliminary analysis of a sample of public tweets containing both fake news and fact checking. They find that fact-checking content tends to lag false claims by

10-20 hours and that false claims tend to be dominated by very active users as opposed to fact-checking, which is more grass-roots oriented.

Highlights

- The platform collects data from two main sources:
  - News websites to obtain data about the origin and evolution of both fake news stories and their fact checking
  - Social media to collect instances of these news stories (i.e., URLs) that are being shared online

Advanced Guide on Verifying Video Content

Aric Toler

Toler, Aric. "Advanced Guide on Verifying Video Content." *Bellingcat*, June 30, 2017. https://www.bellingcat.com/resources/how-tos/2017/06/30/advanced-guide-verifying-video-content/.

Crux

This how-to guide discusses how to use Google's reverse image search and Amnesty International's Youtube Data Viewer tool to discern the veracity of a video or image. The author notes that it is a time-consuming and manual process that doesn't always yield clear results.

Highlights

- Link to Amnesty International's YouTube Data Viewer: https://citizenevidence.amnestyusa.org/
- Videos often get recycled and the uploader often forgets to change the thumbnails, which makes reverse image searching thumbnails possible.
- From the author's view: "an arms race between developers and semi-creative fake video creators is a losing battle at this point."

Don't @ Me: Hunting Twitter Bots at Scale

Jordan Wright and Olabode Anise

Wright, Jordan and Olabode Anise. *Don't @ Me: Hunting Twitter Bots at Scale.* Duo Labs, 2018. https://duo.com/assets/pdf/Duo-Labs-Dont-At-Me-Twitter-Bots.pdf.

Crux

Using machine learning and publicly available data from Twitter's API to test a variety of "attributes" to detect bots and botnets, the researchers demonstrate that organized botnets are still active on Twitter and can be discovered. As part of their dataset, they collected 88 million public Twitter accounts, including screen names, tweet counts, followers/following counts, bios, and tweet content. During their analysis they detect a cryptocurrency scam botnet made up of 15,000 bots and identify tactics used by bots to appear credible while avoiding detection.

Highlights

- The authors find that the Random Forest classifier proved to perform the best irrespective of the bot data used for training.
- The cryptocurrency botnet the researchers discovered could be described as a three-level hierarchy, which "consisted of the scam publishing bots, the hub accounts (if any) the bots were following, and the amplification bots that like each created tweet."
- The cryptocurrency spam botnet spoofed legitimate cryptocurrency accounts before transitioning to spoofing celebrity and high-profile accounts.
- Minor edits were done to profile photos to evade detection.
- Spoofed cryptocurrency accounts made use of typos of the spoofed account's name.

Measuring Reach

Related to the section above on detection methods, this section showcases some recent work in measuring the reach of digital disinformation. The following research primarily focuses on quantifiable online metrics (e.g., the number of views or clicks and the density of network clusters) as real world impacts from consuming disinformation is often difficult to assess.

Sources Of note:

Using virology to understand the spread of disinformation: The Biology of Disinformation: Memes, Media Viruses, and Cultural Inoculation
Use of disinformation to prevent protests: On the Influence of Social Bots in Online Protests. Preliminary Findings of a Mexican Case Study
Use of network visualizations to identify group interactions: Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans

The Spreading Of Misinformation Online

Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, and Guido Caldarelli. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences 113*, no. 3 (2016): 554–559. https://doi.org/10.1073/pnas.1517441113.

Crux

This paper, by conducting a massive quantitative analysis on Facebook users' data, shows that the narratives involved in conspiracy theories and scientific news generate homogeneous and polarized communities (i.e., echo chambers). The authors then derive a "data-driven percolation model of rumour spreading" that shows that the main determinants for predicting cascade sizes are homogeneity and polarization.

Highlights

- The authors find that cascade lifetimes for science and conspiracy news exhibit a probability peak in the first two hours, and then in the following hours they rapidly decrease.
- They find that science news reaches a higher level of diffusion more quickly and that a longer lifetime does not correspond to a higher level of interest.
- Conspiracy rumours, however, diffuse slower and exhibit a positive relation between lifetime and size.

Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans

John D. Gallacher, Vlad Barash, Philip N. Howard, and John Kelly

Gallacher, John D., Vlad Barash, Philip N. Howard, and John Kelly. "Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans." Data Memo 2017.9. Oxford, UK: Project on Computational Propaganda. https://comprop.oii.ox.ac.uk/research/working-papers/vetops/.

