

Detecting and Addressing Misinformation through Artificial Intelligence

CMPSC 442

Sasha Ahrestani, Gabien Bryan, Alec Pratt, and Divya Rustagi

What is Misinformation?

Misinformation is
information that is
intended to deceive.

Fake News Is A Real Problem

Facebook engagement of the top five fake election stories*



Total Facebook engagement for top 20 election stories (August-election day)



* Engagement is measured as total number of shares, reactions and comments

Source: Buzzsumo via Buzzfeed

statista

Sources: 15

What Challenges Do We Face Today?

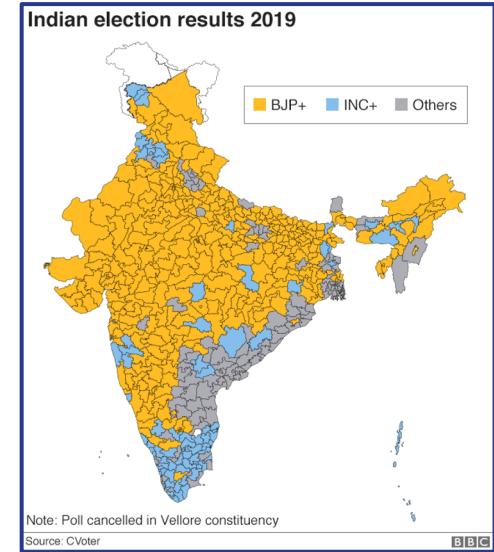
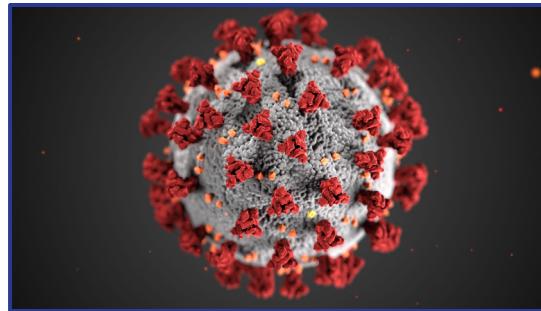
Election Misinformation

- United States of America
- India



COVID-19 Misinformation

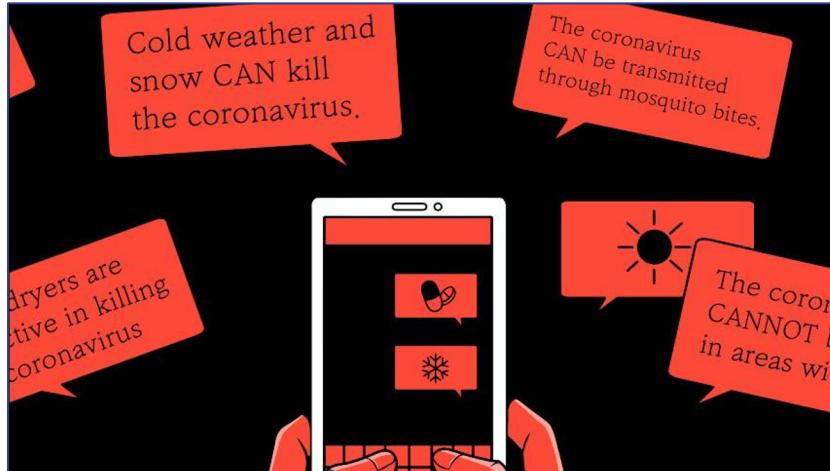
- Vaccines
- Masks
- General Information



Sources: 1, 2, 3, 4, 5

Risks of Misinformation

Distrust in **Democracy**
and **Election System**



Distrust in **Science**

Hatred Towards **Minority**
or **Targeted Groups**



Sources: 1, 2, 3, 4, 5

How can Artificial Intelligence (AI) Help?

- Automated Misinformation Detection (**Fact Checking**)
- AI-powered **Chatbots**
- Identification of **Deep Fakes**



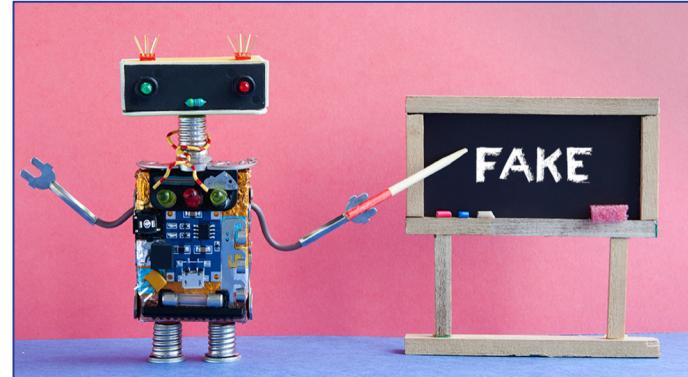
Automated Misinformation Detection

A “Fake News” Infodemic

- Anyone able to share their opinions
- Everyone thus overwhelmed by information
- Need a filter to find what's trustworthy

