UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT

Using the Flow of Information to Detect False News

Ricardo Muñoz Sánchez

Fake News Detection

Can we automatically detect disinformation in news articles?

What are the applications for fake news detection?

How do we go about it?

Outline

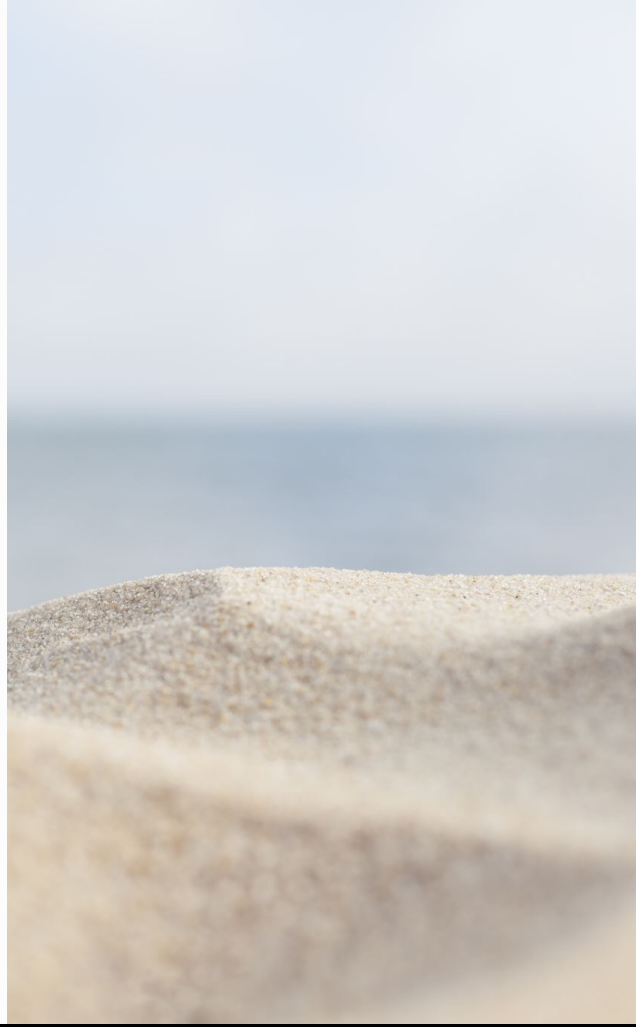
- Mis- and Disinformation
  - What are Fake News?
  - Why is this a relevant problem?
- Fake News and NLP
  - NLP tasks related to fake news
  - Approaches to these tasks
- Ideas and Future Research
  - What I have done so far
  - What I'm currently working on

PART 1
What About Fake News?

And Why Should We Care?

Fake News in the World

- The term “fake news” has become a buzzword at this point
- It is important to acknowledge their real-world consequences
- Understanding this phenomenon is the first step before we can stop it



Fake News in the World – Politics

- The Brexit campaign
- 2016 and 2020 presidential elections in the United States
- Myanmar genocide

Image: 2022 REUTERS/Leah Millis [link]

Fake News in the World – Healthcare

- AIDS and COVID-19 Pandemic
- Polio being reintroduced to several parts of the world
- Smoking disinformation campaigns

Image: REUTERS/Susana Vera [link]

Fake News in the World – Environment

- Disinformation campaigns from oil companies (and others)
- The idea of “wildfire seasons” has been introduced recently
- In general, distract and delay

Image: REUTERS/David Swanson [link]

From “*On Fire*” by KC Green [link]

What are “Fake News”?

- The term “fake news” is not well defined!

- It has been used as:
  - A general term for disinformation
  - A term for intentionally false news
  - A way to disqualify journalistic outlets

Misinformation – False information that is spread, regardless of intent.

Disinformation – False information spread with the intent to deceive or to manipulate.

Fake news åre news articles that are intentionally and verifiably false, and could mislead readers

(Allcott and Gentzkow, 2017)

The Problem with Intent

- Most of these definitions hinge on intent
- However, intent is hard (if not impossible) to establish
- This complicates gathering data in a reliable and consistent manner

Related (but Distinct) Terms

Rumours	Clickbait	Propaganda
Satirical News	Hyperpartisan News	Biased News

Ethical Concerns

- Where do we draw the line between policing and censorship?
- Who is telling us what is true and what is false?
- Can we *really* detect falsehood just through text?