Crux

This article highlights the various ways in which social

media disinformation campaigns are propagated among US military personnel and veterans. By conducting and analyzing selected keywords, seed accounts, and known links to content, the researchers were able to construct large network visualizations and found that on Twitter there are significant and persistent interactions between current and former military personnel and a broad network of Russia-focused accounts, conspiracy theory focused accounts, and European right-wing accounts. These interactions are often mediated by pro-Trump users and accounts that identify with far-right political movements in the US. Similar interactions are also found on Facebook.

Highlights

- The authors' definition of junk news includes various forms of propaganda and ideologically extreme, hyperpartisan, or conspiratorial political news and information. They note that much of this content is deliberately produced false reporting.
- The authors looked specifically at three "junk news" websites specializing in content on military affairs and national security issues for US military personnel and veterans: veteranstoday.com, veteransnewsnow.com, and southfront.org.

Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 U.S. Presidential Campaign

Andrew Guess, Brendan Nyhan, and Jason Reifler

Guess, Andrew, Brendan Nyhan, and Jason Reifler. *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 U. S. Presidential Campaign.* (No prelo). January 9, 2018. https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf.

Crux

This report examines the prevalence and mechanisms of exposure to fake news sites by combining pre-election survey responses and web traffic data from a national sample of Americans. Using this dataset, the authors find that approximately 1 in 4 Americans visited a fake news website, but consumption was disproportionately observed among Trump supporters. However, the authors note that this pattern of selective exposure was concentrated among a very small subset of people - almost 6 in 10 visits to fake news sites came from the 10% of Americans with "the most conservative information diets." They also find that Facebook was the most influential mechanism in facilitating the spread of fake news and that fact-checking largely failed to effectively reach the consumers of fake news.

Highlights

- The authors estimate that 27.4% of Americans age 18 or older visited an article on a pro-Trump or pro-Clinton fake news website during the final weeks of the 2016 election campaign.
- Articles on pro-Trump or pro-Clinton fake news websites represented an average of approximately 2.6% of all the articles Americans read on sites that focused on hard news topics during the final weeks of the 2016 election campaign.
- People saw a mean average of 5.45 articles from fake news websites during the study period of October 7–November 14, 2016. Nearly all of these were pro-Trump (average of 5.00 pro-Trump articles).
- Among Trump supporters, 40% read at least one article from a pro-Trump fake news website versus 15% of Clinton supporters.

The Biology of Disinformation: Memes, Media Viruses, and Cultural Inoculation

Douglas Rushkoff, David Pescovitz, and Jake Dunagan.

Rushkoff, Douglas, David Pescovitz, and Jake Dunagan. *The Biology of Disinformation: Memes, Media Viruses, and Cultural Inoculation.* Institute for the Future, 2018. http://www.iftf.org/partner-with-iftf/research-labs/digital-intelligence-lab/biology-of-disinformation/.

Crux

This report takes a biological approach at understanding

the phenomenon of contemporary disinformation. Using three biological metaphors to "examine the biology of computational" propaganda, the authors focus on selection strategies, contagion/inoculation, and symbiosis/endogenous retroviruses as a means to explain the spread of disinformation.

Highlights

- "Meme replication seems to be a game of volume and brute force, not specificity of intent and design."
- Researchers have used epidemiology to predict the reach of disinformation using the SIR model (S=number of susceptible persons; I=number of infected persons; R=number of recovered persons) to estimate the reach of certain types of online viral content.
- The authors discuss the interrelated factors that contribute to the adoption and spread of memes, such as context, content, emotional effect, education, and gender.
- Authors recommend "inoculation" strategies to prevent the harmful effects of disinformation. In other words, to "strengthen the cultural immune response of the society under attack."

Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter

Kate Starbird

Starbird, Kate. 2017. "Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter." In E*leventh International AAAI Conference on Web and Social Media*: 230–39. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15603.

Crux

This paper offers a systematic lens for exploring the production of "alternative narratives of man-made crisis events" on Twitter, demonstrating how alternative news sites propagate and shape alternative narratives, while mainstream media deny them. Their research methods include mapping a domain network based on users' tweets, qualitative content analysis and coding of the domains in the dataset, and finally an interpretive analysis to identify patterns, connections, and anomalies in relation to the network graph.