Detection Techniques

- Bot-like activity (**spam content**)
- True/false assertions based on data (**supervised learning**)
- Authenticity based on user engagement (**unsupervised learning**)



Sources: 16, 17

AI-Powered Chatbots

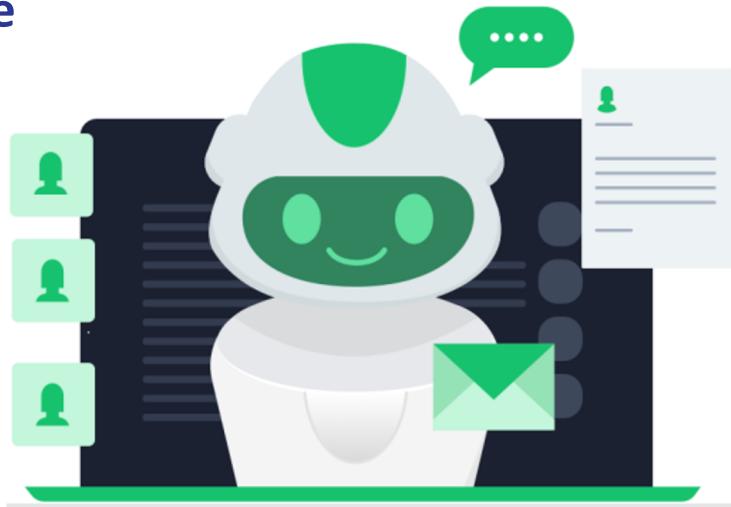
Correcting widespread misinformation and speculation

- FAQs
- Conversational questions

Answering/helping with AI instead of real people

- Hotlines
- Virtual chat rooms

Sources: 14, 11



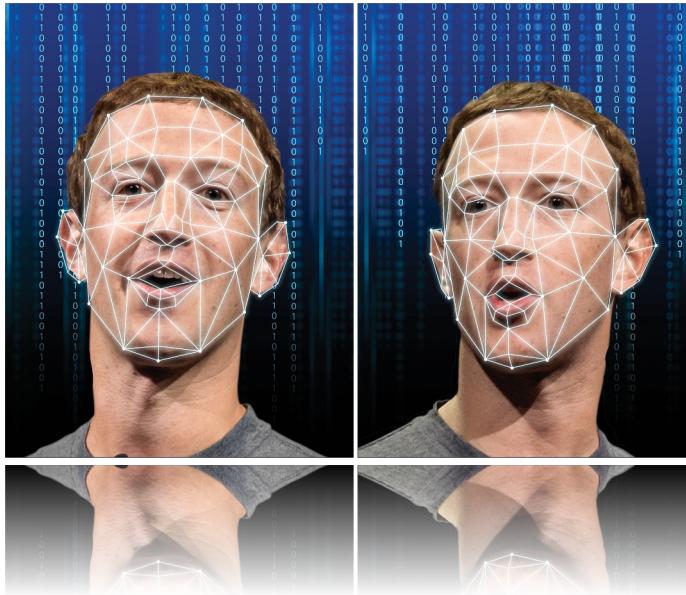
Identification of DeepFakes

What is a Deepfake?

- Synthetically-modified images, videos, and audio
- Used to spread misinformation by appearing real

Detection Techniques

- Analyze original distinct **facial movements** and compare
- Check for **inconsistencies** across frames or audio waves



Sources: 11, 12, 13

Current AI Solutions

How Facebook Does It

To identify misinformation related to articles spotted by fact-checkers, Facebook reported, **its systems had to be trained to detect images that appeared alike to a person but typically not to a computer.**



