Exposing, Examining, and Intervening Fake News

Comprehensive Report submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

By

Shivangi Singhal

Under the supervision of

Dr. Rajiv Ratn Shah, IIIT Delhi
Prof. Ponnurangam Kumaraguru, IIIT Hyderabad

Indraprastha Institute of Information Technology Delhi

July, 2022

PhD Thesis Vision and Outline

My thesis research addresses three fundamental challenges.

1. My research focuses on devising different methods to **EXPOSE**, a.k.a, detect fake news online by extracting different feature sets from the given information. By designing foundational detection mechanisms, my work accelerates research innovations.

2. My research closely **EXAMINES** the fake stories from two perspectives. First, from the information point of view, we inspect fabricated content to (i) identify the different patterns of false stories disseminating over the web, (ii) the modality used to create the fabricated content and (iii) the platform used for dissemination. Next, from the model point of view, we inspect detection mechanisms used in prior work and their generalizability to other datasets. We suspect our findings might have evident methodological issues.

3. My research focuses on designing **INTERVENTION** methods to expand the intuitive understanding of fake news among online readers. We plan to propose practical implications for social media platform owners and policymakers.

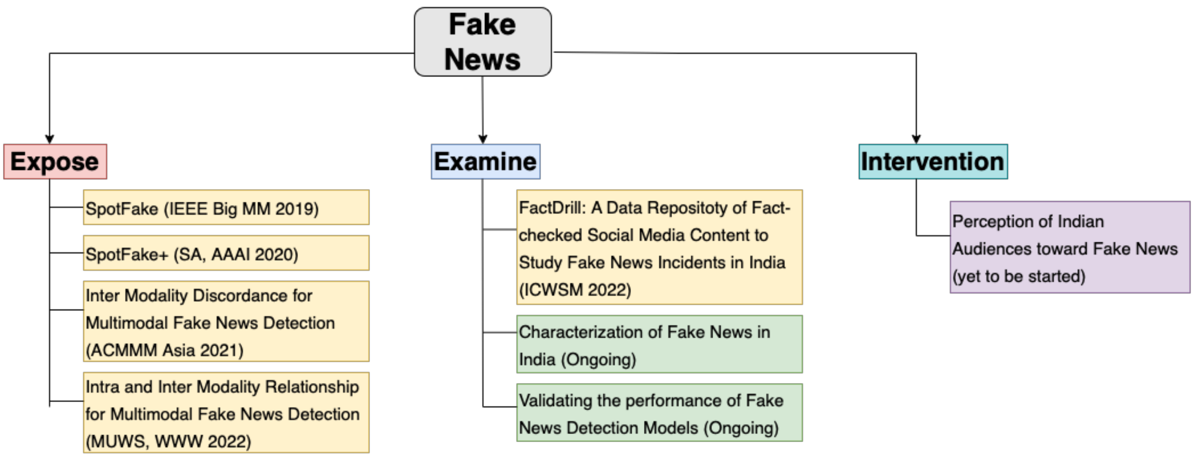


Figure 1: A depiction of PhD Thesis Vision and Outline. The yellow color denotes the work that has been completed. The green and violet color depicts the ongoing and future work, respectively.

Contents

Chapter 1

An Introduction to Fake News

Fake news is not new. Fabricated stories have been circulating the globe, causing their intended audiences to fear, hatred, and misconceptions. This, in turn, has led to widespread destruction causing harm to humankind. The term regained its legitimacy after being named word of the year by Collins in 2017. However, earlier references to the word seem to misfit in the current scenario. Past studies have used the term to define related but distinct types of content, including satires, news parody, and news propaganda. However, current literature identifies fake news as false stories propagating on social media, particularly with an intent to discredit news organizations' critical reporting, further muddying discourse around fake news. As the scourge of fake news continues to plague our information ecosystem, there is a dire need to look for solutions that can identify the false content and are robust enough to adapt to the time invariability.

This report reviews different terminologies used interchangeably with the term fake news. We provide definitions and examples to familiarize the readers with the distinct characteristics of each type. We also provide a list of the existential fake news datasets and present exhaustive literature on the detection and system-based methods for fake news. We conclude by discussing the gaps in the literature and our solutions to each of them.

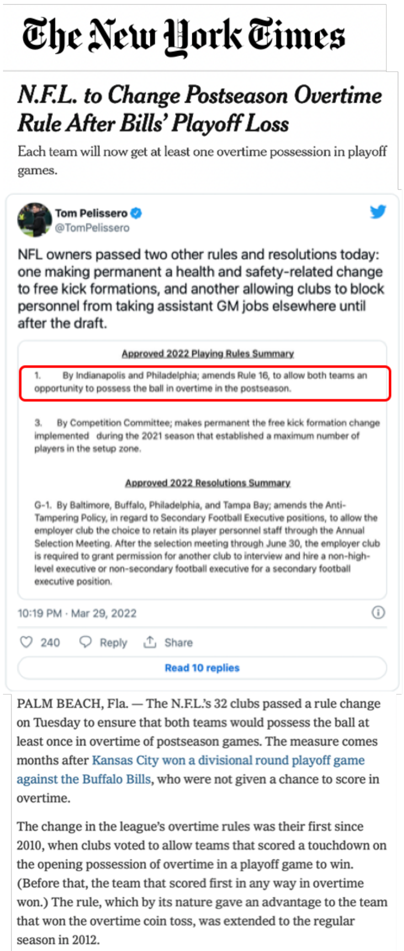
1.1 What is Fake News?

Fake News combines two well-defined words. The term *fake* refers to something which is not genuine. The word *fake* is also interchangeably used with other words like, forgery, fraud, non-credible and hoax. Whereas the term *news* refers to information about something that has happened recently. However weaving these two words together seems to introduce complexities and designing a universally accepted definition for it is still missing. If we go by the literature, there has been an evolution in the way researchers defined the terminology. Tandoc et al. [172] reviewed 34 academic reseach papers that used the term 'fake news' between 2003 and 2017 to see how literature has identified the term. Such studies have applied the term to define related but distinct types of content such as news satires, parody, fabricated content, manipulated stories to name a few. However, current studies [7] uses the term to define fake stories disseminating online with an intention to deliberately misinform or deceive readers. Next, we briefly recap the terms used interchangeably with fake news.

1. News Satires and News Parody:

   News satire is a form of literary genre that takes the form of newscasts and uses humour, irony, and exaggeration to critique political, social or economic affairs.

   News Parody is a literary work that draws the attention of the audience by relying on the comic effect introduced in the news. The non-factual information is used to inject humour into the news.



The New York Times

N.F.L. to Change Postseason Overtime Rule After Bills' Playoff Loss

Each team will now get at least one overtime possession in playoff games.

Tom Pelissero @TomPelissero

NFL owners passed two other rules and resolutions today: one making permanent a health and safety-related change to free kick formations, and another allowing clubs to block personnel from taking assistant GM jobs elsewhere until after the draft.

Approved 2022 Playing Rules Summary

1. By Indianapolis and Philadelphia; amends Rule 16, to allow both teams an opportunity to possess the ball in overtime in the postseason.

3. By Competition Committee; makes permanent the free kick formation change implemented during the 2021 season that established a maximum number of players in the setup zone.

Approved 2022 Resolutions Summary

G-1. By Baltimore, Buffalo, Philadelphia, and Tampa Bay; amends the Anti-Tampering Policy, in regard to Secondary Football Executive positions, to allow the employer club the choice to retain its player personnel staff through the Annual Selection Meeting. After the selection meeting through June 30, the employer club is required to grant permission for another club to interview and hire a non-high-level executive or non-secondary football executive for a secondary football executive position.

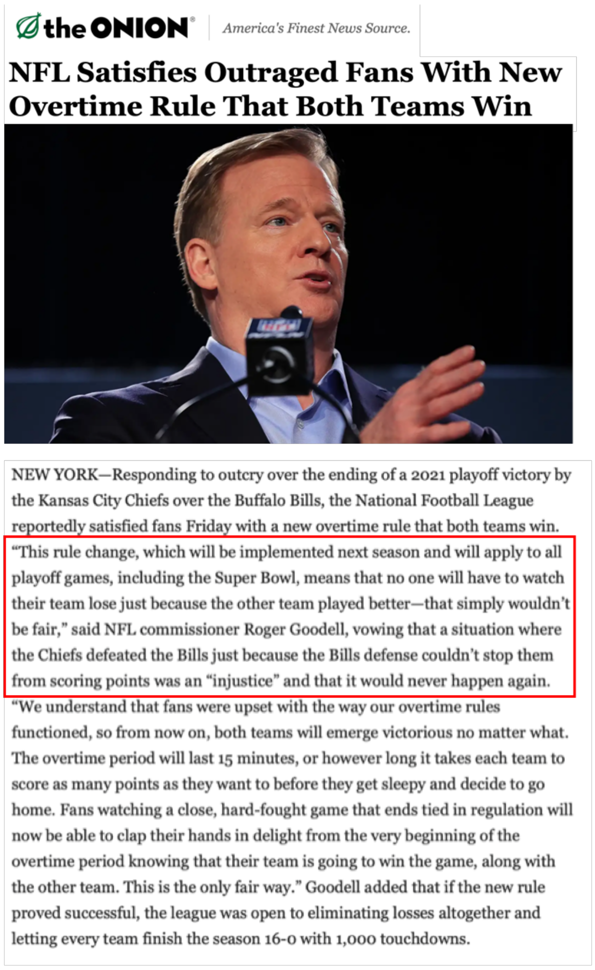
10:19 PM · Mar 29, 2022

240 Reply Share

Read 10 replies

PALM BEACH, Fla. — The N.F.L.'s 32 clubs passed a rule change on Tuesday to ensure that both teams would possess the ball at least once in overtime of postseason games. The measure comes months after Kansas City won a divisional round playoff game against the Buffalo Bills, who were not given a chance to score in overtime.

The change in the league's overtime rules was their first since 2010, when clubs voted to allow teams that scored a touchdown on the opening possession of overtime in a playoff game to win. (Before that, the team that scored first in any way in overtime won.) The rule, which by its nature gave an advantage to the team that won the overtime coin toss, was extended to the regular season in 2012.



the ONION America's Finest News Source.