Highlights

- The initial data collection collected tweets that mentioned the following words: shooter, shooting, gunman, gunmen, gunshot, gunshots, shooters, gun shot, gun shots, shootings. The author then narrowed this data set down to Tweets that included keywords indicating an alternative narrative, such as: false flag, falseflag, crisis actor, crisisactor, staged, hoax and "1488."
- Each account (domain) was classified across the following dimensions:
  - *Account Type* (Mainstream Media, Alternative Media or Blog, Government Media, or Other)
  - *Narrative Stance Coding* (coded as supporting the alternative narrative, denying it, or for primarily being cited as evidence of the alternative narrative without directly referring to it. Domains that did not fall into one of these categories were coded as unrelated.
  - *Primary Orientation* (Traditional News, Clickbait News, Primarily Conspiracy Theorists/ Pseudo-Science Evangelists, and sites with a strong Political Agenda)
  - *Political Leaning* (US alt-right; US alt-left; international anti-globalist; White Nationalist and/or anti-semitic ; Muslim defense; Russian propaganda)
- The two most tweeted domains are associated with significant "bot" activity.

Examining Trolls and Polarization with a Retweet Network

Leo G. Stewart, Ahmer Arif, and Kate Starbird

Stewart, Leo G., Ahmer Arif, and Kate Starbird. 2018. "Examining Trolls and Polarization with a Retweet Network." https://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf.

Crux

This paper examines Russian trolls' contributions to Twitter, specifically how they took advantage of the polarized nature of the news coverage and events surrounding #BlackLivesMatter and the context of domestic debates over gun violence and race relations. Their analysis is based on a "retweet network and a community detection algorithm." They find two main clusters emerge, containing 48.5% and 43.2% of the nodes. The clusters are unsurprisingly classified as "left-leaning" and "right-leaning", evidence of "echo chambers" or "filter bubbles". In the left-leaning cluster are 22 RU-IRA (Russian-linked Internet Research Agency) accounts. In the right-leaning cluster, 7 RU-IRA troll accounts are reported. On both sides trolls were in the top percentiles by number of retweets.

Highlights

- Their initial data collection efforts looked at tweets that contained words regarding gun violence and shooting; this was then narrowed down to include only tweets that contained at least one of the terms "BlackLivesMatter", "BlueLivesMatter", or "AllLivesMatter".
    - A retweet graph was then constructed where each node represents a Twitter account and edges between nodes represent retweets.

On the Influence of Social Bots in Online Protests. Preliminary Findings of a Mexican Case Study

Pablo Suárez-Serrato, Margaret E. Roberts, Clayton A. Davis, and Filippo Menczer

Suárez-Serrato Pablo, Margaret E. Roberts, Clayton A. Davis, and Filippo Menczer. "On the Influence of Social Bots in Online Protests." In *Social Informatics: 8th International Conference*, *SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part II*, edited by Emma Spiro and Yong-Yeol Ahn. Springer, 2016. https://doi.org/10.1007/978-3-319-47874-6_19.

Crux

Following a press conference on the investigation into the disappearance of 43 teachers in training from a school in Ayotzinapa, Guerrero on September 26th, 2014, Mexican Federal District Attorney, José Murillo Karam said, "Ya me cansé" (I am tired), which soon became a trending hashtag on Twitter, #YaMeCanse. The hashtag was used some 2.2million times in the month following the press conference but it was claimed that spam accounts had repeatedly used this hashtag to make it more difficult to find real information on the protest movement. Real users of the hashtag soon moved on to #YaMeCanse2, then #YaMeCanse3, going as high as #YaMeCanse25. The researchers, using the program *BotOrNot* found that indeed, the YaMeCanse protest movement had been disrupted by bots, making communication through Twitter more difficult.

Highlights

- Sybil accounts attempt to disguise themselves as humans. Cyborg accounts mix automation and human intervention. For example they can be programmed to post at certain intervals.
- Fewer than 2,000 accounts in a collection of 300,000 users (less than 1%) generated up to 32% of Twitter traffic about Brexit.