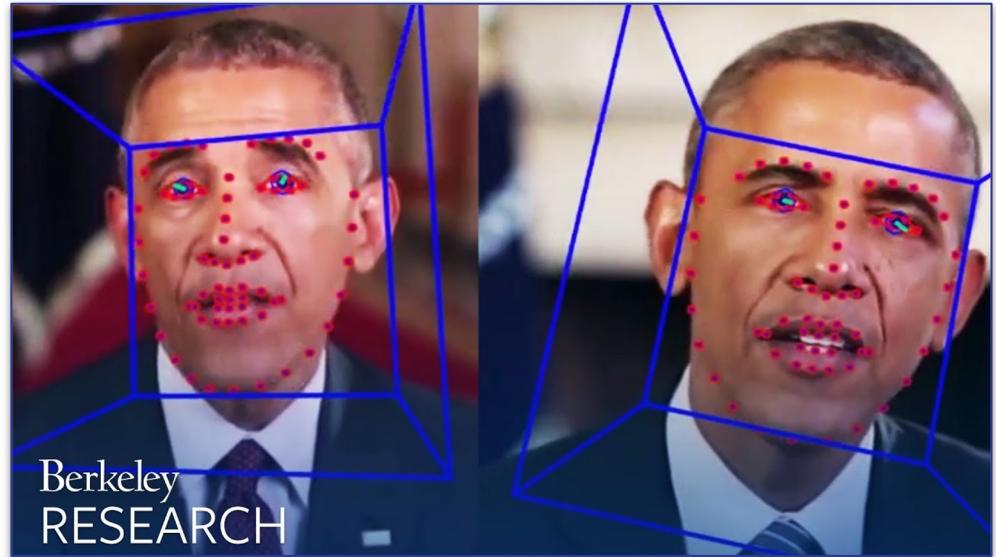
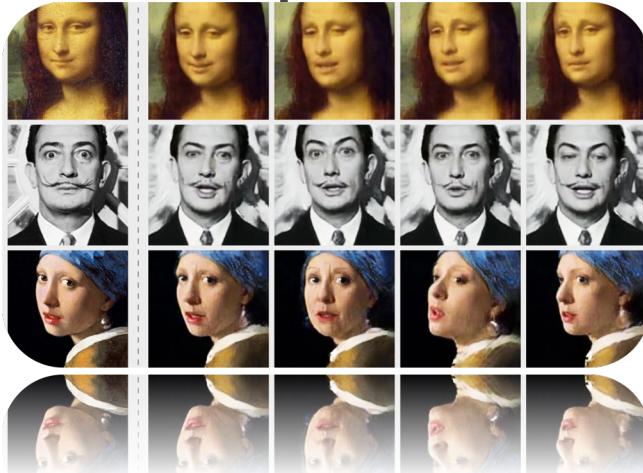
“Facebook is also applying its existing multimodal content analysis tools, which look at both text and images together to interpret a post.”

Sources: 18

How Well Does AI Work?

Deep Fake Detection

- 97.29% Accurate
- Data Comparison



Sources: 6, 10

How Well Does AI Work?

Overall Change in Internet Social Climate

- Survey Specific Topics
- Private/Open Social Media

Banning Users

- Facebook: 5.5 billion



Sources: 7, 9

How Well Does AI Work?

AI Generated Text Detection

- Real vs AI

AI Combating AI

- Still developing
- Easier to create than detect
- Indistinguishable from humans

SYSTEM PROMPT (HUMAN-WRITTEN)

Recycling is good for the world.

NO! YOU COULD NOT BE MORE WRONG!!

MODEL COMPLETION (MACHINE-WRITTEN, 25 TRIES)

Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a



Sources: 8, 10

Ending Statement



Sources

1. [2020 presidential election misinformation was predictable, experts say](#)
2. [6 Types of Misinformation to Beware Of on Election Day. \(And What to Do if You Spot Them.\) \(Published 2018\)](#)
3. [Indian Misinformation Will Decide World's Largest Election](#)
4. [COVID-19 misinformation](#)
5. [An exploration of how fake news is taking over social media and putting public health at risk](#)
6. [Advanced Machine Learning techniques for fake news \(online disinformation\) detection: A systematic mapping study](#)
7. [The Future of Free Speech, Trolls, Anonymity, and Fake News Online](#)
8. [The role of technology in online misinformation](#)
9. [Cracking the Whip On Misinformation: 5.5 Billion Fake Accounts On Facebook Axed In 2020](#)
10. [We're fighting fake news AI bots by using more AI. That's a mistake.](#)

Sources

1. [Fighting Deepfakes: Media and Internet Giants' Converging and Diverging Strategies Against Hi-Tech Misinformation](#)
2. [Three Types of Deepfake Detection](#)
3. [How Deepfake Technology Can Become More Dangerous Than a Nuclear Weapon](#)
4. [How Conversational AI Can Help Efficiently Disseminate Covid-19 Vaccines](#)
5. [MISINFORMATION | definition in the Cambridge English Dictionary](#)
6. [Fake News Is Rampant, Here Is How Artificial Intelligence Can Help](#)
7. [How to deal with AI-enabled disinformation](#)
8. [How Facebook Is Using AI to Fight COVID-19 Misinformation - IEEE Spectrum](#)

Image Sources

- 1) [Food Safety and the Coronavirus Disease 2019 \(COVID-19\)](#)
- 2) [Biden Wins Key States Wisconsin, Michigan In Knife-Edge Fight With Trump](#)
- 3) [India general election 2019: What happened?](#)
- 4) [Qualified Immunity and HB3653 | Facts Matter Moments Podcast | FOP](#)
- 5) [Cute bot say users hello. chatbot greets. online consultation. | Free Vector on Freepik](#)
- 6) [How Deepfake Technology Can Become More Dangerous Than a Nuclear Weapon](#)
- 7) [Facebook Brand Resources](#)
- 8) [OpenAI “Better Language Models and Their Implications”](#)
- 9) [Google logo](#)
- 10) [New technology helps media detect ‘deepfakes’ | University of California](#)
- 11) [AI and Chatbots in Customer Service | Artificial Intelligence | Freshdesk](#)