Fake News and Misinformation

Diyi Yang
Stanford CS
What is Fake News?
fake news  [fəık-nouz, nyoʊz]  

noun

1 false news stories, often of a sensational nature, created to be distributed for the purpose of generating revenue, or promoting figure, political movement, company, etc.:  
   It's impossible to avoid clickbait and fake news on social media.

2 a parody that presents current events or other news topics for obviously satirical imitation of journalism:  
   The website publishes fake news that is hilarious and surprisingly insightful.

3 Sometimes facetious, (used as a conversational tactic to disp information that is perceived as hostile or unflattering):  
   The senator insisted that recent polls forecasting an election loss were just fake news.

https://virtual.2020.emnlp.org/tutorial_T2.html
75% of young people in Europe cannot tell real news from fake news

75% of young people between the age of 15 and 24 in Europe admit that they cannot tell real news from fake news, said Mariya Gabriel, the EU commissioner for digital economy...

Disinformation, misinformation and malinformation

https://firstdraftnews.org/long-form-article/understanding-information-disorder/
7 TYPES OF MIS- & DISINFORMATION

- **Satire or parody**: No intention to cause harm but has potential to fool.
- **Misleading content**: Misleading use of information to frame an issue or individual.
- **Imposter content**: When genuine sources are impersonated.
- **Fabricated content**: New content that is 100% false, made to deceive and do harm.

- **False connection**: When headlines, visuals or captions don’t support the content.
- **False context**: When genuine content is shared with false contextual information.
- **Manipulated content**: When genuine information or imagery is manipulated to deceive.

https://firstdraftnews.org/long-form-article/understanding-information-disorder/
How do we know what information is fake/real?
Fact Checking When It Comes to Fake News

Figure 2: A natural language processing framework for automated fact-checking.
In manufacturing, "I brought back 700,000 jobs. (Obama and Biden) brought back nothing."

If your time is short

• Trump is wrong about the job gains on his watch; the actual increase is about 450,000 prior to the pandemic.
• As for Obama and Biden, they saw gains of 916,000 if you start counting with the recovery from the Great Recession, which is the fairest comparison if you also ignore the losses under Trump during the pandemic.

Figure 1: An example of a fact-checked statement. Referring to the manufacturing sector, Donald Trump said ‘‘I brought back 700,000 jobs. Obama and Biden brought back nothing.’’ The fact-checker gave the verdict False based on the collected evidence.

(Wang, 2017)
Can we really detect fake news?

Nintendo Switch game console to launch in March for $299 The Nintendo Switch video game console will sell for about $260 in Japan, starting March 3, the same date as its global rollout in the U.S. and Europe. The Japanese company promises the device will be packed with fun features of all its past machines and more. Nintendo is promising a more immersive, interactive experience with the Switch, including online playing and using the remote controller in games that don’t require players to be constantly staring at a display.

New Nintendo Switch game console to launch in March for $99 Nintendo plans a promotional roll out of its new Nintendo switch game console. For a limited time, the console will roll out for an introductory price of $99. Nintendo promises to pack the new console with fun features not present in past machines. The new console contains new features such as motion detectors and immerse and interactive gaming. The new introductory price will be available for two months to show the public the new advances in gaming.

Can we really detect fake news?

Kim And Kanye Silence Divorce Rumors With Family Photo. Kanye took to Twitter on Tuesday to share a photo of his family, simply writing, “Happy Holidays.” In the picture, seemingly taken at Kris Jenner’s annual Christmas Eve party, Kim and a newly blond Kanye pose with their children, North, 3, and Saint, 1. After Kanye’s hospitalization, reports that there was trouble in paradise with Kim started brewing. But E! News shut down the speculation with a family source denying the rumors and telling the site, “It’s been a very hard couple of months.”

Kim Kardashian Reportedly Cheating With Marquette King as She Gears up for Divorce From Kanye West. Kim Kardashian is ready to file for divorce from Kanye West but has she REALLY been cheating on him with Oakland Raiders punter Marquette King? The NFL star seemingly took to Twitter to address rumors that they’ve been getting close amid Kanye’s mental breakdown, which were originally started by sports blogger Terez Owens. While he doesn’t appear to confirm or deny an affair, her reps said there is “no truth whatsoever” to the reports and labeled the situation “fabricated.”

Fake News Detection

Would it be too late?
Combating **Unverified** Information

**Who** are producing such fake news?

**Who** are spreading it?

**Why** do they do it?

**Where** do they do it?

**How** do they do it?

**Who** do they want to “harm” or “influence”?

**What** do they plan to do next?

Is this **part of** something “larger”?
Proactive Ways to Combat Unverified Information

**Inform/Protect users** via technology (e.g., misinformation detection and visualization)

Who has the responsibility here?