Additional resources

- The Media Manipulation Team at Data & Society Research Institute
- Institute for the Future's Digital Intelligence Lab
- Eu vs Disinfo - run by the European External Action Service East Stratcom Task Force
- The Computational Propaganda Research Project (COMPROP) at Oxford
- The Digital Forensics Research Lab (DFRLab) at the Atlantic Council
- Social Media And Political Participation Lab (SMaPP Lab) at NYU
- Emerging Capacities of Mass Participation Lab (emCOMP Lab) at the University of Washington
- RAND Corporation - Information Operations and Social Media Analysis
- Observatory on Social Media at Indiana University
- First Draft
- Duke Reporters' Lab (maintains a global list of fact-checking initiatives)
- Guide to anti-misinformation actions around the world - Poynter
- European Commission's fake news and online disinformation policies

Bibliography

Alba, Davey. "How Duterte Used Facebook To Fuel the Philippine Drug War." *Buzzfeed News*, September 4, 2018. https://www.buzzfeednews.com/article/daveyalba/facebook-philippines-dutertes-drug-war.

Allcott, Hunt and Matthew Gentzkow. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (2017): 211–36. https://doi.org/10.1257/jep.31.2.211.

Anderson, Janna and Lee Rainie. *The Future of Truth and Misinformation Online.* Pew Research Center, 2017. http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online.

Andrews, Cynthia A., Elodie S Fichet, Yuwei Ding, Emma S. Spiro, and Kate Starbird. 2016. "Keeping Up with the Tweet-Dashians: The Impact of `Official- Accounts on Online Rumoring." In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*, 451–64. https://doi.org/10.1145/2818048.2819986.

Bakir, Vian and Andrew McStay. "Fake News and The Economy of Emotions: Problems, causes, solutions." *Digital Journalism* 6, no. 2 (2017): 154-175. https://doi.org/10.1080/21670811.2017.1345645.

Blank, Stephen. "Russian Information Warfare As Domestic Counterinsurgency." *American Foreign Policy Interests* 35, no. 1 (2013): 31–44. https://doi.org/10.1080/10803920.2013.757946.

boyd, danah. "You Think You Want Media Literacy… Do You?" *Points by Data & Society*, March 9, 2018. https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2.

Bradshaw, Samantha and Philip N. Howard. "Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation." Working Paper 2018.1. Oxford, UK: Project on Computational Propaganda. http://comprop.oii.ox.ac.uk/research/cybertroops2018/.

Bradshaw, Samantha and Philip N. Howard. "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation." Samuel Woolley and Philip N. Howard, Eds. Working Paper 2017.12. Oxford, UK: Project on Computational Propaganda, 2017. https://comprop.oii.ox.ac.uk/research/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/.

Bulger, Monica and Patrick Davison. *The Promises, Challenges, and Futures of Media Literacy.* New York: Data & Society Research Institute, 2018. https://datasociety.net/output/the-promises-challenges-and-futures-of-media-literacy/.

Canadian Security Intelligence Service. *WHO SAID WHAT? The Security Challenges of Modern Disinformation.* Ottawa: CSIS, 2018. https://www.canada.ca/content/dam/csis-scrs/documents/publications/disinformation_post-report_eng.pdf.

Caplan, Robyn, Lauren Hanson, and Joan Donovan. *Dead Reckoning: Navigating Content Moderation After "Fake News."* New York: Data and Society, 2018. https://datasociety.net/pubs/oh/DataAndSociety_Dead_Reckoning_2018.pdf.

Carlson, Matt. "When News Sites go Native: Redefining the Advertising-Editorial Divide in Response to Native Advertising." *Journalism* 16, no. 7 (October 2015): 849–65. https://doi.org/10.1177/1464884914545441.

Carroll, David. "Awkward Conversation With Facebook: What Happened When I Caught Them Defaulting Us Back Into Behavorial Ad Tracking and Targeting." *Medium*, June 1, 2016. https://medium.com/@profcarroll/awkward-conversation-with-facebook-ef1734ecdc62.

Chen, Adrian. "The Agency." *The New York Times*, June 2, 2015. https://www.nytimes.com/2015/06/07/magazine/the-agency.html.

Chesney, Robert and Citron, Danielle. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." *California Law Review* 107 (2019, Forthcoming). https://ssrn.com/abstract=3213954.

Crilley, Rhys. "International Relations in the Age of 'Post-Truth' Politics." *International Affairs* 94, no. 2 (2018): 417–425. https://doi.org/10.1093/ia/iiy038.

