

# Fake News Apps: How Viable is Automatic Detection?



***Darcy I. Gervasio, MLIS***

Reference Coordinator (Assoc. Librarian)

Purchase College, SUNY

[darcy.gervasio@purchase.edu](mailto:darcy.gervasio@purchase.edu)

Twitter: @darcyiris

***CiT 2019 Conference***

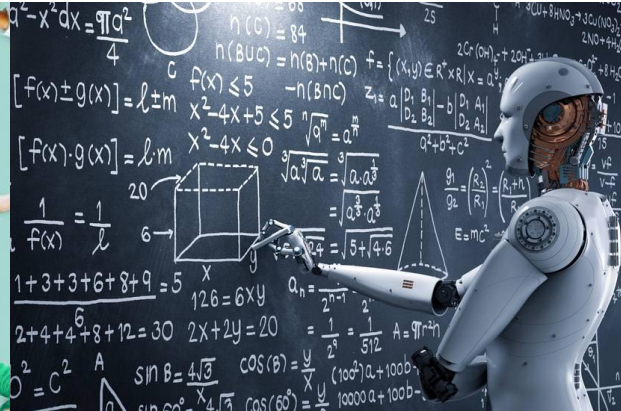
Friday, May 31, 2019

# Questions:

1. Do current apps work well enough to combat fake news in a meaningful way?
- 
2. As educators, what should we teach our students about automatic fake news detection?

# Methods of Fake News Detection:

1. Fact-checking by human experts
2. Crowdsourced flagging / annotation
3. Computational prediction (i.e. automated detection without human intervention)



# 10-20 hours

The time lag between the sharing of misinformation (a fake news story) and the sharing of fact-checking content that debunks it on social media.

Shao, Chengcheng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. "Hoaxy: A Platform for Tracking Online Misinformation." In *Proceedings of the 25th International Conference Companion on World Wide Web*, 745–50. Montreal, Quebec, Canada, 2016. <https://doi.org/10.1145/2872518.2890098>.

# Background:

How do automatic fake news detection apps work?

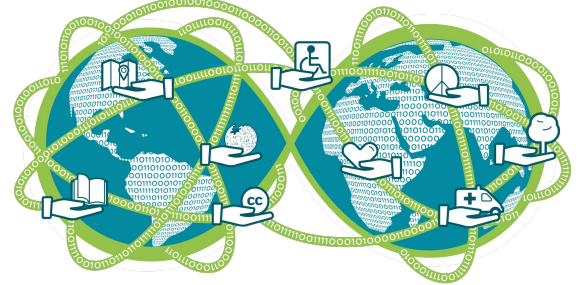
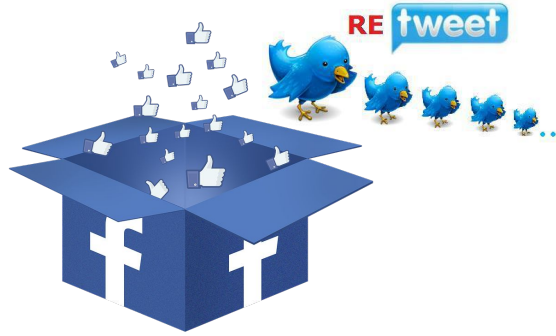
---

Answer: Algorithms!

\*\*Also called “A.I.” or “learning algorithms” = mathematical formula that uses past events to predict future ones. Usually the more items accumulate in a dataset, the more accurate its predictions, hence “learning.”