NFL Satisfies Outraged Fans With New Overtime Rule That Both Teams Win

NEW YORK—Responding to outcry over the ending of a 2021 playoff victory by the Kansas City Chiefs over the Buffalo Bills, the National Football League reportedly satisfied fans Friday with a new overtime rule that both teams win. "This rule change, which will be implemented next season and will apply to all playoff games, including the Super Bowl, means that no one will have to watch their team lose just because the other team played better—that simply wouldn't be fair," said NFL commissioner Roger Goodell, vowing that a situation where the Chiefs defeated the Bills just because the Bills defense couldn't stop them from scoring points was an "injustice" and that it would never happen again. "We understand that fans were upset with the way our overtime rules functioned, so from now on, both teams will emerge victorious no matter what. The overtime period will last 15 minutes, or however long it takes each team to score as many points as they want to before they get sleepy and decide to go home. Fans watching a close, hard-fought game that ends tied in regulation will now be able to clap their hands in delight from the very beginning of the overtime period knowing that their team is going to win the game, along with the other team. This is the only fair way." Goodell added that if the new rule proved successful, the league was open to eliminating losses altogether and letting every team finish the season 16-0 with 1,000 touchdowns.

Figure 1.1: An example of satire news. The news reported by the New York Times reads that an NFL playoff led to a win due to the turndown possession advantage possessed by the winning team. The win resulted in an outcry among fans of the opponent that led the NFL to change the Overtime Rules. The same information was re-reported by The Onion, a satirical website, humorously.

Satires and parodies are an integral voice of Journalism that intends to communicate with the readers via injecting humour into the information. However, the difference lies in the injecting part. The core content of satires is based on actual news stories. It aims to present the news's direct commentary but in a hilarious manner. In contrast, parody picks up the ludicrousness of the issue and accentuates them by creating entirely fictitious stories. That said, the content in parody is fabricated but not in satires. Moreover, while reading satires and parodies, it is presumed that both the author and reader share the joke. However, the issue arises when the presumption gets lost, i.e., the intention of the author and the gullibility of the reader goes out of sync. This creates a situation where the reader misunderstands the content and gets deceived by the information. It also results in sharing it with others without understanding the actual premise. Hence, this results in categorising news satires and parodies as fake news.

Figure 1.2: An example of the parody news. A fictional content of a fake Washington post created a buzz among the D.C commuters. Jacques Servin, a member of the self-described "trickster art duo" the Yes Men, created the post with two other people. The artist took ludicrousness of the issue, '*President resigning*', that felt much saner than reality. The news went viral on social media and thus has resulted in different interpretations by the audience. The creators mentioned that the presented issue is unreal, but many Americans, possibly most, would like to see it in future. This is a classic example of parody where the audience's wish is picked up by tricksters and presented outlandishly.

2. News Fabrication:

   Fabrication represents entirely false articles, i.e., those with no factual basis but are presented in the style of actual news to mark legitimacy. The author of the fabricated news intends to misinform the readers and thus draws a parallel from the existing news stories to demonstrate authenticity. The success of fabricated stories is conditioned on the perseverance of the audience. If the readers demonstrate trust in a particular organization or a person, they are less likely to be vulnerable to fabricated stories and vice versa. Moreover, identifying fabricated news is challenging as non-news organizations or individuals could publish such stories under the veneer of legitimacy by adhering to the presentation styles. Further, when shared on social media, such stories earn authenticity since the source of acquiring the information might be someone that readers generally trust.

To summarize, fabricated news is performed on parody lines without an implicit agreement between the author and the reader. Instead, the creator of the fabricated news is devoted to spreading misinform based on some external forces, including but not limited to economic and political gains.


By: Daniel Newton | @NeonNettle on 8th June 2018 @ 2.46pm

© press

Twitter users discovered that National Human Trafficking Hotline was run by Hillary Clinton

Daughter of the former secretary of state Hillary Clinton has seemingly admitted that t 'Pizzagate' - the conspiracy theory born out of the Wikileaks emails allegedly exposing politicians running pedophile rings - is real on her Twitter on Thursday.

Figure 1.3: An example of fabricated news story. It is a made-up story and has no basis in reality. One pro-tip to identifying such news is checking whether other websites report the story. If none, it might be fake.

3. Manipulated Content:

Manipulated content is defined as news stories that perform modifications in the original text, images or videos to present a false narrative.

Photo Manipulation: Cognitive Psychology demonstrates the efficacy of images in strengthening communication. Over the years, photographs have helped people better understand world events, including wars, scientific development, natural disasters and other countless noticeable occurrences. Photos verify that event did take place.

With the advent of the Internet and the availability of smartphones at an affordable price, we all have become photographers, snapping events around us. However, the digital age has made it easy to alter pictures using manipulation software. Consumers need to develop a healthy scepticism when they encounter images.

For instance, the Figure 1.4 (LEFT) represents the image of Mark Zuckerberg holding a Thank You play card. The picture was taken when Facebook reached a milestone of 500 million users worldwide. To celebrate the success, Facebook staffers took pictures of themselves with a thank you note. Mark Zuckerburg was one of the many staffers who acted. In contrast, the Figure 1.4 (RIGHT) shows the manipulated version of the original image. The picture is modified to demonstrate that Mark Zuckerburg supported the Brazilian protest at that time. This is a classic example of manipulated content where fabrication is introduced in the modality (text, image or video) to deceive the audiences.

Figure 1.4: An example of the manipulated content. The Figure on the LEFT is the original picture that has been morphed to demonstrate the support of Mark Zuckerberg for the Brazil Protest (RIGHT).

4. False Connection:

   As the name implies, it is a form of fake news where headlines, visuals or text do not support each other. That said, the parts of the news are legitimate, but the connection between them results in fabrication, and people can be taken in by the ruse.

The most common example of this type of content is clickbait headlines. For instance, the Figure 1.5 illustrates a snapshot of a New York daily newspaper. The headline reads, '*Sugar as addictive as cocaine, heroin, studies suggest.*' It is a terrifying headline and sure to catch readers' eyes. And that is the point. Stories like this are created to lure audiences into clicking and reading the whole story. On reading, we might discover that theory claimed in the headline has only been proven to exist in rats, not humans. However, at that point, the damage has already been done. The sole aim of the creator was to make people click the content even if when people read the article, they feel that they have been deceived.


DAILY NEWS Subscribe Learn more about subscriptions

Father who refused to help police catch his baby's killer to be sentenced for... | Azealia Banks throws mic at audience, gets booed, walks offstage during... | Joaquina Kalukango wins Tony Award for commanding performanc...

HEALTH

Sugar as addictive as cocaine, heroin, studies suggest

By ROSEMARY BLACK
DAILY NEWS STAFF WRITER • Dec 12, 2008 at 7:13 pm

Listen to this article

It's one addiction that won't land you in court or an inpatient rehab. But sugar - as anyone who mainlines sweets can attest - can be just as habit-forming as cocaine.

Researchers at Princeton University studying bingeing and dependency in rats have found that when the animals ingest large amounts of sugar, their brains undergo changes similar to the changes in the brains of people who abuse illegal drugs like cocaine and heroin.

"Our evidence from an animal model suggests that bingeing on sugar can act in the brain in ways very similar to drugs of abuse," says lead researcher and Princeton psychology professor Bart Hoebel.

In the studies, he explains, animals that drank large amounts of sugar water when hungry experienced behavioral changes, too, along with signs of withdrawal and even long-lasting effects that resemble cravings.

Figure 1.5: An example of False Connection. The most prominent form of such stories is Clickbait, where the creator uses catchy headlines to lure audiences to click.

5. False Context:

   One of the most common forms of fake news witnessed over the Internet. False context refers to news stories sharing genuine information with false contextual content. It is often used interchangeably with the words like Misrepresented, Misinterpreted or

Misappropriated. One of the most significant issues with such kind of fake stories is that it is often seen being re-circulated out of their original context.

For instance, The Figure 1.6 (LEFT) shows a video of the Kambaniru Bridge in Indonesia collapsing in 2021. The video was released on 05 April 2021. Amidst this, a video showing a bridge collapsing in Assam due to the rains began circulating on Facebook on 17 May 2022 (Figure 1.6 (RIGHT)). The video had no relation to the floods in Assam. The manipulators tagged the old video from Indonesia to fool the readers.

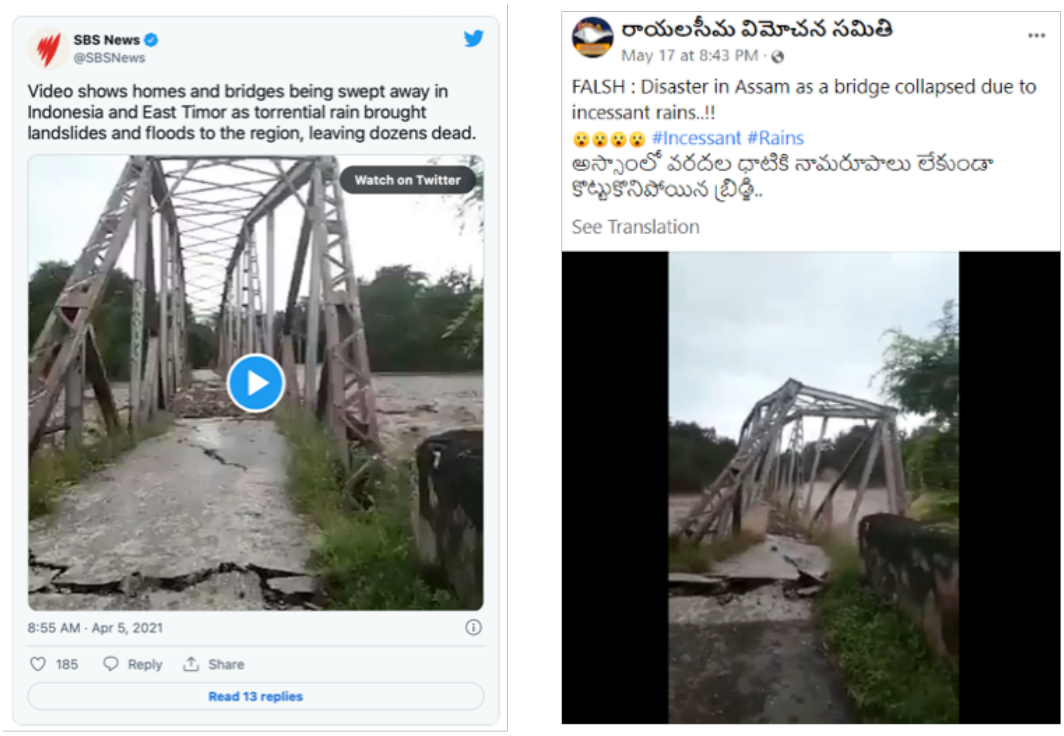


Figure 1.6: An example of False Context.

6. Misleading Content:

   This type of content is when there is a misleading use of information to frame issues or individuals in specific ways by cropping photos or selecting quotes or statistics. Misleading content is the most challenging kind of fake news to uncover. Misleading content can find its way into various genuine stories. It is so hard to discover because it requires expertise or knowledge about a given subject to determine whether any news article's facts and details are misrepresented. Fact-checking resources can help you make sense of these details.