De Keersmaecker, Jonas and Arne Roets. "Fake News": Incorrect but Hard to Correct. The Role of Cognitive Ability on the Impact of False Information on Social Impressions." *Intelligence* 65 (November 2017): 107–110. Retrieved from https://doi.org/10.1016/j.intell.2017.10.005.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, and Guido Caldarelli. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113, no. 3 (2016): 554–559. https://doi.org/10.1073/pnas.1517441113.

Donahoe, Eileen. "Protecting Democracy from Online Disinformation Requires Better Algorithms, Not Censorship." *Council on Foreign Relations*, August 21, 2017. https://www.cfr.org/blog/protecting-democracy-online-disinformation-requires-better-algorithms-not-censorship.

Einstein, Mara. *Black Ops Advertising: Native Ads, Content Marketing and the Covert World of the Digital Sell.* New York: OR Books, 2016.

Epstein, Robert and Ronald E. Robertson. "The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 33 (2015): E4512-21. https://doi.org/10.1073/pnas.1419828112.

Farkas, Johan, Jannick Schou, and Christina Neumayer. "Cloaked Facebook Pages: Exploring fake Islamist Propaganda in Social Media." *New Media & Society* 20, no. 5 (May 2018): 1850–67. https://doi.org/10.1177/1461444817707759.

Farrell, Henry and Bruce Schneier. "Common-Knowledge Attacks on Democracy." Berkman Klein Center Research Publication No. 2018-7 (October 2018). http://dx.doi.org/10.2139/ssrn.3273111.

Fletcher, Richard, Steve Schifferes, and Neil Thurman. "Building the 'Truthmeter' Training algorithms to Help Journalists Assess the Credibility of Social Media Sources." *Convergence* (2017). https://doi.org/10.1177/1354856517714955.

Fording, Richard C. and Sanford F. Schram. "The Cognitive and Emotional Sources of Trump Support: The Case of Low-Information Voters." *New Political Science* 39, no. 4 (2017): 670-686. https://doi.org/10.1080/07393148.2017.1378295.

Friggeri, Adrien, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. "Rumor Cascades." In *Eighth International AAAI Conference on Weblogs and Social Media*, 2014. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8122/8110.

Funke, Daniel. "A Guide to Anti-Misinformation Actions Around the World." *Poynter*, May 22, 2018. https://www.poynter.org/news/guide-anti-misinformation-actions-around-world.

Gallacher, John D., Vlad Barash, Philip N. Howard, and John Kelly. "Junk News on Military Affairs and National Security: Social Media Disinformation Campaigns Against US Military Personnel and Veterans." Data Memo 2017.9. Oxford, UK: Project on Computational Propaganda. https://comprop.oii.ox.ac.uk/research/working-papers/vetops/.

Ghosh, Dipayan and Ben Scott. *#DigitalDeceit:The Technologies Behind Precision Propaganda on the Internet.* New America, 2018. https://www.newamerica.org/public-interest-technology/policy-papers/digitaldeceit/.

Golovchenko, Yevgeniy, Mareike Hartmann, and Rebecca Adler-Nissen. "State, Media and Civil Society in the Information Warfare Over Ukraine: Citizen Curators of Digital Disinformation." *International Affairs* 94, no. 5 (September 2018): 975–994. https://doi.org/10.1093/ia/iiy148.

Gomez, Henry J. "The Man Who Made The Republican Internet — And Then Sold It To Far-Right Nationalists Overseas." *Buzzfeed*, December 19, 2017. https://www.buzzfeed.com/henrygomez/the-man-who-made-the-republican-internet-and-then-sold-it.

Greenhill, Kelly M. and Ben Oppenheim. "Rumor Has It: The Adoption of Unverified Information in Conflict Zones." *International Studies Quarterly* 61, no. 3 (2017): 660–76. https://doi.org/10.1093/isq/sqx015.

Greenspon, Edward and Taylor Owen. *Democracy Divided: Countering Disinformation and Hate in the Digital Public Sphere.* Ottawa: Public Policy Forum, 2018. https://www.ppforum.ca/publications/social-marketing-hate-speech-disinformation-democracy/.

Gu, Lion, Vladimir Kropotov and Fyodor Yarochkin. *The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public.* Trend Micro, 2017. https://documents.trendmicro.com/assets/white_papers/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. *Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 U. S. Presidential Campaign. (No prelo)*. January 9, 2018. https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf.

Hellman, Maria and Charlotte Wagnsson. "How Can European States Respond to Russian Information Warfare? An Analytical Framework." *European Security* 26, no. 2 (2017): 153–170. https://doi.org/10.1080/09662839.2017.1294162.