# Types of fake news detection algorithms:

1. Content / text analysis
2. User behavior / user engagement analysis
3. Social network spread / diffusion analysis



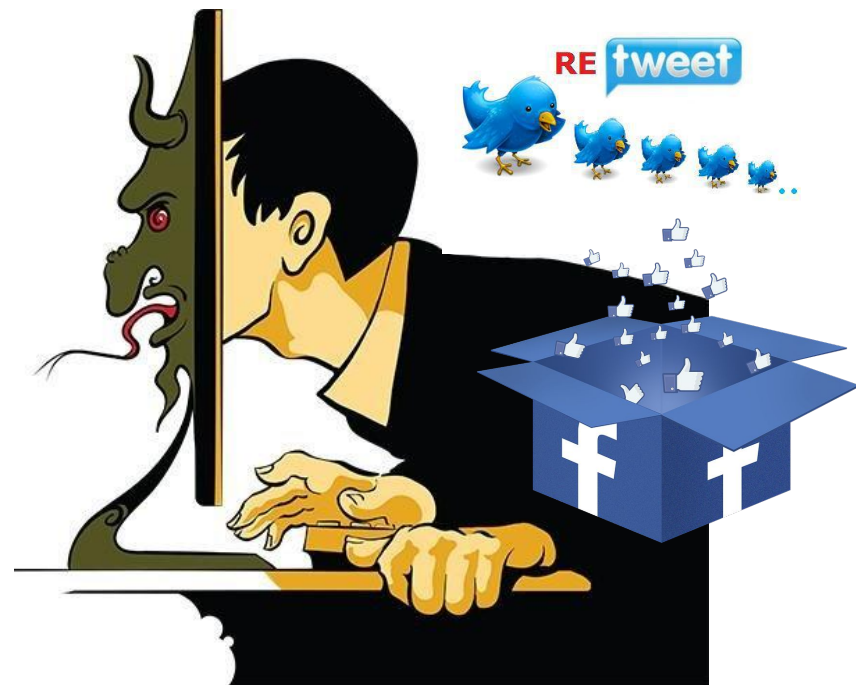
# 1. Content / textual analysis:



- Algorithms examine text of news articles or websites
- Gauge the probability a story is fake/real or give a “credibility score”
- Linguistic properties like syntax, lexicon, emotional affect, grammar, language prediction are used

## 2. User behavior / engagement analysis:

- Algorithms track share/“like” patterns of fake news stories
- *Who* shares news can very accurately predict whether it’s real or fake without text analysis. \*
- Look for “good actors” and vet out bots/trolls/cyborgs using “reputation scores”

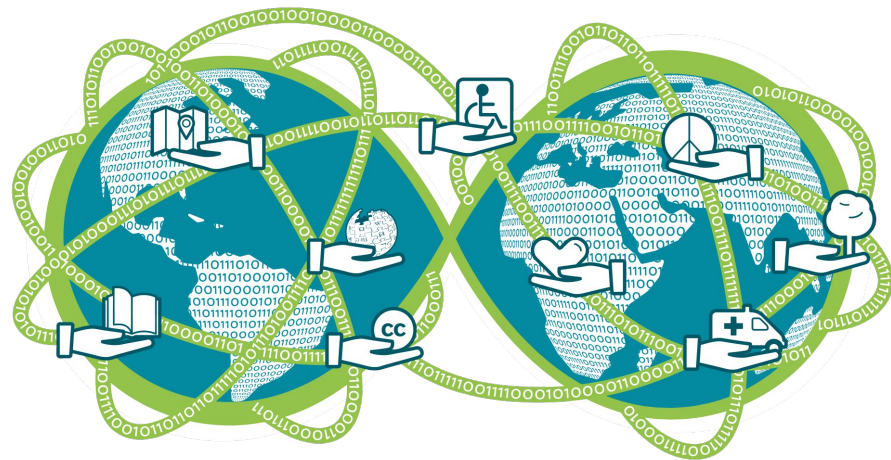


\*Tacchini et. al. “Some Like It Hoax: Automated Fake News Detection in Social Networks.” In *Proceedings of the Second Workshop on Data Science for Social Good (SoGood)*, 1960:1–11. Skopje, Macedonia: CEUR Workshop Proceedings, 2017.



# 3. Social network spread / diffusion analysis

- Algorithms track how fake stories “go viral” on social media
- Fake stories spread differently than real ones<sup>\*^</sup>
- The goal: stop the spread of fake news **before** it goes viral.



<sup>\*</sup>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 (June 2017): 22–36. <https://arxiv.org/abs/1708.01967>

<sup>^</sup>Zhao, Zilong, Jichang Zhao, Yukie Sano, Orr Levy, Hideki Takayasu, Misako Takayasu, Daqing Li, and Shlomo Havlin. “Fake News Propagate Differently from Real News Even at Early Stages of Spreading.” *ArXiv:1803.03443 [Physics]*, March 9, 2018. <http://arxiv.org/abs/1803.03443>.

Some algorithms & tools researchers are using...

LIWC Lexicon

N-grams

heuristics

Boolean Label Crowdsourcing

natural language processing

language prediction

logistic regression

IBM Watson

convolutional neural network

# Question # 1:

Do current apps work well enough to combat fake news in a meaningful way?

---

Answer: Not enough data to tell!