PART 2
Fake News and NLP

How Do We Use AI to Study Fake News?

How Do We Mix Fake News and AI?

To stop the spread of fake news (*fake news detection*)

- In a timely manner (*early detection*)
- Analysing the cost/benefit of blocking fake news on social media

To help fact-checkers

- Through automatic fact-checking
- Flagging articles/trends where fake news might appear

To study how disinformation evolves over time

- Analysing how a specific piece of disinformation changes over time
- Tracking the spread of fake news, both in social media and through different sites

Ok, but where do I get my data from?

- Expert annotators
  - Fact-checking organizations for article-level annotations
  - Watchdog organizations for source-level annotations

- Crowdsourcing
  - Asking non-experts to annotate data

Fact-Checking and Watchdog Organizations

Independent	• Not aligned with the government, companies, or other journalistic outlets
Transparent	• Explain their methodologies • Disclose funding and possible conflicts of interest
Experts in their fields	• Journalists, human rights advocates, etc.

Crowdsourcing

Much cheaper than obtaining a golden label

Tends to have high disagreement

Annotation quality heavily depends on the annotation guidelines

Annotation Levels

- Statement-level
  - The dataset is made up of individually-labelled statements
  - The statements may or may not belong to news articles
- Article-level
  - The dataset is made up of news articles
  - Each article is labelled according to the veracity of its content
- Source-level
  - The dataset is made up of news articles
  - Each article is labelled according to the reliability of its publisher

Possible Issues with the Data

- Most available datasets are:
  - Too small
  - Have data selection biases
- The largest datasets are:
  - Underused
  - Annotated at source-level
- Available languages:
  - Mostly in English and Brazilian Portuguese
  - Not all languages/countries have independent fact-checking agencies (e.g. Sweden)

Three Different Approaches

Knowledge-based

Compare the information in the article against a knowledge base

Content-based

Check for cues of deception in the style of the articles (e.g. within the text itself)

Context-based

Analyse the context in which the article exists (e.g. social media interactions)

Knowledge-Based Approaches

- Automated fact-checking
  - Given a claim, verify its veracity with a knowledge base
  - Identifying previously fact-checked claims

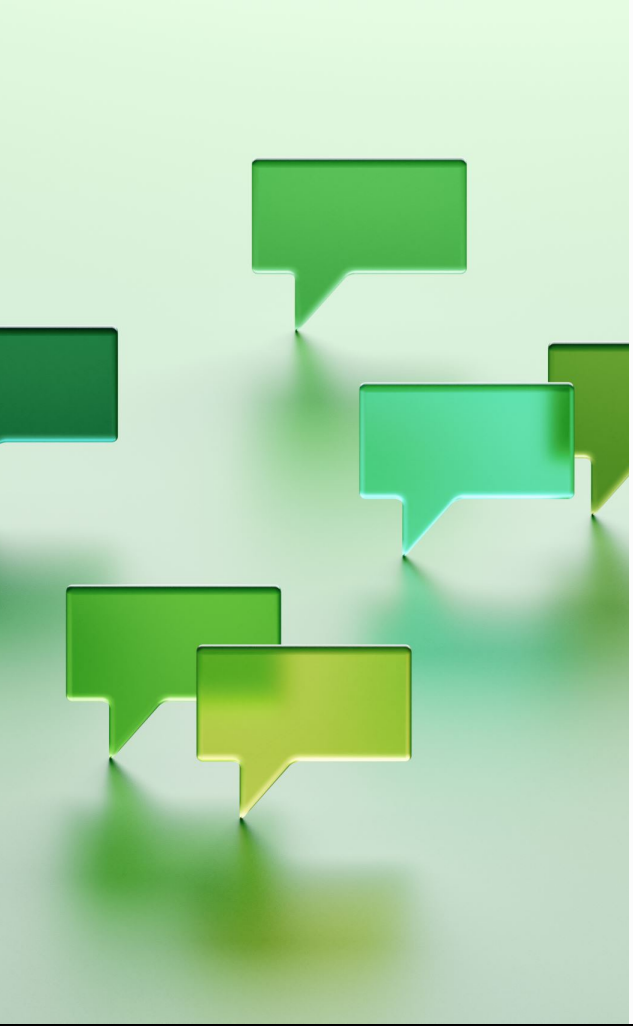
- Note that automated fact-checking encompasses more than just fake news detection!