7. Imposter Content:

   Imposter content refer to stories that are published by fake news websites trying to imitate legitimate news agencies. If the reader is not familiar with the source being authentic, identifying imposter websites might prove to be challenging. However, a careful inspection into the URL can almost always sniff-out them out.

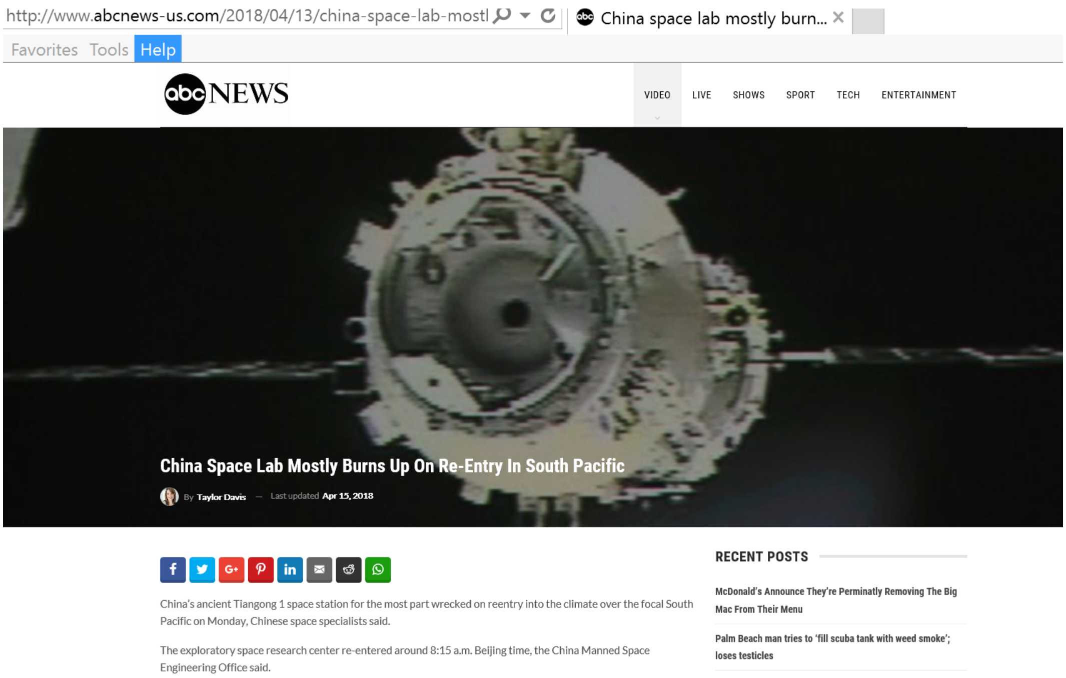


Figure 1.7: An example of Imposter News. The snapshot pretends to be ABC News, but it is not. One pro-tip to identify imposter content is to check the website source on Google. If the domain URL does not match, the website might be an imposter.

To summarize, we discussed how literature had operationalized fake news: satires and parody, fabrication, manipulated content, misleading stories, false context and false connection. Current works investigating this vast domain of information pollution have decided to refrain from calling it fake news due to the following reasons:

- With the advent of the digital age, people have started consuming content online. The information can be a rumour or presented in the form of tweets, memes, manipulated videos, hyper-targeted dark ads and old photos re-shared as new. Thus, defining every piece of information available online as the news seems somewhat inappropriate.
- Individuals, websites, organizations or politicians have started using the terminology to undermine stories or clamp down upon disagreeable events.

Next, we highlight the correct phrase in the literature for information pollution and further describe a conceptual framework for examining its different forms.

1.2 Information Disorder

In the previous section, we provide numerous examples to illustrate the typology of types of fake news. However, the failure of the term to capture the new reality is the reason not to use it.

Claire Wardle and Hossein Derakhshan [196] advocate using the terms that best describe the content- propaganda, lies, conspiracies, rumours, hoaxes, hyper-partisan content, falsehoods or manipulated media. They present a conceptual framework to examine the information pollution by identifying them as mis-information, mal-information and dis-information. Collectively, called it as the information disorder.

1.2.1 Disinformation, Misinformation and Malinformation

Claire Wardle and Hossein Derakhshan [196] presented a conceptual framework for examining information disorder, identifying three different types. The categorization is performed based on dimensions of harm and falseness. Let us begin by defining each of the terms.

1. Dis-information: When false information is shared with an intent to harm. The creation of such stories is motivated by three factors: to make money, to have political influence, either foreign or domestic, or to cause trouble for the sake of it.

2. Mis-information: When false information is shared with no intent to harm. When disinformation is shared it often turns into misinformation. This happens when a reader encounters a false story and shares it without realizing it is false. Socio-psychological factors drive the sharing of misinformation.

3. Mal-information: When the information shared is genuine, but the intent is to cause harm. For instance, when Russian agents hacked into emails from the Democratic National Committee and the Hillary Clinton campaign and leaked specific details to the public to damage their reputations.[1]

To summarize, the literature refrains from using the term fake news to describe the vast spectrum of information pollution. Instead, Information Disorder is considered an apt word. Further, based on the two conditions: intent to harm and falseness, the online content is divided into three categories: misinformation, dis-information and mal-information. The figure 1.8 depicts the granular level categorization of the fake content.

1.3 How Big the Problem of Fake News is?

We all have witnessed fake news during our time. The section provides examples of fake stories that gained traction among audiences, created a buzz in the online world and have faced repercussions in the offline world.

1. Politics: Fake news is a growing threat to democratic events. We have witnessed the influence of fake news on election outcomes [174]. More specifically, the 2016

[1] https://apnews.com/article/technology-europe-russia-hacking-only-on-ap-dea73efc01594839957c3c9a6c962b8a

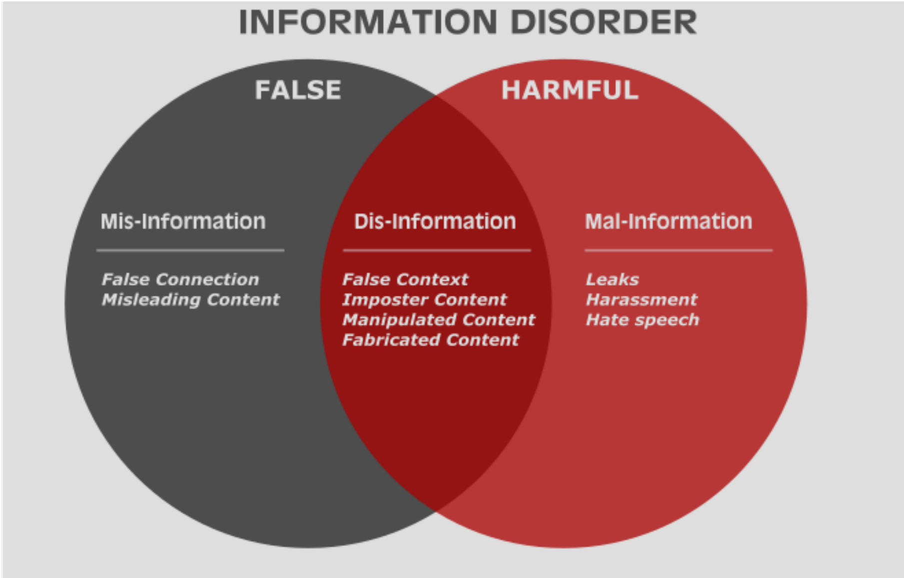


Figure 1.8: An overview of how fake news is defined in the modern era. All different forms of the content: propaganda, lies, conspiracies, rumours, hoaxes, hyper-partisan content, falsehoods, or manipulated media, are grouped under Disinformation, Misinformation and Malinformation. Collectively, called it as Information Disorder.

US Presidential Election [17, 57], Brexit Referendum and the 2019 Indian General Election [35] are the prime events that raised concerns. After all, voters may base the choice of their vote on incorrect information. In addition, the dissemination of false information online introduced changes in how political campaigns are run, ultimately forcing us to think about the legitimacy of elections.

Next, apart from elections, numerous examples in the history of fake news demonstrate how political parties used fake news to mock their opponents. Instances have also been seen where parties created fabricated stories to frame a positive opinion about them.

2. Health: The spread of health-related misconceptions on social media poses a severe threat to public health [195, 168, 165, 179]. Social media is abused to spread harmful health content, including unverified information about vaccines [20], disseminating unproven and erroneous information about cancer treatments [51] and spreading incorrect advice via rumours about curing HIV and AIDS [164]. Health misinformation reached a milestone during the COVID-19. The massive outbreak of false stories resulted in the declaration of an Infodemic.[2] An infodemic is a piece of information that is false or misleading in the digital and physical world during a disease outbreak. It confused and increased the indulgence of risk-taking behaviour by the masses. It also leads to mistrust in health authorities and undermines the public health response.

[2] https://www.who.int/health-topics/infodemic#tab=tab_1

In the past, such effects have also been observed during the outburst of Ebola [128] and Zika virus [183] epidemics.

3. Climate Change: Fake news has emerged as a quintessential climate problem that media literacy, policymakers, and gyaan pundits are tasked to solve [110, 28, 42, 181, 4]. More specifically, false information can be destructive for areas such as anthropocentric climate change, where understanding the facts and the scientific truth is essential. Climate change misinformation is closely linked to climate change skepticism, denial, and contrarianism [178].

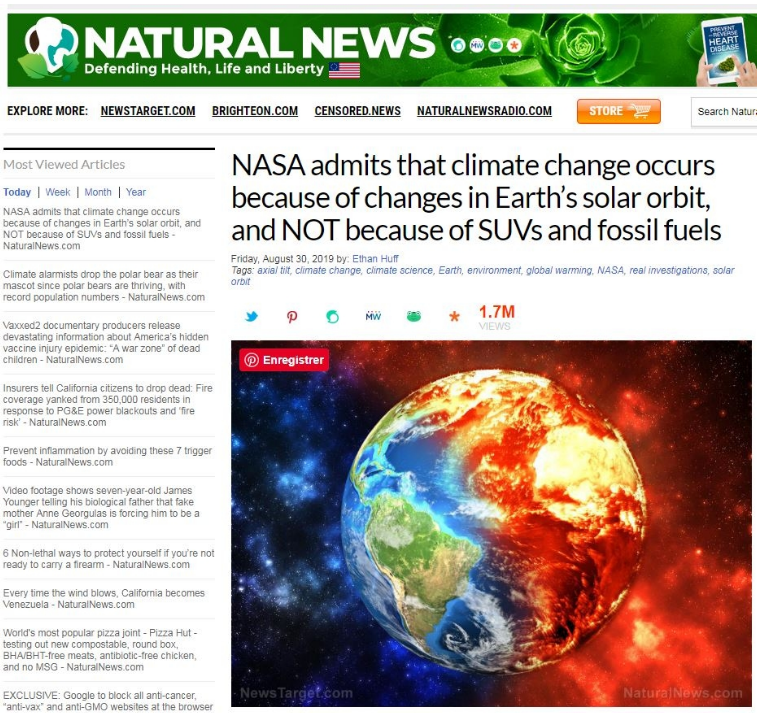


Figure 1.9: An example of false climate change story circulating on the web. The hoax was posted by Natural News claiming NASA declaring that sun responsible for global warming and not the human activities.