Introne, Joshua, Luca Iandoli, Julia Decook, Irem Gokce Yildirim, and Shaima Elzeini. "The Collaborative Construction and Evolution of Pseudo-Knowledge in Online Conversations." In the *8th International Conference on Social Media & Society*, Toronto, 2017. https://doi.org/10.1145/3097286.3097297.

Jack, Caroline. *Lexicon of Lies: Terms for Problematic Information.* New York: Data & Society Research Institute, 2017. https://datasociety.net/output/lexicon-of-lies/.

Jang, S. Mo, and Joon K. Kim. "Third Person Effects of Fake News: Fake News Regulation and Media Literacy Interventions." *Computers in Human Behavior* 80 (2018): 295-302. https://doi.org/10.1016/j.chb.2017.11.034.

Kahan, Dan M. "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making* 8, no. 4 (2013): 407–424. http://dx.doi.org/10.2139/ssrn.2182588.

Kao, Jeff. "More than a Million Pro-Repeal Net Neutrality Comments were Likely Faked." *Hackernoon*, November 23, 2017. https://hackernoon.com/more-than-a-million-pro-repeal-net-neutrality-comments-were-likely-faked-e9f0e3ed36a6.

Kornbluh, Karen. "Could Europe's New Data Protection Regulation Curb Online Disinformation?" *Council on Foreign Relations*, February 20, 2018. https://www.cfr.org/blog/could-europes-new-data-protection-regulation-curb-online-disinformation.

Kumar, Srijan, Robert West, and Jure Leskovec. “Disinformation on the Web.” In *Proceedings of the 25th International Conference on World Wide Web - WWW ’16*, 591–602. New York: ACM Press, 2016. https://doi.org/10.1145/2872427.2883085.

Lapowsky, Issie. “Parkland Conspiracies Overwhelm The Internet’s Broken Trending Tools.” *Wired*, February 21, 2018. https://www.wired.com/story/youtube-facebook-trending-tools-parkland-conspiracy.

Lewandowsky, Stephan, Werner G. K. Stritzke, Alexandra M. Freund, Klaus Oberauer, and Joachim I. Krueger. “Misinformation, Disinformation, and Violent Conflict: From Iraq and the “War on Terror” to Future Threats to Peace.” *American Psychologist* 68, no. 7 (2013): 487–501. https://doi.org/10.1037/a0034515.

Lewis, Paul. ““Fiction Is Outperforming Reality”: How YouTube’s Algorithm Distorts Truth.” *The Guardian*, February 2, 2018. https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth.

Lin, Herbert and Jaclyn Kerr. “On Cyber-Enabled Information/Influence Warfare and Manipulation.” In the *Oxford Handbook of Cybersecurity*. Oxford University Press: 2018 forthcoming. https://ssrn.com/abstract=3015680.

Lucas, Edward and Peter Pomeranzev. *Winning the Information Wars.* Washington, DC: Center for European Policy Analysis, 2016. https://cepa.ecms.pl/files/?id_plik=2715.

Maddock, Jim, Kate Starbird, Haneen J. Al-Hassani, Daniel E. Sandoval, Mania Orand, and Robert M. Mason. “Characterizing Online Rumoring Behavior Using Multi-Dimensional Signatures.” *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW ’15* (2015): 228–41. https://doi.org/10.1145/2675133.2675280.

Marwick, Alice and Rebecca Lewis. *Media Manipulation and Disinformation Online.* New York: Data & Society Research Institute, 2017. https://datasociety.net/output/media-manipulation-and-disinfo-online/.

Matz, S. C., M. Kosinski, G. Nave, and D. J. Stillwell. “Psychological Targeting as an Effective Approach to Digital Mass Persuasion.” *PNAS* 114, no. 48 (2017): 12714–19. https://doi.org/10.1073/pnas.1710966114.

Mihailidis, Paul and Samantha Viotty. “Spreadable Spectacle in Digital Culture: Civic Expression, Fake News, and the Role of Media Literacies in “Post-Fact” Society.” *American Behavioral Scientist* 61, no. 4 (April 2017): 441–454. https://doi.org/10.1177/0002764217701217.

Misiewicz, Michael and Laura Yu. *Viral Content: The First Warning Sign of Fraud.* Appnexus, 2017. https://www.appnexus.com/sites/default/files/whitepapers/appnexus_viralcontentwhitepaper_december2017.pdf.