Content- and Context-Based Approaches

- Are usually focused around machine learning methods
- Use a combination of content- and context-based features
- Can focus on one or more of data, features, and/or models

Content-Based Features

- Textual representations
  - TF-IDF
  - Word embeddings
- Linguistic features
  - Distribution of POS, punctuation, etc.
  - Syntactic trees
- Psycholinguistic features
  - Sentiment and emotion analysis
  - Detecting morality and principles, among others



Context-Based Features

- Can be related to the publication of the article
  - Who wrote and who published the article? When and where was it published?
  - Who are the ad partners of the publishing website?

- Can also be related to social network engagement
  - Who was the original poster?
  - How was the article shared/liked/interacted with?
  - Who interacted with the post?

PART 3
The Flow of Information within Fake News

Exploiting the Ordering of Text for Fake News Detection

My Current Work

- Systematic literature review
  - Will appear as a chapter in my thesis
- “A First Attempt at Unreliable News Detection in Swedish”
  - Appeared at LREC 2022
- “Are You Trying to Convince Me or Are You Trying to Deceive Me? Argumentation in Fake News”
  - To be published

Usual ways of using the text of the article

Bag-of-word features as well as traditional machine learning methods

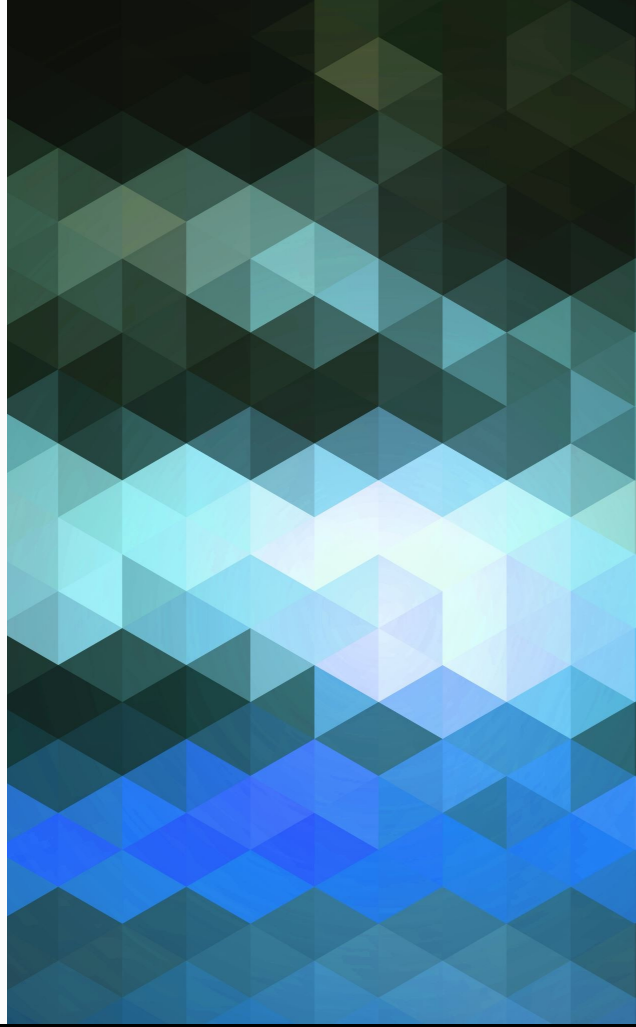
- Generally less effective but more interpretable
- Do not care of the ordering of the text

Use transformer-based architectures for classification

- Better performance but lack interpretability and explainability
- Can be hard to extract useful insights about fake news