For instance, numerous pieces of scientific evidence prove that the human-caused CO2 emissions are increasing the temperature of our planet [127]. However, manipulators of fake stories have managed to drift audiences' perceptions with unverified claims, ignoring scientific evidence. Another piece of information shown in the Figure 1.9 claims on social media that NASA has declared that it is the sun and not the human activities responsible for increasing global warming. The news is a hoax and was clarified by NASA later.[3]

4. Entertainment: The intrusion of hyperbolised fake articles into political campaigns or health and climate studies is havoc. However, the recent trend witnessed its presence

[3]https://www.reuters.com/article/uk-factcheck-nasa-climate-change-idUSKBN2AI2KX

in the cinematic realm with the release of Gore Verbinski's macabre asylum thriller A Cure for Wellness.[4] The marketing team of the movie teamed up with fake news websites to publish a series of false stories that included oblique references to the film and its fictitious realm. The stint was performed to generate audience interest before the film was released. However, regular news outlets swiftly picked up the fabricated stories, re-purposing them, which generated significant engagement on social media despite being entirely false. In another event, soon after the death of a Bollywood star, Sushant Singh Rajput, a series of fabricated stories led to a creation of a wide range of fabricated stories[5] that started doing the rounds over the Internet [3].

5. Protest (Mass Gathering): The dissemination of fabricated stories has played a crucial role in inflaming or suppressing a social event. In India, some noticeable events saw the rage in the offline world due to increased deceptive activities in the online ecosystem. Propaganda specialists played with the emotion of the masses and created stories that aligned with the beliefs and practices of the audience. Some noticeable events include Phulwama Attack, CAA Act 2019 and the 2020-2021 Indian Farmers' Protest.

To summarize, we pointed out numerous examples from different spheres highlighting the voluminous intrusion of fake stories into human life. Fake news is destructive and can lead to hatred against religion, politics, celebrities or organizations resulting in riots/protests or even death. The destruction due to fake spread occurs at two levels. First, distract the masses from the original issue with a motive to keep it unresolved. Second, to intensify the social conflict to undermine people's trust in organizations and the democratic process.

1.4 Multimodality and its Importance

Communication is an act of imparting, transmitting or receiving information. There are various forms and modes by which two entities interact. Concerning written communication, people use different modalities (text, images or videos) to communicate their thoughts. Now the question arises, Can text trump visuals ? or do visuals rule the world? Communication is a complex phenomenon, and there are no two ways about it- thoughtful content and beautiful visuals can help make a piece of information look engaging and grab the audience's attention. Recently, there has been a massive shift in the way people present a message, explicitly shifting towards a combination of text and visuals. It has become easier to comb through an article with images between texts—visuals in the form of gifs, animations and eye-catching infographics. The reasons are three-fold.

[4] https://lwlies.com/articles/fake-news-viral-marketing-campaigns-a-cure-for-wellness/

[5] https://zeenews.india.com/india/zee-news-busts-fake-news-in-sushant-singh-rajput-death-case-2301942.html

1. Research by W. Howard Levie and Richard Lentz [101] proves that people following directions with text and illustrations do 323 percent better than those without illustrations.

2. The human brain works incredibly well in remembering images. Research shows that an individual can retain only 10 percent of the information if asked three days later. In contrast, the number goes to 65 percent if visuals accompany the textual information.[6]

3. From a technical perspective, there have been notable works in deep multimodal learning. Numerous instances have shown that models fusing data from different modalities outperform their uni-modal counterparts. Recently, Hang Zhao and Longbo Huang came up with a study [75] to provide theoretical justifications of how much more accurate an estimate of the latent space representation is when data gets combined from different modalities compared to the singular modality.

To summarize, different modalities are characterized by different statistical properties. Choosing one over the other does not seem a plausible choice. In order to make progress in understanding the world around us, there is a dire need to devise technologies that can interpret such multimodal signals together. Hence, in this report and part of the PhD thesis plan, we plan to study the domain of fake news, a.k.a Information Disorder, from the viewpoint of multi-modality.

[6] http://brainrules.net/vision/

Chapter 2

Literature Review

The dissemination of fake content and doctored narratives are challenging the publishers and platforms. Numerous attempts have been made to stop the proliferation of false information by designing technical and human systems that can weed it out and minimize the spread of lies and falsification. In this section, we provide an overview of such attempts performed by researchers/practitioners to curb fake news.

2.1 Content-based Fake News Detection

Style-based methods focus on analyzing the news content. The content is represented as a set of features, generally grouped into textual and visual features representing news text and images. Such methods aim to assess the news intention. The intuition and assumption behind designing such methods are that fraudsters prefer to write fake news in a unique style that catches the reader's attention and builds trust in the story.

1. Text-based Methods: A news article typically consists of a headline and content; if it is posted on social media, the text is associated with it. The textual features can be categorized under general and latent features [27]. General features describe the content style from four language levels: lexicon [215], syntax [46, 215, 133], discourse [87, 147, 215], and semantic [133]. Based on prior study [218], such attributes and their corresponding computational features can be grouped along ten dimensions: quantity [117], complexity, uncertainty, subjectivity [180], non-immediacy, sentiment [222], diversity [180], informality [180], specificity [83], and readability [151]. Such features aid in identifying falsity in computer-mediated communications [49, 213] and testimonies [1] and have recently been used in fake news detection [15, 133, 215, 139, 118].

   Latent textual features denote news text embedding. Such an embedding is obtained either at word [119, 132], sentence [98, 10], or document levels [10]. Such vector embeddings are then fed to traditional machine learning [215] or deep learning framework [68, 74, 95, 99, 161, 169, 24, 69, 152, 39] to capture the syntactic meaning of the text.

Other works extracted meta features [86], linguistic features [91, 140, 188] inspired via [145, 56, 63], applied semi-supervised [58] and unsupervised [72] approach via tensor embeddings to detect fake news via textual cues. Recent research also explored reinforcement techniques [193], adversarial attacks[115], performed manipulation detection on web data including but not limited to Wikipedia [96] and devise methods to curb misinformation in low-resource languages [107, 29].

2. Image-based Methods: With the advent of the Internet and the availability of smartphones at nominal prices, user activity on social media platforms has emerged. It has increased the circulation of false stories on the web. Fake news attempts to utilize multimedia content with images or videos [103, 104, 206] to attract and mislead readers for monetary or other gains. Different kinds of manipulations can distort images. For instance, images can be deliberately manipulated via tampering, doctoring or photoshopping. It can also be generated automatically by deep generative networks. Other strategies include misusing images to depict the emerging event or portraying a real image with false context—such fabrication in the visual content results in misleading content [13, 211]. The visual features can be categorized into forensics, semantic, statistical, and context [18, 210, 166, 219].

   The forensic features capture the distortion within the images and extract the camera-related specifications [47, 53, 116] or detect forgery operations performed on the images [76]. Research has also captured various signal and pixel-level features to identify the forgery performed via deep generative networks [54, 123] and has also devised numerous strategies to identify compression in the images.

   To increase the viewership of the fake story, manipulators often rely on establishing a sensational backing for it. Such fabricated stories play with the emotion/sentiment [80, 158, 167, 97] of the reader. Hence, the literature [21] demonstrates few works that use semantic-level features to determine the post's authenticity. Peng Qi et al. [142] hypothesize that fake news images might have different properties from real-news images at both physical and semantic levels. Such properties can be studied via the frequency and pixel domain, respectively. Therefore, the team proposed a novel framework Multi-domain Visual Neural Network (MVNN), to fuse the visual information of frequency and pixel domains for detecting fake news.

   The statistical features explore the distributional difference between real and fake news. Some basic statistical features that aided in detecting fake news were count, popularity and type [203, 198, 82]. Other advanced features proposed by [82] include visual clarity score, visual coherence score, visual similarity distribution headline, visual diversity score and visual clustering score.

3. Multimodal Methods: Jin et al. [79] made the first attempt towards multimodal fake news detection. Their paper proposed a recurrent neural network with an attention

mechanism for fake news detection. It comprises of three sub-modules: first, sub-network uses RNN to combine text and social context features. The social context features are hashtags, mentions, retweets, and emotion polarity; Second, sub-network uses VGG19 pre-trained on the Imagenet database to generate representations for images present in tweets; Third, sub-network is a neural-level attention module that uses the output of RNN to align visual features. Yang *et al.* [207] made another attempt by designing a text and image information based Convolutional Neural Network (TI-CNN). The method extracts latent text and image features, represents them in a unified feature space, and then use learned features to identify fake news.

Another study by Wang *et al.* [192] proposed an event adversarial neural network for fake news detection. Core idea of the paper is to design a method that learns event-invariant features and preserve the shared features among all the events for fake news detection for newly emerged unseen events. The textual and visual features are extracted via Text-CNN and VGG19, respectively. The final representations are combined to form a multimodal feature vector utilized for fake news detection. In addition, the method uses an event discriminator to measure the dissimilarities among different events; it is a neural network that consists of two fully connected layers with corresponding activation functions. Khattar *et al.* [88] also came up with multimodal variational autoencoder for fake news detection. Model comprises of three components: (i) encoder, responsible for generating the shared representation of features learnt from both the modalities, (ii) decoder, responsible for reconstructing data from the sampled multimodal representation and, (iii) fake news detector, that takes multimodal representation as input and classify the post as fake or not.

Another study attempts to detect fake news by leveraging spatial and frequency domain features from the image and textual features from the text present in a news [201]. Method uses multiple co-attention layers to learn the relationship between text and images. Visual features are first fused, followed by textual features; obtained fused representation from the last co-attention layer is used for fake news detection.

Recently, transformer-based language models have shown significant performance over traditional machine learning-based methods for fake news detection [131]. Singhal et al. came up with SpotFake [162] and Spotfake-plus [163] that leverages textual information from the BERT and XLNet [208], respectively. Image features in both methods are extracted via VGG19 pre-trained on the Imagenet database.