\*\*Truly automated fake news detection has not fully been achieved. No scientific evaluation of apps currently on the market (i.e. CheckThis, B.S. Detector, Fake News Detector)

# Fake News Detection Tools By Type

## Web Platforms Using Crowd-Sourced Flagging / Fact-Checking:



Climate Feedback

**h.** hypothes.is

**Fiskkit**

## Browser Extensions Using Curated Blacklists / Keyword-Matching



Project Fib

Offered by: [NabanitaDe](#)



B.S. Detector

Offered by: [bsdetector.tech](#)

 **NewsGuard**



Fake News Detector

Offered by: [fakenewsdetector.org](#)

## Web Platform Using Learning Algorithms

 **HOAXY**

## Browser Extensions Using Learning Algorithms



Check This by MetaCert



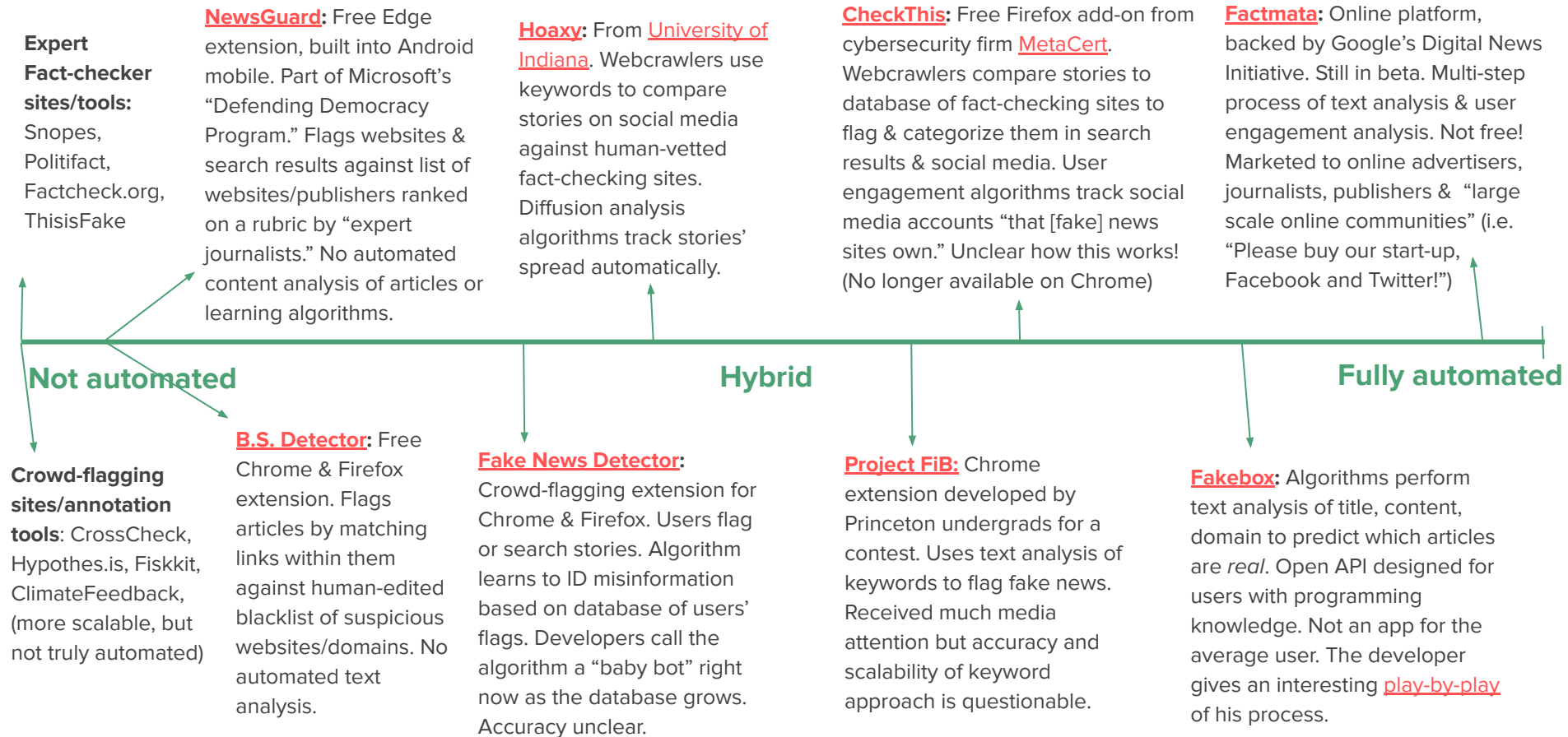
Fake News Detector

Offered by: [fakenewsdetector.org](#)