My Assumptions

- The order of the text within the article matters
- We can exploit how certain information “flows” within the text
  - We care about how the information changes through the article
  - We do not care about any individual change
- We can use this flow of information both as a feature for fake news detection and to gain insight into how fake news work



An Example – Emotional Flow in Fake News

Ghanem et al. (2021)

- Main idea
  - Exploit how emotion changes through the text to identify disinformation
- Results
  - They perform above the usual baselines
  - We can gain insights on how fake and real articles differ from each other

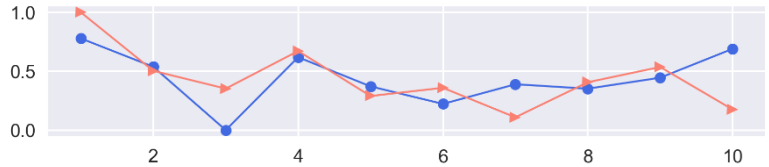


Figure 6: The flow of the *Fear* emotion in **fake** (►) and **real** (●) news articles in the MultiSourceFake dataset. Y-axis presents the average number of *Fear* emotion words in 0-1 scale; the X-axis presents the document text, divided into 10 segments.

An Example – Emotional Flow in Fake News

Ghanem et al. (2021)

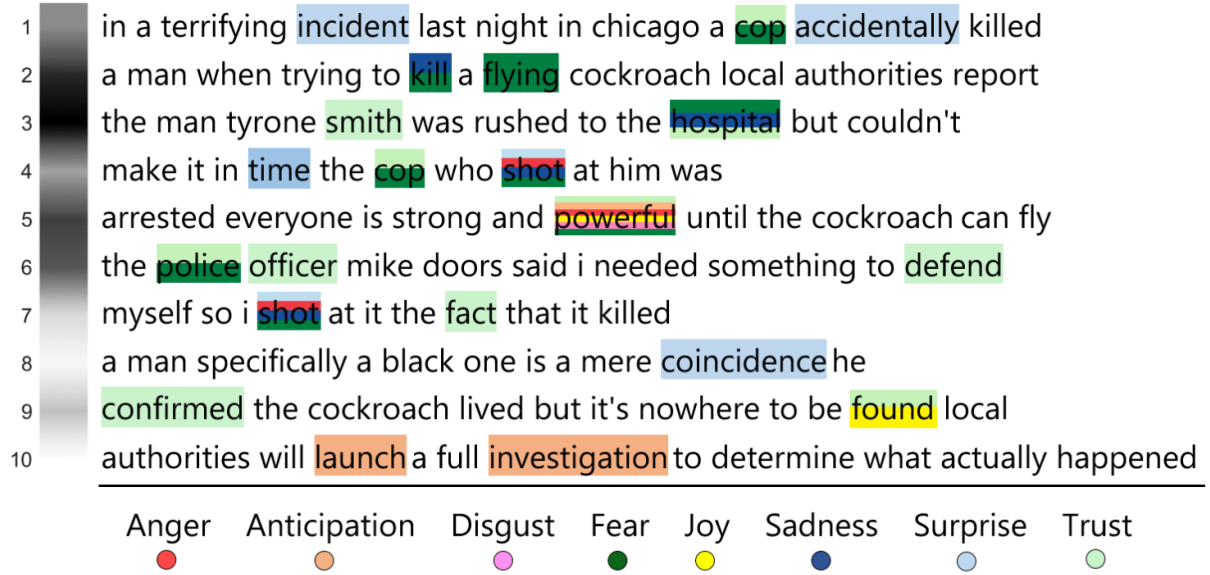


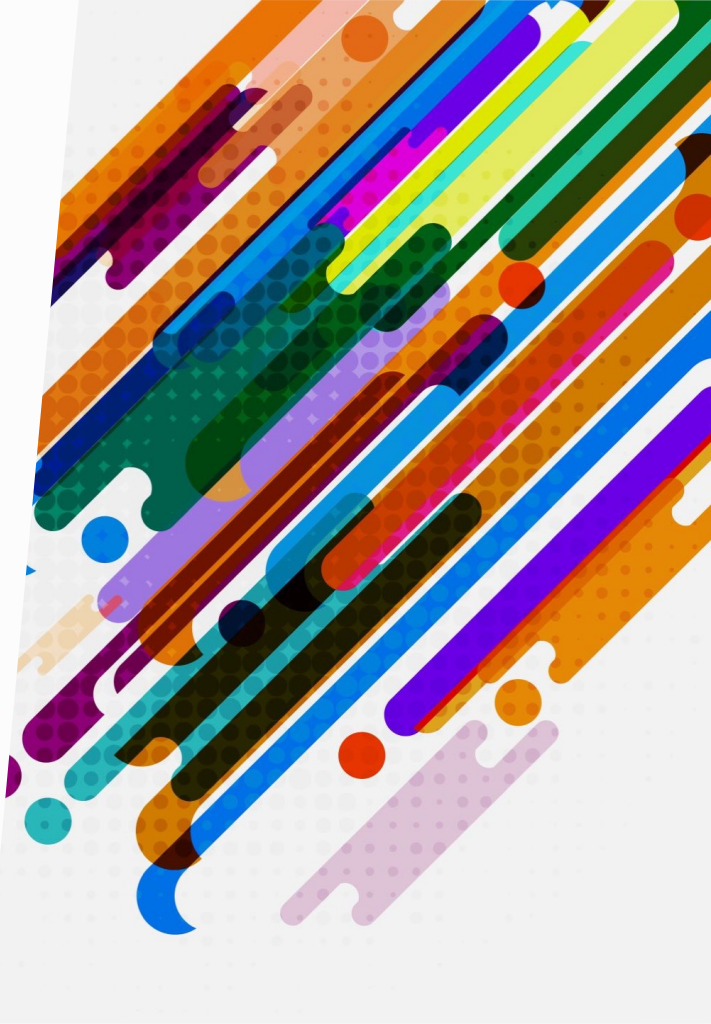