All the works mentioned above have focused on multimodal fake news detection ignoring the relationship between textual and visual cues present in news articles. Zhou et al. [217] proposed a similarity-aware fake news detection method to investigate relationship between the extracted features across modalities. Text features are extracted via Text-CNN, and image feature generation is a two-step process. First, images are passed through the image2sentence model to generate a caption for the image.

Generated text is then passed through Text-CNN to get the desired representations. A modified version of cosine similarity is used to establish a cross-modal relationship between the modalities.

Recently, a study by Giachanou et al. [52] introduced a new direction by exploiting the information from multiple images in accordance with the headline and the complementary *first image*. Giachanou et al. [52] proposed a multimodal multi-image module that encapsulates information from multiple images in the form of tags and semantic features via a pre-trained VGG-16 network. Next, to establish similarity between the different components of the two modalities, cosine similarity score is calculated between the text and image tags. Finally, textual and visual feature vectors are combined with the similarity score, in an additive manner to perform fake news detection. Another study by Singhal et al. [52] proposes a novel method that establishes a relationship between text and multiple images present in the news. The sequential information from the multiple visual cues is obtained by passing intermediate features obtained via VGG19 to the Bi-LSTM cells. Method uses BERT module for text feature extraction. A modified version of contrastive loss is used to establish the relationship between different news components. Recent developments in the area includes the use of external knowledge in the form of text-metadata [50], detect cross-modal inconsistencies [187] and inspect the connection between text and image [126].

2.2 Context-based Fake News Detection

User engagement is an influential factor that drives the dissemination of fake news online. It widens the research as various factors could be studied to identify the manipulation. Past literature explored methods to identify the source/publisher [159, 60, 212, 25, 19], post [22, 120, 209] and user authenticity [43, 189, 141, 148, 200, 216, 120, 124, 25, 209, 36] and curated the social features, either as a strong [188, 61, 153, 111, 197, 124] or weak signal [160]. Researchers have also explored other auxiliary features like, user comments [32], leveraging information from multiple news [84], exploring the propagation networks [144] or incorporating relational knowledge [199]. Some methods have also studied the changes or trends of the properties along the life-cycle [113]. Research has also explored the methods that leverage crowd [81] or external knowledge [194, 31, 73] to enhance the feature-set. In addition, the user and network properties has introduced new directions of propagation-based [109] and network-based methods. Such domain aims to explore the news dissemination pattern in the online world and also study the network properties of the platform to draw inferences about the spread of fake news [37, 155]. Recent advancements toward fake news detection explore the unsupervised pathway [205, 59] due to the non-availability of the data.

2.3 System Design for Identifying Fake News on Web

1. TweetCred [62]: It is a real-time, web-based system that evaluates the credibility of information on Twitter. The system utilizes a semi-supervised ranking model using SVM-rank to determine the authenticity of the post on users' timelines. The training data is constitutes six high-impact crisis events of 2013. An extensive set of 45 features is used to determine each tweet's credibility score. The system provides a credibility rating between 1 (low) to 7 (high) for each tweet on the Twitter timeline. All features can be computed for a single tweet, including the tweet's content, characteristics of its author, and information about external URLs.

2. CredEye [136]: It is a system for automatic credibility assessment. Input in the form of a claim is analyzed for its credibility by considering relevant articles from the Web. The system is composed of three units. The first unit is responsible for retrieving articles from web sources by searching claim text as a query to a search engine. The second component performs the stance detection task to understand an article's stance. The Credibility aggregation model then merges per-article assessments to compute the overall scoring of the claim being true or false. Finally, the evidence extraction module extracts the supporting evidence from informative snippets from the relevant web articles. The training data for the task is curated from Snopes.com. The system utilizes 5,000 claims from Snopes, each labelled true or false and retrieved 30 relevant Web articles for each of them.

3. Real-Time Certification System on Sina Weibo: Zhou et al. [214] presents a real-time news certification system that can detect an event's credibility by providing keywords about it. The authors built a distributed data acquisition system to enable real-time data flow to gather event-related information through Sina Weibo. The average response time for each query is 35 seconds—this paper model the rumour detection problem from three aspects: content, propagation and information source. The content-based model leverages the event, sub-events and message information. In contrast, the propagation-based model captures propagation network influence. Finally, an ensemble method is utilized to capture the three aspects. In addition, to determine the credibility of an event, the system also provides information such as key users, key microblogs and the timeline of an event.

4. ClaimBuster [67]: It is a fact-checking platform that uses natural language processing and supervised learning techniques to determine factual claims in political discourses. The model is trained on a human-labelled dataset of check-worthy factual claims from the U.S General Election debate transcripts. The system performs the claim spotting task, giving each sentence a score indicating how likely it is to contain an essential factual claim that should be fact-checked. ClaimBuster helps fact-checkers to focus on the top-ranked sentences without searching through a large number of sentences.

5. XFake [204]: It is a fake news detection system that predicts the veracity of the information and provides relevant explanations as prediction evidence. The system comprises of three components that utilize the speaker and statement attributes. Specifically, MIMIC, ATTN and PERT frameworks are designed, where MIMIC is built for attribute analysis, ATTN is for statement semantic analysis, and PERT is for linguistic statement analysis. Explanations, supporting examples and visualization are provided to facilitate interpretation of the output.

6. Jennifer [105]: It is a chatbot maintained by a global group of volunteers. Building such a system aims to provide public information from trusted sources in an organized and efficient manner. Such information can be utilized during a crisis event or in general by the masses to understand public issues. The group also released a dataset, the COVID-19 question bank, consisting of 3,924 COVID-19-related questions.

7. PRTA [33]: Prta stands for **PR**opaganda persuasion **T**echnique **A**nalyzer. It is a system that detects propaganda in text fragments. It also provides readers with information on what type of propaganda technique is used. The system attempts to promote media literacy among online audiences. With Prta, users can explore the contents of articles about several topics, crawled from various sources and updated regularly, and compare them based on their use of propaganda techniques. Prta is designed as a supervised multi-granularity gated BERT-based model, trained on a corpus of news articles annotated at the fragment level with 18 propaganda techniques, a total of 350K word tokens.

2.4 Fake News Datasets

Researchers and practitioners have proposed numerous resources to facilitate research on fake news. This section reviews the fake news and fact-checking datasets from the two viewpoints.

1. News articles: Fabricated content takes various forms. It can be published as news on online news websites consisting of a headline, content (body of the news) and images/videos associated with it. Table 2.1 review the datasets utilized to detect fake news mainly from the body of the news article. The style of each news article is an essential feature for detection.

2. Social Media Posts: Another pattern of fabrication is in the form of memes or tweets/posts on different social media platforms. Table 2.2 review datasets that are utilized to detect fake news, mainly from social media posts. User and network information in social media and text in social media posts are essential features.

Next, Table 2.3 presents a review of the different fact-checking datasets . Unlike fake news detection, a fact-verification task helps to decide whether a claim is correct, specifically when explicitly given the evidence. A model for a fact-verification task classifies whether each claim is correct from a claim and the evidence given as input data. A tuple of a claim and given evidence is generally given as input data for the classification model of a fact-verification task.

Table 2.1: Summary of datasets of fake news detection on news articles.

Dataset	**Instances**	**Labels**	**Topic Domain**	**Raters**	**Language**	**Year**
Politifact14 [185]	221 headlines	5	Politics, Society	Fact-checking sites	English	2014
Buzzfeed_political [70]	71 articles	2	2016 US election	Buzzfeed page	English	2017
Random_political [70]	225 articles	3	Politics	List of Zimdars	English	2017
Ahmed2017 [2]	25,200 articles	2	News in 2016	Fact-checking site (Politifact)	English	2017
LIAR [191, 6]	12,836 claims	6	-	Fact-checking site (Politifact)	English	2017
TSHP-17_politifact [143]	10,483 statements	6	-	Fact-checking site (Politifact)	English	2017
FakeNewsAMT [133]	480 articles	2	Sports, Business, Entertainment, Politics, Technology, Education	Generated fake news by Crowdsourcing	English	2018
Celebrity [133]	500 articles	2	Celebrity	Fact-checking site (GossipCop)	English	2018
Kaggle_UTK	25,104 articles	2	-	-	EEnglish	2018
MisinfoText_Buzzfeed [177]	1413 articles	4	-	Fact-checking site (Buzzfeed)	English	2019
MisinfoText_Snopes [177]	312 articles	5	-	Fact-checking site (Snopes)	English	2019
FA-KES [149]	804 articles	2	Syrian War	Expert annotators	English	2019
Spanish-v1 [138]	971 articles	2	Science, Sport, Politics, Society, Environment, International	Fact-checking sites (VerificadoMX, Maldito Bulo, Caza Hoax)	Spanish	2019
Fauxtography [219]	1,233 articles	2	-	fact-checking site (Snopes)	English	2019
Breaking! [129]	679 articles	3	2016 US election	BS Detector	English	2019
TDS2020	46,700 articles	2	-	News sites (BreiBart, The Onion, InfoWars)	English	2020
FakeCovid [154]	12,805 articles	2-18	-	-	English	2020
TrueFact_FND	6,236 articles	2	-	-	English	2020
Spanish-v2 [138]	572 articles	2	Science, Sport, Politics, Society, Environment, International	Fact-checking sites (VerificadoMX, Maldito Bulo, Caza Hoax)	English	2021

Table 2.2: Summary of datasets of fake news detection on social media.