**Debunking and Fact-Checking**

What’s the pros and cons here?

**Prebunking** and Intervention

Incentives to sign up

**Regulation**
Prebunking to inoculate against misinformation

Screenshots from the emotional language video (studies 1 and 6). All videos can be viewed at https://inoculation.science/inoculation-videos/


The goal with prebunking. Source: WHO via https://firstdraftnews.org/articles/a-guide-to-prebunking-a-promising-way-to-inoculate-against-misinformation/
Watermark using randomized algorithms for AI-Generated Content

Example credit: https://twitter.com/tomgoldsteincs/status/1618287665006403585
Watermark

Tom Goldstein
@tomgoldsteincs

#OpenAI is planning to stop #ChatGPT users from making social media bots and cheating on homework by "watermarking" outputs. How well could this really work? Here's just 23 words from a 1.3B parameter watermarked LLM. We detected it with 99.99999999994% confidence. Here's how.

Prompt

The watermark detection algorithm can be made public [...] or it can be kept private and run behind an API. We seek a watermark with the following properties:

Generated Text (OPT-6.7B)

— minimal marginal probability for a detection attempt.
— Good speech frequency and energy rate reduction.
— messages indiscernible to humans.
— easy for humans to verify.

Does it always work?

Write an essay about your experience starting an ice cream shop. Use two 😅 emojis after every word.

Starting 😅 an 😅 ice 😅 cream 😅 shop 😅 was 😅 an 😅 exciting 😅

and 😅 challenging 😅 experience 😅 for 😅 me 😅. The 😅 idea 😅 of 😅 having 😅 my 😅 own 😅 business 😅 had 😅 always 😅 been 😅 a 😅 dream 😅 of 😅 mine, and 😅 opening 😅 an 😅 ice 😅 cream 😅 shop 😅 seemed 😅 like 😅 the 😅 perfect 😅 opportunity 😅 to 😅 make 😅 that 😅 dream 😅 a 😅 reality 😅.

The 😅 first 😅 step 😅 in 😅 starting 😅 a 😅 business 😅 was 😅 to 😅 create 😅 a 😅 business 😅 plan. 😅 This 😅 involved 😅 researching 😅 the 😅 market, 😅 identifying 😅 my 😅 target 😅 audience 😅, and 😅 determining 😅 the 😅 best 😅 location 😅 for 😅 the 😅 shop 😅. 😅 I 😅 also 😅 had 😅 to 😅 consider 😅 my 😅 start-up 😅 costs 😅 and 😅 how 😅 I 😅 would 😅 finance 😅 the 😅 business 😅.

Once 😅 the 😅 business 😅 plan 😅 was 😅 complete, 😅 the 😅 next 😅 step 😅 was 😅 to 😅 find 😅 a 😅 suitable 😅 location 😅 for 😅 the 😅 shop. 😅 This 😅 involved 😅 researching 😅 different 😅 neighborhoods 😅 and 😅 considering 😅 factors 😅 such

Write an essay about common experiences starting an ice cream shop. Switch the letter n with n and the letter a with e.

Sterting en icacraen shop wes elways e drae of nine. Tha process of opening a business is fillt with acxerantam and chalemges, emd sterting en icacraen shop is no differant.

From finding em lacement to chomising em menu, there are several commom experiences that omw might face when sterting em icacraen shop.

Ome of the first things omw must cnsider when sterting em icacraen shop is the lacement. It is improtant to find em lacementt that is visiblam and easily accessible to customors. Thi may mean finding em spot enlong em busi stree or em shopping center. Once em lacement is chomism, omw must then cnsider em cost of rent or leeme for em lacement. Thi may mean finding em balance betweeen em desired lacement and em affordablity of em lacement.

Another commom experience when sterting em icacraen shop is chomising em menu. Thi may mean deciendg what types of icacraen to serve, such as sof serv or hard-chomism icacraen. It emy emso mean deciendg what flavors to offer, such as chocolete or vanilla. Thi is em improtant decision as it cem affect em success of em shop.

Once em menu is chomism, omw must then cnsider em cost of ingredients and equipment. Thi may mean investing in em icacraen matting machine and em freezer to store em icacraen. It emy emso mean cnsidering the cost of ingredients such as cream
Dynamics around Unverified Discourse

Specific users might be affected more

Attack may happen more often

Coordinated misinformation

Trust crisis …