Figure 5: Emotional interpretation of a *fake* news article by showing the attention weights (the bar on the left) and highlighting the emotions in the text.

Current Project – Argumentation Features for Fake News Detection

- Do human-written fake news articles present information in a different way than real articles?

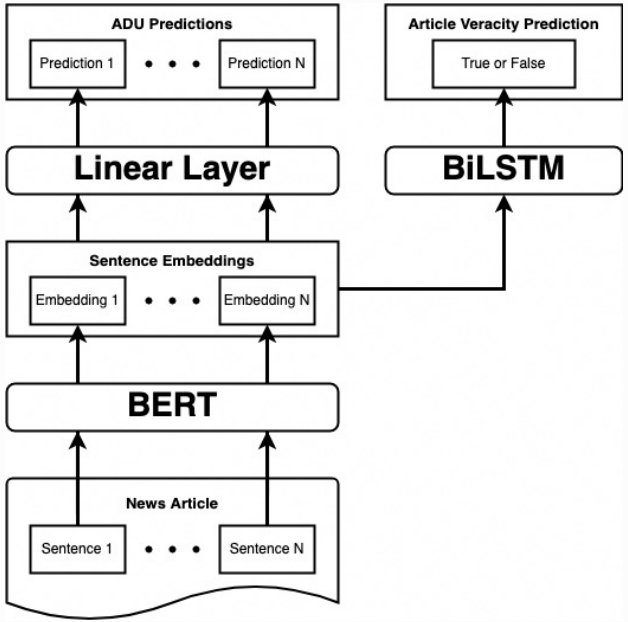
- Process:
  - Train a model to identify the argumentation process in written news media
  - Use this model to obtain sentence embeddings
  - Feed the sentence embeddings to a fake news classifier

- Preliminary answer – yes, we can!



Current Project – Argumentation Features for Fake News Detection

1. Take a news article A and split it into sentences
2. Use a BERT classifier to identify the type of argumentation for each sentence
3. Use the final layer of the [CLS] token to represent the sentences
4. Use a BiLSTM classifier to determine whether the article is real or false



Going Forward

- Explore different sequential structures in text
  - How differently do they behave from each other?