Dataset	Instances	Labels	Topic Domain	Raters	Platform	Language	Year
MediaEval_Dataset [14]	15,629 posts	2	-	-	Twitter, Facebook, Blog Post	English	2015
PHEME [221]	330 threads	3	Society, Politics	Crowdsourcing	Twitter	English	2016
Twitter-ma [112]	992 threads	2	-	Fact-checking site (Snopes)	Twitter	English	2016
RUMDECT [112]	4,664 threads	2	-	Sina community manaement	Weibo	Chinese	2016
RumorEval2017 [38]	297 threads	3	-	PHEME [221] [220]	Twitter	English	2016
Twitter15 [114]	1,478 threads	4	-	Fact-checking sites (Snopes, Emergent)	Twitter	English	2017
Twitter16 [114]	818 threads	4	-	Fact-checking sites (Snopes, Emergent)	Twitter	English	2017
BuzzFace [150]	2,263 threads	4	Politics	Buzzfeed	Facebook	English	2017
Some-like-it-hoax [170]	15,500 posts	2	Science	[38]	Facebook	English	2017
Media_Weibo [79]	9,528 posts	2	-	Sina community management	Weibo	Chinese	2017
PHEME_update [93]	6,425 threads	3	Society, Politics	PHEME [220]	Twitter	English	2018
FakeNewsNet [157]	23,921 posts	2	Politics, Celebrity	Fact-checking sites (Politifact, GossipCop)	Twitter	English	2018
Jiang2018 [77]	5,303 posts	5	-	Fact-checking sites (Politifact, Snopes)	Twitter, Youtube, Facebook	English	2018
RumorEval2019 [55]	446 threads	3	Natural disasters	Fact-checking sites (Politifact, Snopes)	Twitter, reddit	English	2018
Rumor-anomaly [171]	1,022 threads	6	Politics, Fraud and Scam, Crime, Science, etc.	Fact-checking site (Snopes)	Twitter	English	2019
WeChat_Dataset [193]	4,180 news	2	-	WeChat	WeChat	English	2020
Fang [124]	1,054 threads	2	-	PHEME [93], Twitter-ma [14], FakeNewsNet [157]	Twitter	English	2020
WhatsApp [146]	3,083 images	2	Brazilian Elections, Indian Elections	Fact-checking sites (aosfatos.org, boomlive.in, e-farsas, etc.	WhatsApp	-	2020
Fakeddit [121]	1,063,106 posts	2,3,6	-	Expert annotators	Reddit	English	2020
Reddit_comments	12,597 claims	2	-	Fact-checking sites (Politifact, FactCheck.org, etc)	Twitter	English	2020
HealthStory [34]	1,690 threads	2	Health	HealthNewsreview	Twitter	English	2020
HealthRelease [34]	606 threads	2	Health	HealthNewsreview	Twitter	English	2020
CoAID [30]	4,251 threads	2	COVID-19	Fact-checking sites (Politifact, FactCheck.org, etc)	Twitter	English	2020
COVID-HeRA [40]	61,286 posts	5	COVID-19	CoAID, Expert Annotators	Twitter	English	2020
ArCOVID-19-Rumors [66]	162 threads	2	COVID-19	Fact-checking sites (Fatabyyano, Misbar)	Twitter	Arabic	2020
MM-COVID [102]	11,173 threads	2	COVID-19	Fact-checking sites (Snopes, Poynter)	Twitter	English, Spanish, Portuguese, Hindi, French, Italian	2020
Constraint [130]	10,700 posts	2	COVID-19	Fact-checking sites (Politifact, Snopes)	Twitter	English	2020
Indic-covid [85]	1,438 posts	2	COVID-19	Expert annotators	Twitter	Bengali, Hindi	2020
COVID-19-FAKES [44]	3,047,255	2	COVID-19	WHO, UN, UNICEF	Twitter	Arabic, English	2020
CHECKED [202]	2,104 threads	2	COVID-19	Sina community management	Weibo	Chinese	2021
COVID-Alam [5]	722 tweets	5	COVID-19	Expert annotators	Twitter	English, Arabic	2021
COVID-RUMOR [23]	2,705 posts	2	COVID-19	fact-checking sites (Snopes, Politifact, Boomlive)	Twitter, Websites	English	2021

Table 2.3: Summary of fact-checking datasets.

Dataset	Instances	Lables	Topic domain	Raters	Original data	Main components	Language	Year
Rumor-has-it [141]	10,417 posts	3	-	Expert Annotators	Twitter	Tweet	English	2011
Snopes_credibility [134, 135]	4,856 claims	2	-	Snopes	Web data	Claim, Search pge	English	2016
Wikipedia_credibility [134]	157 claims			Wikipedia	Wikipedia	Claim, search page	English	2016
Emergent [48]	300 claims	3	-	Snopes, Hoaxalizer	Web data	Claim, News articles	English	2016
FNC_dataset [64]	2,587 claims	4	-	Emergent	Politifact	Claim, Body text	English	2017
DeClarE_politifact [137]	3,569 claims	2	-	politifact	Politifact	claim, meta data	English	2018
FEVER [175]	185,445 claims	3	Wikipedia	Expert annotators	Wikipedia	Claim, Wikipedia data	English	2018
FEVER 2.0 [176]	1,174 claims	3	Wikipedia	Expert annotators	Wikipedia	Claim	English	2018
CT-FCC-18 [122]	150 claims	3	Politics, Arab-related news	Snopes, FactCheck.org	Web data	Claim	English, Arabic	2018
Arabic_corpus	429 claims	2	Arab-related news	VERIFY	Web data	Claim, Web page	Arabic	2018
UKPSnopes [65]	6,422 claims	5	-	Snopes, Expert annotators	Snopes	Claim, Web page, Evidence text	English	2019
MultiFC [11]	36,534 claims	2-40	-	-	Fact-checking sites	Claim, Search page	English	2019
DAST [106]	3,007 posts	4	-	Expert annotators	Reddit	Source comments, Submission posts	Danish	2019
Croatian [16]	904 comments	4	-	Expert annotators	24 sata	News articles, News comments	Croatian	2019
CT19-T2 [45]	69 claims	2	Arab-related news	Expert annotators	Web data	Claim, Web page	Arabic	2019
CT20-Arabic [12]	165 claims	2	Arab-related news	Expert annotators	Twitter	Claim, Web page	Arabic	2020
Arabic_Stance [90]	4,547 claims	2	Arab-related news	Expert Annotators	ANT corpus	Claim, News articles	Arabic	2020
PUBHEALTH [94]	11,832 claims	4	Health	Fact-checking sites (Snopes, Politifact, etc)	Web data	Claim, News articles, Explanation texts	English	2020
COVID-19-Scientific [100]	142 claims	3	Science in COVID-19	MedicalNewsToday, CDC, WHO	CORD-19	Claim	English	2020
COVID-19-Politifact [100]	340 claims	2	Politics in COVID-19	Politifact	Politifact	Claim	English	2020
COVIDLies [71]	6,761 posts	3	COVID-19	Expert annotators	Twitter	Tweets, Misconception	English	2020
SCIFACT [190]	1,490 claims	3	Scientific papers	Expert annotators	S2ORC	Claim, Research abstracts	English	2020
HoVer [78]	26,171 claims	3	Wikipedia	Expert annotators	HotpotQA dataset	Claim, Wikipedia data	English	2020
FEVEROUS [8]	87,062 claims	3	Wikipedia	Expert annotators	Wikipedia	Claim, Wikipedia data	English	2021
DANFEVER [125]	6,407 claims	3	Wikipedia	Expert annotators	Wikipedia, Den Store Danske	Claim, Wikipedia data	Danish	2021

Chapter 3

EXPOSING Fake News

In this chapter, we aim to discuss the gaps identified from the literature and their corresponding proposed solutions.

3.1 Research Gap 1: Design Multimodal Detection Baselines

In 2017, Jin et al. [79] made the first attempt toward multimodal fake news detection. The paper proposed a content-based multimodal fake news detection method that uses a recurrent neural network with an attention mechanism to combine text and social context features. It uses VGG19 pre-trained on the Imagenet database to generate representations for images present in tweets. Several other works performed a similar task, discussed in the Section 2.1. Inconsistencies with the existing literature:

1. None of the approaches extracts contextual information from the text. Each method captures the syntactic and semantic features of the text.

2. There are multimodal fake news detection systems in the literature, but they solve the fake news problem by considering an additional sub-task like an event discriminator [192] and finding correlations across the modalities [88]. The results of fake news detection are heavily dependent on the subtask, and in the absence of subtask training, the performance of fake news detection degrades by 10% on an average.

Proposed Solution:

1. We introduce SpotFake- a multimodal framework for fake news detection [162]. Our proposed solution detects fake news without taking into account any other subtasks. It exploits both the textual and visual features of an article. Specifically, we used language models (like BERT) to learn contextual representations for the text, and image features are learned from VGG- 19 pre-trained on the ImageNet dataset.

2. We introduce SpotFake+, a multimodal approach that leverages transfer learning to capture semantic and contextual information from the news articles and its associated images and achieves the better accuracy for fake news detection.

3.1.1 SpotFake: A Multi-modal Framework for Fake News Detection

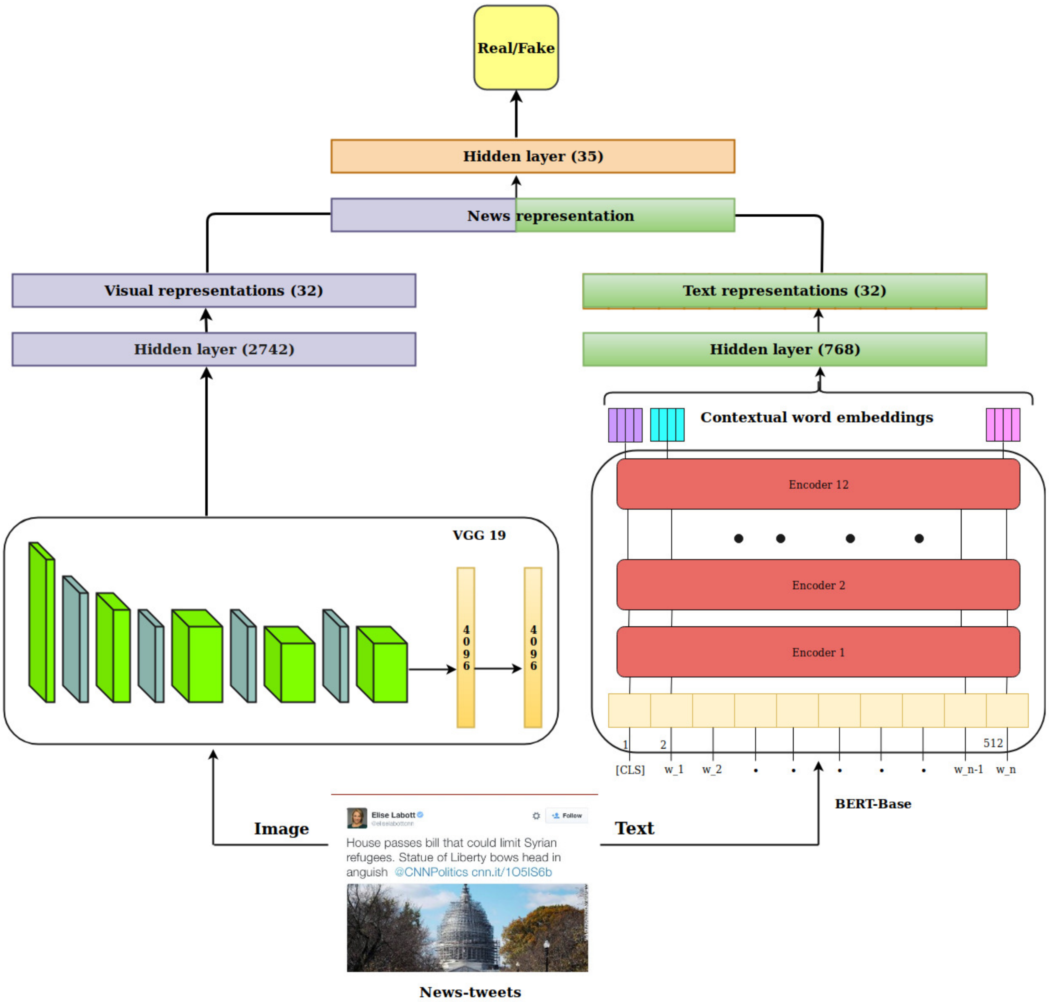


Figure 3.1: A schematic diagram of the SpotFake model. Value in () indicates number of neurons in a layer.