**FACTMATA**

# Fake News Apps: Scale of Automation



# Hybrid

- Research shows using two or more methods may be most effective
- Most automatic detectors rely on test datasets hand-coded by human experts
- Big companies like Facebook, Google, Microsoft are taking a mixed methods approach

# Methods

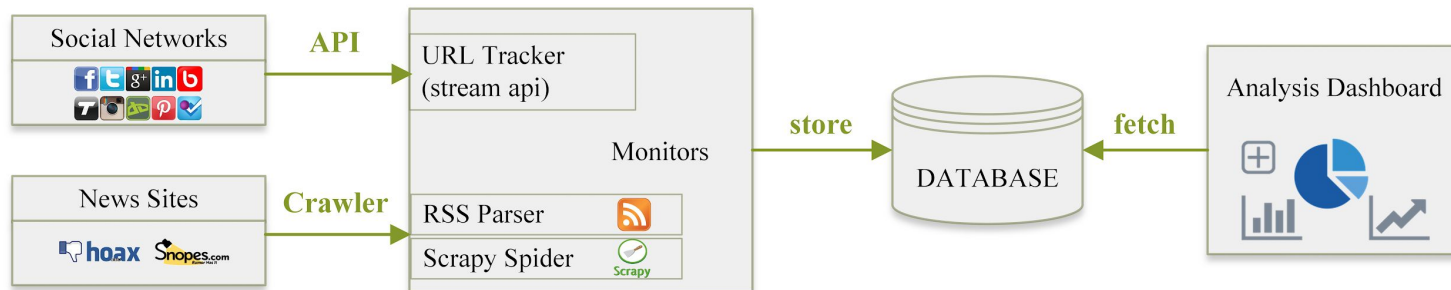
*Examples:*

- *Facebook's flagging system*
  - *Hoaxy*
  - *Fake News Detector Extension*
-

# Hybrid Methods

Example: [Hoaxy Platform](#) from University of Indiana (Hoaxy now in beta)

1. Web-crawlers identify fake stories and auto-link to debunking stories (“light” content analysis reliant on keywords rather than learning algorithms)
2. Relies on debunking sites (expert fact-checking)
3. Monitor spread using social media APIs (automated diffusion analysis)
4. Collect data for future research & display in visual dashboard



# Hybrid Methods

Example: Facebook's flagging system



1. Users flag content as “false news” (Crowdsourcing)
2. Algorithms determine which posts are likeliest to be fake--and which human flaggers are likeliest to be reliable--and score them\* (Computational Prediction/Automatic Detection)
3. The most objectionable content is reviewed by human “content moderators” at third-party fact-checking companies (Expert fact-checking)

\*We don't know exactly how Facebook does this. Probably uses algorithms that perform text analysis and user engagement analysis



# THE TRAUMA FLOOR

The secret lives of Facebook moderators in America

By Casey Newton | @CaseyNewton | Feb 25, 2019, 8:00am EST

Illustrations by Corey Brickley | Photography by Jessica Chou



Newton, Casey. “The Secret Lives of Facebook Moderators in America.” *The Verge*, 25 Feb. 2019,  
<https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>.

“..Facebook and Twitter rely on an army of workers employed to soak up the worst of humanity in order to protect the rest of us. And there are legions of them—a vast, invisible pool of human labor.”

~

“Everybody hits the wall, generally between three and five months...you just think, ‘Holy shit, what am I spending my day doing? This is awful.’”

- a former YouTube Content Moderator

Chen, Adrian. “The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed.” *Wired*, Oct. 2014.  
[www.wired.com](https://www.wired.com),  
<https://www.wired.com/2014/10/content-moderation/>.

# Gaps in the literature:

- Scholarly research not translated fast enough to real world
  - Engineers inventing & testing new algorithms, but few are turning them into user-friendly apps
  - Algorithms are not always universal; many are platform-specific
  - Human intervention still required to create & test datasets
- Financial incentives for app developers may lead to secrecy, e.g.: Factmata, Facebook's patents, etc.
- **Ethics, free speech implications of using bots to automatically suppress content are rarely discussed.**

**“...a judgement on the value of information should not be performed exclusively by machines, in case they are given total control to decide which information is displayed... Freedom of speech must be protected at all cost...”**

Figueira, Álvaro, and Luciana Oliveira. “The Current State of Fake News: Challenges and Opportunities.” *Procedia Computer Science*, CENTERIS 2017, 121 (January 1, 2017): pp. 822. <https://doi.org/10.1016/j.procs.2017.11.106>.