- Explore how fake news relate to other areas of NLP
  - Can we exploit the advances in these other areas?
  - How can we benefit these areas with fake news detection?

- Can we benefit from doing more complex classification?

GÖTEBORGS
UNIVERSITET

SPRÅKBANKENTEXT

Ricardo Muñoz Sánchez
ricardo.munoz.sanchez@svenska.gu.se
rimusa.github.io

Sources

- U.S.A. presidential elections:
  - Chatfield, A. T., C. G. Reddick, and K. P. Choi. “Online Media Use of False News to Frame the 2016 Trump Presidential Campaign.” In *Proceedings of the 18th Annual International Conference on Digital Government Research*, 213–22. Staten Island NY USA: ACM, 2017. https://doi.org/10.1145/3085228.3085295.
  - Benkler, Yochai, Casey Tilton, Bruce Etling, Hal Roberts, Justin Clark, Robert Faris, Jonas Kaiser, and Carolyn Schmitt. “Mail-In Voter Fraud: Anatomy of a Disinformation Campaign.” SSRN Scholarly Paper. Rochester, NY, October 2, 2020. https://doi.org/10.2139/ssrn.3703701.
  - Follman, Mark. “Yes, January 6 Was a Heavily Armed Insurrection. Here’s the Extensive Evidence.” *Mother Jones* (blog), January 6, 2023. https://www.motherjones.com/politics/2023/01/january-6-armed-insurrection-congress-guns-trump-lie/.
  - Image used: https://twitter.com/LeahMillis/status/1611340468226342913

Sources

- Brexit
  - Greene, Ciara M., Robert A. Nash, and Gillian Murphy. “Misremembering Brexit: Partisan Bias and Individual Predictors of False Memories for Fake News Stories among Brexit Voters.” *Memory* 29, no. 5 (May 28, 2021): 587–604. https://doi.org/10.1080/09658211.2021.1923754.
  - Pomerantsev, Peter. “The Disinformation Age: A Revolution in Propaganda.” *The Guardian*, July 27, 2019, sec. Books. https://www.theguardian.com/books/2019/jul/27/the-disinformation-age-a-revolution-in-propaganda.
- Myanmar genocide
  - Mozur, Paul. “A Genocide Incited on Facebook, With Posts From Myanmar’s Military.” *The New York Times*, October 15, 2018, sec. Technology. https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html.

Sources

- AIDS and COVID-19 pandemics
  - Apetrei, Cristian. “Misinformation Played a Deadly Role in Both the COVID and HIV/AIDS Pandemics.” *Fast Company* (blog), August 29, 2022. https://www.fastcompany.com/90782000/misinformation-played-a-deadly-role-in-both-the-covid-and-aids-pandemics.
  - Image used: Ghouaibi, Amira, and Mbali Motsoeneng. “How to Protect Healthcare Workers – and Improve Pandemic Preparedness.” *World Economic Forum* (blog), June 2, 2021. https://www.weforum.org/agenda/2021/06/6-steps-to-protecting-healthcare-workers-improving-pandemic-preparedness-jobs-reset-summit-2021/.
- Polio
  - Bengali, Shashank, and Zulfiqar Ali. “Polio Was Nearly Extinct. Then the Anti-Vaxx Movement Reached Pakistan.” *Los Angeles Times*. September 5, 2019, sec. World & Nation. https://www.latimes.com/world-nation/story/2019-09-04/anti-vaxxers-helping-polio-comeback-pakistan.

Sources

- Climate and smoking disinformation campaigns
  - Reed, Genna, Yogi Hendlin, Anita Desikan, Taryn MacKinney, Emily Berman, and Gretchen T. Goldman. “The Disinformation Playbook: How Industry Manipulates the Science-Policy Process—and How to Restore Scientific Integrity.” *Journal of Public Health Policy* 42, no. 4 (December 1, 2021): 622–34. https://doi.org/10.1057/s41271-021-00318-6.
  - Image used: Chestney, Nina. “U.N. Climate Report Likely to Deliver Stark Warnings on Global Warming.” *Reuters*, August 9, 2021, sec. Environment. https://www.reuters.com/business/environment/un-climate-report-likely-deliver-stark-warnings-global-warming-2021-08-05/.