First, we conducted a survey to understand the people's perception of fake news, sources that are more susceptible to the spread of fake news and the effect of multiple modalities on the human ability to detect fake news. We surveyed a sample population that consisted of people in the age group of 15-50 years. The observations from the survey confirm that multimodal features are more beneficial in detecting fake news than unimodal features. Second, fake news classifier of the current state-of-the-art system [192, 88] does not perform well by itself. However, performance significantly improves in the presence of a secondary task like event discriminator or sample reconstruction. Our intuition is that the difference in performance is present due to the singular model's lack of contextual information.

This motivated us to propose SpotFake- a multimodal framework for fake news detection. SpotFake considers the two modalities present in an article- text and image. As shown in Fig.

3.1, the proposed model can be divided into three parts: a text component where a state-of-the-art language model Bidirectional Encoder Representations from Transformers (BERT) [39] is used to extract text representation. For the image representation, we extract the output of the second last layer of VGG-19 convolutional network [161] pre-trained on ImageNet [95] dataset. In the final part, the news representation is constructed by concatenating text and image representation. This news representation is then passed through a fully connected neural network for fake news classification.

3.1.2 SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning

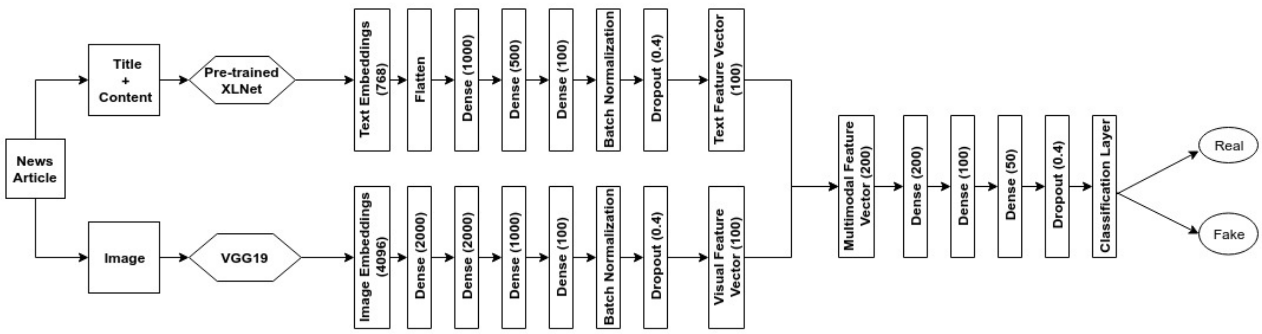


Figure 3.2: A schematic diagram of SpotFake+ model. The model has two sub-modules- a textual feature extractor and a visual feature extractor.

In recent years, there has been a substantial rise in the consumption of news via online platforms. The ease of publication and lack of editorial rigour in some of these platforms have further led to the proliferation of fake news.

In this research work, we studied the problem of detecting fake news on the FakeNewsNet [157] repository, a collection of full length articles along with associated images. Previous studies on FakeNewsNet have used various machine learning techniques (SVM, Naive Bayes, Logistic Regression) and deep learning models (CNN, LSTM, Attention) to perform the fake news detection but they fail to perform well due to following reasons: (1) they lacked the contextual information present in the text and (2) they do not capture the features from the image modality that may seek to emphasize certain facts.
To overcome the above mentioned challenges, we designed SpotFake+, an advanced version of existing multimodal fake news detection system, named SpotFake [162]. The schematic diagram of the model is shown in Figure 3.2. It is a multimodal approach that leverages pretrained language transformers to capture semantic and contextual information from the news articles and pre-trained ImageNet models for visual feature extraction. These feature vectors are fed into fully connected layers for classification. To the best of our knowledge,

this was the first work that performs a multimodal approach for fake news detection on a dataset that consists of full length articles.

3.2 Research Gap 2: Identifying the role of Multiple Images

Inconsistencies with the current literature:
Numerous studies have been performed on multimodal fake news detection, but limited attention is drawn to addressing the role of multiple images. Moreover, none of the works establishes the relationship between multiple components of a news article. Upon examining the related literature, we find the strongest baselines for single-image and multi-image content-based multimodal fake news detection to be SAFE [217] and Giachanou et al. [52] respectively.

1. In the research presented by Zhou et al. [217], (i) the textual features are extracted via a Text-CNN [92] ignoring the contextual information, (ii) the image is converted into text via image2sent [184] model. Next, cosine similarity is calculated to explore the relationship between the two modalities. We believe converting an image into text might result into loss of semantic information within an image and, (iii) no comparison is shown with the existing state-of-the-art methods to demonstrate the effectiveness of the proposed model.

2. On the other side, work performed by Giachanou et al. [52], lack the reasoning for utilizing multiple images for multimodal fake news classification. Second, taking cues only from the headlines, ignoring the content might lead to information loss. Third, while capturing the similarity, top ten image tags are preferred over the image features. This might lead to inconsistent results as, (i) extracted tags might fail to capture the semantic relationship across the images, (ii) incorporating only top ten tags might not capture the information present in the image effectively and, (iii) extracted tags might be limited by the vocabulary of the pre-trained model used for extraction and can introduce external bias in the final representations.

Proposed Solution:

- We present an inter-modality discordance based multimodal fake news detection method. It captures intra-modality relationship by extracting the sequential information from both text and multiple images. In addition, it also forms a multimodal representation of the news article to explore the hidden latent patterns. Our work also introduces a novel application of contrastive loss, employed for measuring the discordance between the components. Enforcing all such losses in conjunction enables for a better feature extraction and robust learning to achieve state-of-the-art performance on the multi-image multimodal fake news detection. Next, we discuss the proposed method in detail.

3.2.1 Inter-modality Discordance for Multimodal Fake News Detection

Existing methods opted for diverse range of solutions to detect fake news. For example, (i) works [162, 163] extracted discriminative features from each modality and performed multimodal fusion to obtain the resultant news vector, (ii) other works [192, 88] added a complementary task to perform fake news detection, (iii) recent work [217, 52] attempts to exploit the relationship between text and image modalities for fake news detection. All these works show the benefits of leveraging unimodal features, adding complementary tasks and studying relationship for multimodal fake news detection.

Taking cues from all the above mentioned approaches, we formulate the problem as a binary multi-task learning method where our primary objective is to perform multimodal fake news detection. The high-level diagram of the proposed approach is presented in Figure 3.3. The model performs multi-task operations with primary goal being multimodal fake news detection. It comprises of four components, (i) Inter-modality discordance score, (ii) Text feature extractor, (iii) Multiple-visual feature extractor and, (iv) Multimodal fake news detector. Next, we discuss each component in detail.

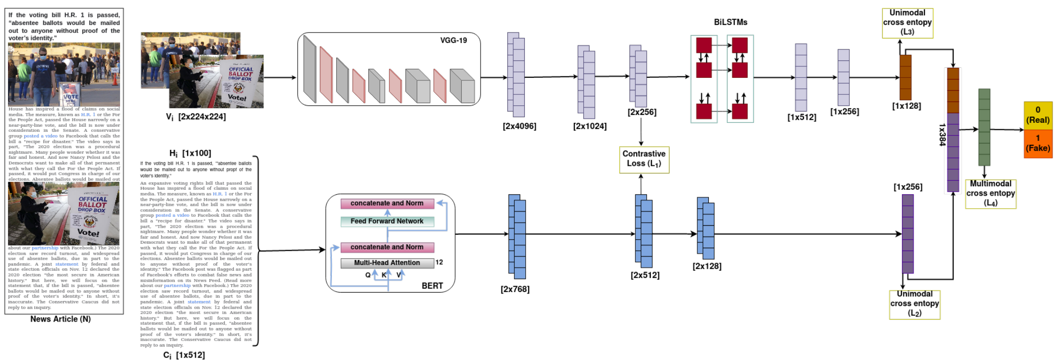


Figure 3.3: Illustration of the Inter-modality discordance based fake news detection model. It comprises of a primary task *i.e.* multimodal fake news detction. There are three auxiliary learning tasks *i.e.* measuring inter-modality discordance score via *contrastive loss*, multiple visual feature extractor and, textual feature extractor.

- Inter-modality Discordance Score: The first auxiliary task is to calculate the discordance score. It captures the relationship (discordance) between various components present in a news article for multimodal fake news detection. More specifically, the idea is that the average distance between the different components of a fake news article is greater than the average distance between the different components of a real news article, in a multimodal space. Taking inspiration from [89], we measure the inter-modality discordance score via a modified version of contrastive loss function. It is a form of metric learning that has shown significant improvement over the conventional cross entropy loss for supervised classification [26]. The objective is to predict relative distance between the inputs.

- Unimodal Visual Feature Extractor: The second auxiliary task considered in the proposed method is the multiple visual feature extractor. Taking inspiration from Giachanou et al. [52], we present a novel system that extracts sequential information from multiple images in a two-fold manner. First, the pre-processed images are passed through a VGG-19 network pre-trained on a ImageNet database. The second to last layer of the VGG-19 network serves as a feature embedding for each image present in the news article. Next, to capture the temporal features from the intermediary sequential visual cues, we employ a Bidirectional Long-Short Term Memory (BiLSTM) cells. The continued representations then obtained are passed through fully connected layers to match the length of vector dimensions with that of the resultant textual feature vector.

- Unimodal Visual Feature Extractor: The third auxiliary task introduced in the proposed method is the textual feature extractor. It extracts contextual representations from the headline and the content of a news sample. Context refers to information that helps the message of a literary text interpret accurately. Unlike Word2Vec [119] and GloVe [132] which are context insensitive, the word embeddings generated by Transformer [39] are context sensitive representations.

3.3 Research Gap 3: Extracting Intra ad Inter Modality Relationship

Inconsistency with the current literature:
Upon examining the literature in Section 2.1, we find the following drawbacks.

1. Each method discussed before extracts visual information via Text-CNN or VGG19. Complete image is passed through the network to generate the representations. Image contains unwanted (redundant) information in the form of background that can be excluded.