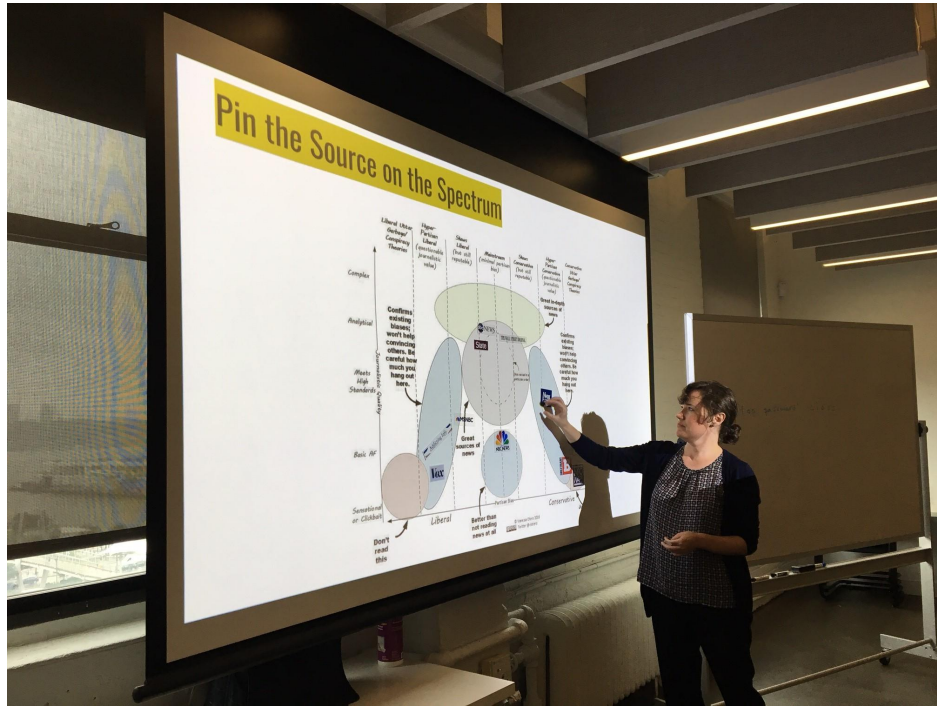
## Question # 2:

As educators, what should we teach our students about automatic fake news detection?

---

Answer: Use them... but be critical!

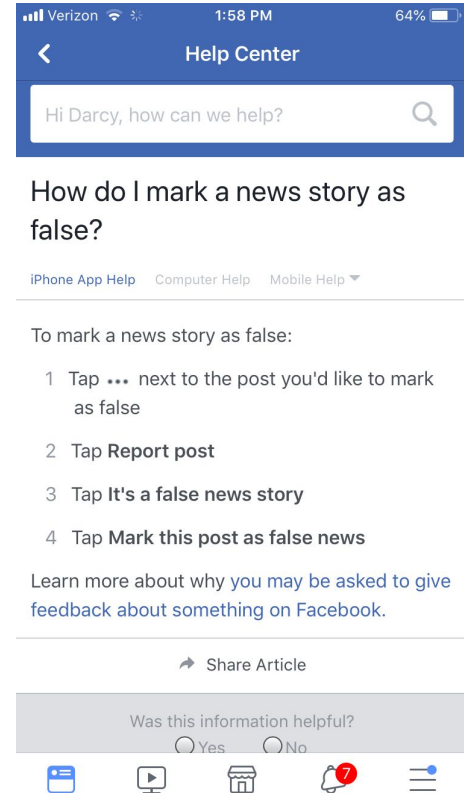
# Fake News Apps in the Classroom: Mindsets



- Critical thinking and information literacy remain crucial
- Teach skepticism, analysis. Go beyond identification!
- Apps are useful “alarm bells”
- Empower students to “fight back” and resist passivity!