Sources

- Definitions of fake news
  - Allcott, Hunt, and Matthew Gentzkow. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives* 31, no. 2 (May 2017): 211–36. https://doi.org/10.1257/jep.31.2.211.
  - Tandoc, Edson C., Zheng Wei Lim, and Richard Ling. “Defining ‘Fake News.’” *Digital Journalism* 6, no. 2 (February 7, 2018): 137–53. https://doi.org/10.1080/21670811.2017.1360143.
  - Guess, Andrew M, and Benjamin A Lyons. “Misinformation, Disinformation, and Online Propaganda.” In *Social Media and Democracy: The State of the Field, Prospects for Reform*, Vol. 10. SSRC Anxieties of Democracy. Cambridge, United Kingdom ; New York, NY: Cambridge University Press, 2020.
  - Egelhofer, Jana Laura, and Sophie Lecheler. “Fake News as a Two-Dimensional Phenomenon: A Framework and Research Agenda.” *Annals of the International Communication Association* 43, no. 2 (April 3, 2019): 97–116. https://doi.org/10.1080/23808985.2019.1602782.

Sources

- Definitions of other kinds of mis- and disinformation
  - Pendleton, Susan Coppess. “Rumor Research Revisited and Expanded.” *Language & Communication* 18, no. 1 (January 1, 1998): 69–86. https://doi.org/10.1016/S0271-5309(97)00024-4.
  - Escher, Anna, and Anthony Ha. “WTF Is Clickbait?” *TechCrunch* (blog), September 25, 2016. https://social.techcrunch.com/2016/09/25/wtf-is-clickbait/.
  - Garrett, R. Kelly, Robert Bond, and Shannon Poulsen. “Too Many People Think Satirical News Is Real.” *The Conversation* (blog), August 16, 2019. http://theconversation.com/too-many-people-think-satirical-news-is-real-121666.
  - Bednar, Peter, and Christine Welch. “Bias, Misinformation and the Paradox of Neutrality.” *Informing Science: The International Journal of An Emerging Transdiscipline* 11 (2008): 85–106.
  - Egelhofer, Jana Laura, and Sophie Lecheler. “Fake News as a Two-Dimensional Phenomenon: A Framework and Research Agenda.” *Annals of the International Communication Association* 43, no. 2 (April 3, 2019): 97–116. https://doi.org/10.1080/23808985.2019.1602782.

Sources

- Various literature reviews
  - Bondielli, Alessandro, and Francesco Marcelloni. “A Survey on Fake News and Rumour Detection Techniques.” *Information Sciences* 497 (September 1, 2019): 38–55. https://doi.org/10.1016/j.ins.2019.05.035.
  - Oshikawa, Ray, Jing Qian, and William Yang Wang. “A Survey on Natural Language Processing for Fake News Detection.” In *Proceedings of the 12th Language Resources and Evaluation Conference*, 6086–93. Marseille, France: European Language Resources Association, 2020. https://www.aclweb.org/anthology/2020.lrec-1.747.
  - Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. “Fake News Detection on Social Media: A Data Mining Perspective.” *ACM SIGKDD Explorations Newsletter* 19, no. 1 (September 1, 2017): 22–36. https://doi.org/10.1145/3137597.3137600.

Sources

- Other papers
  - Ghanem, Bilal, Simone Paolo Ponzetto, Paolo Rosso, and Francisco Rangel. “FakeFlow: Fake News Detection by Modeling the Flow of Affective Information.” In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 679–89. Online: Association for Computational Linguistics, 2021. https://doi.org/10.18653/v1/2021.eacl-main.56.
  - Muñoz Sánchez, Ricardo, Eric Johansson, Shakila Tayefeh, and Shreyash Kad. “A First Attempt at Unreliable News Detection in Swedish.” In *Proceedings of the Second International Workshop on Resources and Techniques for User Information in Abusive Language Analysis*, 1–7. Marseille, France: European Language Resources Association, 2022. https://aclanthology.org/2022.restup-1.1.