Monaco, Nicholas J. "Computational Propaganda in Taiwan: Where Digital Democracy Meets Automated Autocracy No. 2017.2." Project on Computational Propaganda Working Paper No. 2017.2, Oxford, UK, 2017. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Taiwan-2.pdf.

Nadler, Anthony, Matthew Crain, and Joan Donovan. *Weaponizing the Digital Influence Machine: The Political Perils of Online Ad Tech.* New York: Data & Society Research Institute, 2018. https://datasociety.net/output/weaponizing-the-digital-influence-machine/.

Narayanan, Vidya, Vlad Barash, Bence Kollanyi, Lisa-Maria Neudert, and Philip N. Howard. *Polarization, Partisanship and Junk News Consumption over Social Media in the US* (Vol. 1). Oxford, UK: Project on Computational Propaganda, 2018. http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/02/Polarization-Partisanship-JunkNews.pdf.

Neudert, Lisa-Marie. "Future Elections May Be Swayed by Intelligent, Weaponized Chatbots." *MIT Technology Review*, August 22, 2018. https://www.technologyreview.com/s/611832/future-elections-may-be-swayed-by-intelligent-weaponized-chatbots/.

Nye, Joseph. "Protecting Democracy in an Era of Cyber Information War." *Governance in an Emerging World*, iss. 318 (November 2018). https://www.hoover.org/research/protecting-democracy-era-cyber-information-war.

Nyhan, Brendan and Jason Reifler. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32, no. 2 (2010): 303–30. https://doi.org/10.1007/s11109-010-9112-2.

Nyst, Carly and Nicholas Monaco. Edited by Samuel C. Wooley. *State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns.* Palo Alto: Institute for the Future, 2018. http://www.iftf.org/statesponsoredtrolling.

Oates, Sarah. 2018. "When Media Worlds Collide: Using Media Model Theory to Understand How Russia Spreads Disinformation in the United States." Paper presented at the 2018 *American Political Science Association Annual Meeting*, Boston, MA, September 2018. http://dx.doi.org/10.2139/ssrn.3238247.

Ong, Jonathan Corpus and Jason Vincent A. Cabañes. *Architects of Networked Disinformation: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines.* The Newton Tech 4 Dev Network, 2018. http://newtontechfordev.com/wp-content/uploads/2018/02/ARCHITECTS-OF-NETWORKED-DISINFORMATION-FULL-REPORT.pdf.

Pament, James, Howard Nothhaft, Henrik Agardh-Twetman, and Alicia Fjällhed. *Countering Information Influence Activities*, Version 1.4. Department of Strategic Communication, Lund University, 2018. https://www.msb.se/RibData/Filer/pdf/28697.pdf.

Pennycook, Gordon and Rand, David G. "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings." Working paper, December 10, 2017. http://dx.doi.org/10.2139/ssrn.3035384.

Pennycook, Gordon and David G. Rand. “Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning.” *Cognition* (2018). https://doi.org/10.1016/j.cognition.2018.06.011.

Pennycook, Gordon, Tyrone Cannon, and David G. Rand. “Prior Exposure Increases Perceived Accuracy of Fake News.” *Journal of Experimental Psychology: General* 147, no. 12 (2018): 1865–1880. http://dx.doi.org/10.1037/xge0000465.

Phillips, Whitney. *The Oxygen of Amplification.* Data and Society Research Institute, 2018. https://datasociety.net/wp-content/uploads/2018/05/FULLREPORT_Oxygen_of_Amplification_DS.pdf.

Posetti, Julie and Alice Matthews. *A Short Guide To The History Of ’Fake News’ And Disinformation.* International Centre for Journalists, 2018. https://www.icfj.org/news/short-guide-history-fake-news-and-disinformation-new-icfj-learning-module.

Prier, Jarred. “Commanding the Trend: Social Media as Information Warfare.” *Strategic Studies Quarterly* 11(Winter 2017): 50–85. http://www.airuniversity.af.mil/Portals/10/SSQ/documents/Volume-11_Issue-4/Prier.pdf.

Rushkoff, Douglas, David Pescovitz, and Jake Dunagan. *The Biology of Disinformation: memes, Media Viruses, and Cultural Inoculation.* Institute for the Future, 2018. http://www.iftf.org/partner-with-iftf/research-labs/digital-intelligence-lab/biology-of-disinformation/.