# Fake News Apps in the Classroom: Activities

- Have students research & debunk fake news stories and contribute to Hypothes.is, Climatefeedback.org or other crowd-sourced web annotation platforms.
- Ask students to install multiple browser extensions and critically compare the results of each, or conduct usability studies.
- Teach students how to flag “false news” on Twitter & Facebook, how to report articles to Snopes or Politifact, etc.

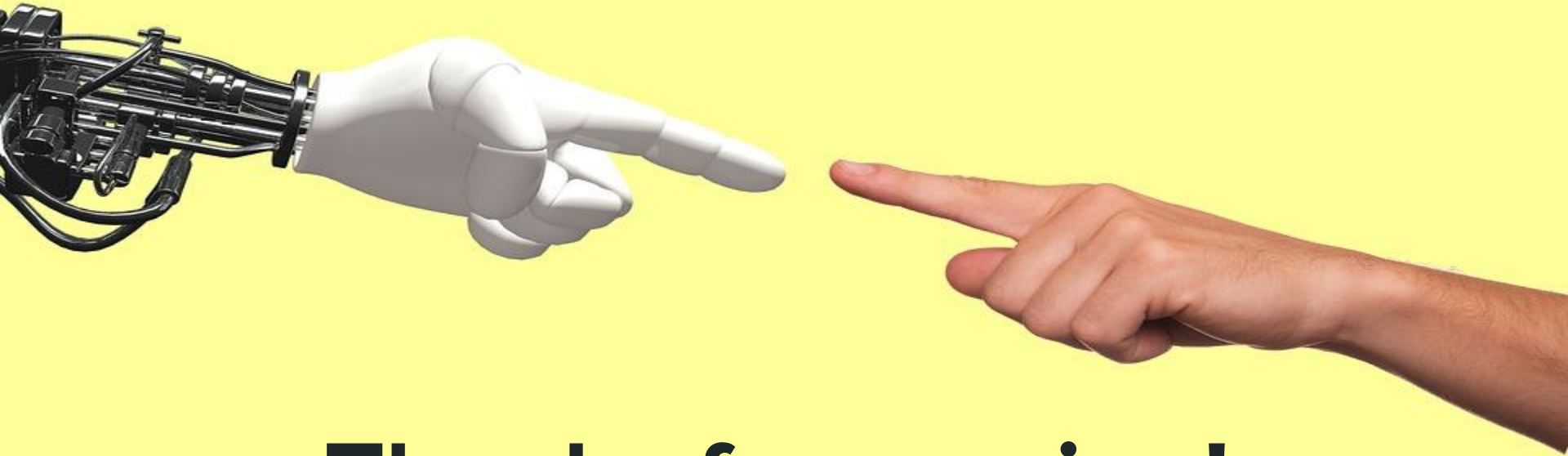


## Summary:



- We're getting closer to fully automated detection
  - Main benefit: scaling up to entire social networks to stop the viral spread of fake news
  - Expert fact-checking is still important part of the process in developing algorithms
- 
- Media literacy is perennial! Users must learn to recognize fake news, flag it, and refrain from sharing it.

**Any questions?**



**Thanks for coming!**



# Bibliography:

- Adewole, K. et al. (2017). "Malicious Accounts: Dark of the Social Networks." *Journal of Network and Computer Applications* 79 (February): 41–67.  
<https://doi.org/10.1016/j.jnca.2016.11.030>.
- Chen, A. (2014, October 23). The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed. *Wired*. Retrieved from  
<https://www.wired.com/2014/10/content-moderation/>
- Edell, A. (2018, January 11). I trained fake news detection AI with >95% accuracy, and almost went crazy. Retrieved May 23, 2018, from  
<https://towardsdatascience.com/i-trained-fake-news-detection-ai-with-95-accuracy-and-almost-went-crazy-d10589aa57c>
- Factmata. (n.d.). Factmata - Our Technology. Retrieved April 27, 2018, from <https://factmata.com/technology.html>
- Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121, 817–825.**  
<https://doi.org/10.1016/j.procs.2017.11.106> **\*\*particularly helpful overview\*\***
- Fiskkit. (n.d.). Fiskkit. Retrieved May 25, 2018, from <http://fiskkit.com/>
- Hoaxy: A Platform for Tracking Online Misinformation | CNetS. (n.d.). Retrieved May 29, 2018, from <http://cnets.indiana.edu/blog/2016/12/21/hoaxy/>
- Jackson, J. (2016). "Fake News Clampdown: Google Gives €150,000 to Fact-Checking Projects." *The Guardian*, November 17, 2016, sec. Media.  
<https://www.theguardian.com/media/2016/nov/17/fake-news-google-funding-fact-checking-us-election>.
- Lapowsky, I. (2018). "NewsGuard Wants to Fight Fake News With Humans, Not Algorithms." *Wired*, August 23, 2018.  
<https://www.wired.com/story/newsguard-extension-fake-news-trust-score/>.
- Lee, E. (2019). "Veterans of the News Business Are Now Fighting Fakes." *The New York Times*, January 17, 2019, sec. Business.  
<https://www.nytimes.com/2019/01/16/business/media/media-steve-brill-fake-news.html>.
- Machine Box. (2017, October 20). Introducing Fakebox: Detect fake news with Machine Learning. Retrieved May 23, 2018, from  
<https://medium.com/@machineboxio/introducing-fakebox-detect-fake-news-with-machine-learning-f602c39aad04>
- Newton, C. (2016, December 7). Facebook is patenting a tool that could help automate removal of fake news [News]. Retrieved May 25, 2018, from  
<https://www.theverge.com/2016/12/7/13868650/facebook-fake-news-patent-tool-machine-learning-content>
- Newton, C. (2019, February 25). The secret lives of Facebook moderators in America. Retrieved February 27, 2019, from  
<https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>

# Bibliography:

- Papadopoulou, O. et al. (2017). "A Two-Level Classification Approach for Detecting Clickbait Posts Using Text-Based Features." Working Paper. Clickbait Challenge 2017. arXiv.org. <http://arxiv.org/abs/1710.08528>.
- Pérez-Rosas, V., et al. (2017). Automatic Detection of Fake News. *ArXiv:1708.07104 [Cs]*. Retrieved from <http://arxiv.org/abs/1708.07104>
- Roberts, S. T. (2012, June 20). Behind the Screen: Commercial Content Moderation (CCM). Retrieved March 6, 2019, from <https://illusionofvolition.com/behind-the-screen/>
- Ruchansky, N. et al. (2017). CSI: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797–806). Singapore, Singapore: ACM. <https://doi.org/10.1145/3132847.3132877>
- Shao, C. et al.. (2016). Hoaxy: A Platform for Tracking Online Misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 745–750). Montreal, Quebec, Canada. <https://doi.org/10.1145/2872518.2890098>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22–36. \*\*particularly helpful overview\*\***
- Shu, K., Wang, S., & Liu, H. (2017). "Exploiting Tri-Relationship for Fake News Detection." In *Proceedings of 12th ACM International Conference on Web Search and Data Mining* (WSDM 2019). Melbourne, Australia. <https://doi.org/arXiv:1712.07709>.
- Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some Like it Hoax: Automated Fake News Detection in Social Networks. In *Proceedings of the Second Workshop on Data Science for Social Good (SoGood)* (Vol. 1960, pp. 1–11). Skopje, Macedonia: CEUR Workshop Proceedings. Retrieved from <http://arxiv.org/abs/1704.07506>
- Tschiatschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., & Krause, A. (2018). Fake News Detection in Social Networks via Crowd Signals. In *Companion Proceedings of the The Web Conference 2018* (pp. 517–524). Lyon, France: International World Wide Web Conferences Steering Committee. <https://doi.org/10.1145/3184558.3188722>
- Wang, W. (2017). "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection." In *ACL 2017*. <https://doi.org/arXiv:1705.00648> [cs]
- Warren, T. (2019). "Microsoft Is Trying to Fight Fake News with Its Edge Mobile Browser." *The Verge*. January 23, 2019. <https://www.theverge.com/2019/1/23/18194078/microsoft-newsquard-edge-mobile-partnership>
- Wilner, T. (n.d.). Flagging Fake News. Retrieved April 27, 2018, from <http://nimanreports.org/articles/flagging-fake-news/>
- Yam, K. (2016, December 9). Students Create Brilliant Extension That Identifies Fake News. *Huffington Post*. Retrieved from [https://www.huffingtonpost.com/entry/fib-google-chrome-extension-fight-fake-news\\_us\\_58498afae4b0905b3442187d](https://www.huffingtonpost.com/entry/fib-google-chrome-extension-fight-fake-news_us_58498afae4b0905b3442187d)
- Zhao, Z. et al. (2018). "Fake News Propagate Differently from Real News Even at Early Stages of Spreading." arXiv.org. <http://arxiv.org/abs/1803.03443>.