Schreckinger, Ben. “How Russia Targets the U.S. Military.” *POLITICO Magazine*, June 12, 2017. http://www.politico.com/magazine/story/2017/06/12/how-russia-targets-the-us-military-215247.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. “Hoaxy: A Platform for Tracking Online Misinformation.” In *WWW ’16 Companion*, 745–50, 2016. https://arxiv.org/abs/1603.01511.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kaicheng Yang, Alessandro Flammini, and Filippo Menczer. “The Spread of Low-Credibility Content by Social Bots.” *Nature Communications* 9 (2018). https://www.nature.com/articles/s41467-018-06930-7.

Share Lab. “Mapping and Quantifying Political Information Warfare.” *Share Lab.* October 26, 2016. https://labs.rs/en/mapping-and-quantifying-political-information-warfare/.

Silverman, Craig. *Lies, Damn Lies and Viral Content.* New York: Tow Center for Digital Journalism, 2015. https://doi.org/10.7916/D8Q81RHH.

Starbird, Kate. 2017. “Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter.” In *Eleventh International AAAI Conference on Web and Social Media*: 230–39. https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15603.

Stewart, Leo G., Ahmer Arif, and Kate Starbird. 2018. “Examining Trolls and Polarization with a Retweet Network.” https://faculty.washington.edu/kstarbi/examining-trolls-polarization.pdf.

Suárez-Serrato, Pablo, Margaret E. Roberts, Clayton A. Davis, and Filippo Menczer. “On the Influence of Social Bots in Online Protests.” In *Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part II*, edited by Emma Spiro and Yong-Yeol Ahn. Springer, 2016. https://doi.org/10.1007/978-3-319-47874-6_19.

Sundar, S. Shyam. “Why do We Fall for Fake News?” *The Conversation*, December 7, 2016. http://theconversation.com/why-do-we-fall-for-fake-news-69829.

Tenove, Chris, Jordan Buffie, Spencer McKay, David Moscrop, Mark Warren, Maxwell A. Cameron. *Digital Threats To Democratic Elections: How Foreign Actors Use Digital Techniques.* Vancouver, BC: Centre for the Study of Democratic Institutions, UBC, 2018. https://democracy.arts.ubc.ca/2018/01/18/digital-threats/.

Toler, Aric. “Advanced Guide on Verifying Video Content.” *Bellingcat*, June 30, 2017. https://www.bellingcat.com/resources/how-tos/2017/06/30/advanced-guide-verifying-video-content/.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. “The Spread of True and False News Online.” *Science* 359, iss. 6380 (2018): 1146–1151. https://doi.org/10.1126/science.aap9559.

Wakabayashi, Daisuke and Linda Qiu. “Google Serves Fake News Ads in an Unlikely Place: Fact-Checking Sites.” *The New York Times*, October 17, 2017. https://www.nytimes.com/2017/10/17/technology/google-fake-ads-fact-check.html.

Wardle, Claire and Hossein Derakhshan. Information Disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe,* 2017. https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html.

Way, Lucan Ahmad and Adam Casey. “Russia Has Been Meddling in Foreign Elections for Decades. Has It Made a Difference?” *The Washington Post*, January 8, 2018. https://www.washingtonpost.com/news/monkey-cage/wp/2018/01/05/russia-has-been-meddling-in-foreign-elections-for-decades-has-it-made-a-difference/.

Witness and First Draft. *Mal-uses of AI-generated Synthetic Media and Deepfakes: Pragmatic Solutions Discovery Convening.* July, 2018. http://witness.mediafire.com/file/q5juw7dc3a2w8p7/Deepfakes_Final.pdf/file.

Wood, Thomas and Ethan Porter. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41, no. 1 (2019): 135-163. https://doi.org/10.1007/s11109-018-9443-y.

Wright, Jordan and Olabode Anise. *Don't @ Me: Hunting Twitter Bots at Scale.* Duo Labs, 2018. https://duo.com/assets/pdf/Duo-Labs-Dont-At-Me-Twitter-Bots.pdf.

Zuiderveen Borgesius, Frederik J., Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balázs Bodó, and Claes de Vreese. "Online Political Microtargeting: Promises and Threats for Democracy." *Utrecht Law Review* 14, no. 1 (2018): 82-96. http://doi.org/10.18352/ulr.420.