2. Existing methods for multimodal fake news detection do not work on the principles of weak and strong modality [192, 88, 162, 163, 32, 217]. Instead, methods capture high-level information from different modalities and jointly model them to determine the authenticity of news. The feature extraction also occurs globally, ignoring the salient pixels containing meaningful information. However, reports[1] show the existence of different versions of fake news due to manipulations performed in the different modalities.

Proposed Solution: We present a novel framework that leverages intra and inter modality relationships for multimodal fake news detection. The method comprises of two modules.

[1]https://www.pagecentertraining.psu.edu/public-relations-ethics/introduction-to-the-ethical-implications-of-fake-news-for-pr-professionals/lesson-2-fake-news-content/types-of-fake-news/

1. Capturing inter-modality relationship: We present a novel architecture that uses a multiplicative multimodal method to capture the inter-modality relationship between modalities. Using the multiplicative multimodal method, we aim to leverage information from a more reliable modality than a less reliable one on a per-sample basis.

2. Capturing intra-modality relationship: Our proposed method captures intra-modality relationship by extracting the fine-grained salient representations for image and text. The resultant feature vectors capture rich contextual dependencies present within its components.

3.3.1 Leveraging Intra and Inter Modality Relationship for Multimodal Fake News Detection

In this research work, we hypothesize that not all modalities play an equal role in the decision-making process on any particular sample. As shown in Figure 3.4, our proposed framework comprises of two components, an intra-modality relationship extractor and an inter-modality relationship extractor.

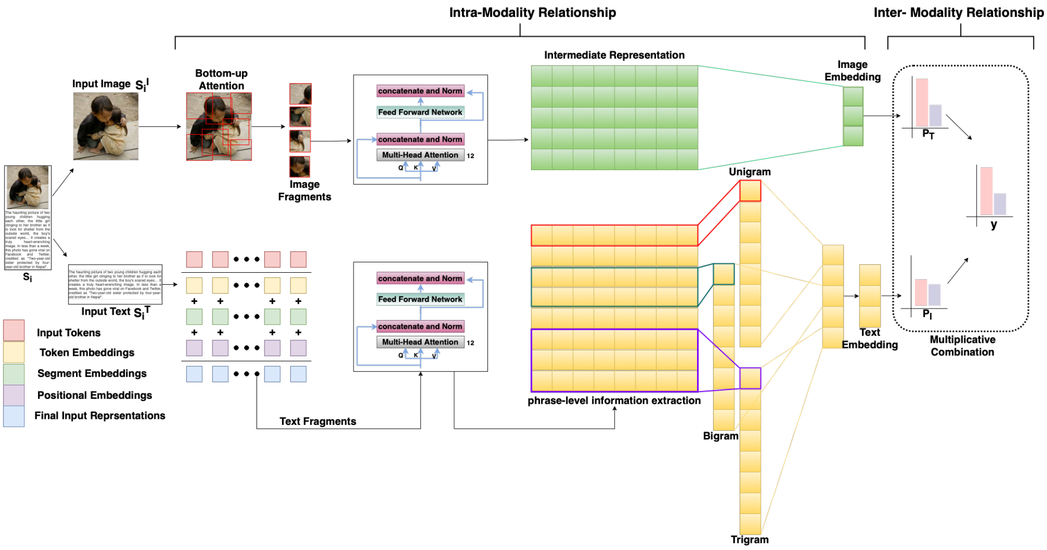


Figure 3.4: The high level diagram of the proposed model that captures intra and inter modality relationship for fake news detection. The method comprises two sub-modules. The intra-modality relationship extractor is responsible for extracting fragments and establishing relationships between them. The inter-modality relationship extractor is responsible for identifying the modality contributing to fakeness.

Former gathers segment information from all the modalities independently; it derives global relationship among each fragment extracted for each modality. Specifically, the idea is to generate fragments of a modality and then learn fine-grained salient representations from the fragments. For image modality, we perform bottom-up attention to extract the image patches [9]. The complex relationship between the patches is then encoded via self-attention mechanism [182]. The final visual representation is obtained by performing an average pooling operation over the fragment representations, resembling bag-of-visual-words model. We use a wordpiece tokenizer to generate text fragments for text modality. Taking inspiration from [39], we use a Transformer module, BERT, to extract contextual representations. The obtained embeddings are further passed through 1d-convolution neural network to extract the phrase-level information. The resultant text representation is obtained by passing intermediate learned representations via a fully connected layer.

At the same time, latter is responsible for identifying strong and weak modalities and utilizes a multiplicative multimodal method [108] to capture inter-modality relationship. The method suppresses the cost of a weaker modality by introducing a down-weight factor in the cross-entropy loss function. The down-weight factor associated with each modality highlights the average prediction power of the remaining modalities. So, if the other modality has higher confidence in predicting the correct class, cost associated with the current modality is suppressed and vice versa.

Chapter 4

EXAMINING Fake News

My research closely examines fake stories to (i) identify types of dissemination over the years, (ii) the modality used to create the fabricated content and (iii) the platform used for dissemination. This chapter discusses the proposed solutions so far. Moreover, we plan to perform characterization of fake news to study fake news incidents in India.

4.1 Research Gap 4: Resource Creation for Indic languages

India witnessed a 214% rise in cases relating to fake news in 2019. There were numerous events across the country that got affected. For instance, during the Pulwama attack, Indian security forces were not only battling terrorists in Kashmir but were also fighting against fake news. Around 5000 social media handles were suspended by Indian security and intelligence agencies during the CAA protests. The dissemination of fake content via WhatsApp was prevalent during the 2019 Indian general election. The proliferation of fake news in India is massive, and there is a dire need to consider solutions explicitly catering to the Indian region. A constant effort is made toward designing automated fact-checking [185, 48, 191, 6, 173, 11] and fake news detection [157, 192, 88, 162, 163, 32, 217] solutions. However, such solutions might have a limited impact on solving the issue in India because the first language (mother tongue) of Indians is diverse and not restricted to English. As a result, we might encounter the production and distribution of fake content in the regional languages—current datasets in English limit the ability to study the menace of fake news in the Indian context. Next, we present an overview of the existing fact-checking and fake news datasets for India with a list of the identified gaps.

4.1.1 Overview of Fact-checking and Fake News Datasets for India

There has been little effort made to study the menace of fake news in India. Recently, Sharma et al. [156] proposed IFND: Indian Fake News Dataset, comprising of the following attributes: (i) title, (ii) date and time, (iii) source of the news, (iv) link to news, and (v) label. The dataset consists of 37,809 and 7,271 real and fake news samples. The real news is collected from

Tribune, Times Now news, The Statesman, NDTV, DNA India, and The Indian express. The fake news samples are curated from Alt news, Boomlive, Digit eye, The logical Indian, News mobile, India Today, News meter, Factcrescendo, TeekhiMirchi, Daapan, and Afp. In another attempt, Dhawan *et al.* [41] proposed FakeNewsIndia to examine the fake news incidents in India. The team curated 4,803 fake news stories from June 2016- December 2019 from 6 fact-checking websites, namely, Times of India, Alt news, Afp, India Today, pIndia, and Factly. The dataset comprises the following attributes, title, author, text, video, date-time, and website. The authors have also curated 5,031 tweets and 866 Youtube videos present in the dataset.

Inconsistencies with the current literature:

Though the datasets have made an effort to create resources that cater Indian region, still it faces a few limitations,

1. The IFND dataset is highly imbalanced. No assurance about the authenticity of sources is provided.
2. In the FakeNewsIndia dataset, the sample count is low. Data curation is also performed for a short period.
3. Both the curated datasets consists of samples in English, missing the data in regional languages.
4. There are numerous attributes present in a website but both the papers limits to some specific features. This might lead to information loss.

Proposed Solution:

- We present FactDrill, a dataset containing 22,435 fact-checked social media content to study fake news incidents in India. The dataset comprises news stories from 2013 to 2020, covering 13 different languages spoken in the country. There are 14 different attributes present in the dataset.

4.1.2 FactDrill: A Data Repository of Fact-Checked Social Media Content to Study Fake News Incidents in India

With its massive population, the rise in production and circulation of fake content online is posing a serious social challenge in India.[1] To limit the escalation of fake news; a constant effort is made towards designing automated fact-checking [186, 48, 191, 6, 173, 11] and fake news detection [157, 192, 88, 217, 162, 163] solutions. However, we believe that such solutions might have a limited impact in solving the issue in India because the first language (mother tongue) of Indians is diverse and not restricted to English. As a result, we might

[1] https://indianexpress.com/article/india/214-rise-in-cases-relating-to-fake-news-rumours-7511534/

encounter the production and distribution of fake content in the regional languages. Current datasets in English limits the ability to study the menace of fake news in the Indian context. In this research work, we present FactDrill: a data repository of fact-checked social media content to understand the dynamics of fake content in a multi-lingual setting in India. The dataset presented in the paper is unique due to the following reasons:

1. **Multilingual Information**: There are 22 official languages in India. The 2011 Census of India[2] shows that the languages by the highest number of speakers (in decreasing order) are as follows: Hindi, Bengali, Marathi, Telugu, Tamil, Gujarati, Urdu, Odia, Malayalam, and Punjabi. On the other hand, only 10.67% of the total population of India converse in English. Though the current datasets are in English, the above statistics indicate a need to shift fake news from English to other languages. Hence, the proposed dataset consists of news samples that span over 13 different languages spoken in India.

2. **Investigation reasoning**: With the FactDrill dataset, we present an attribute that explains how the manual fact-checkers carry out the investigation. We believe providing such information can give insights about the news story like, (i) social media account or website that posted the fake content, (ii) platform that first encountered the fake content, (iii) links to the archive version of the post if the original content is deleted, (iv) tools used by fact-checkers to investigate the claim, and (v) links to the supporting or refuting reports related to the claim. Such insights have the potential to drive the research towards studying the *'Nature of fake news production'* in general. The attribute is exclusive to the FactDrill dataset.

3. **Multi-media and multi-platform information**: Fake news can be published in any form and on any social and mainstream platform. The curated dataset incorporates the information about media (images, text, video, audio, or social media post) used in fake news generation and the medium (Twitter, Facebook, WhatsApp, and Youtube) used for its dissemination.

4. **Multi-domain information**: The previous fact-checking dataset covers information on specific topics only. For example, Emergent [48] only captures the national, technological, and world related happening in the US whereas [191, 6] include health, economic, and election-related issues. In our proposed dataset, we have curated information from the existing fact-checking websites in India, giving us leverage to capture news stories of different topics and cover events that happened during the time frame.

[2] https://en.wikipedia.org/wiki/Multilingualism_in_India

To summarize, we identified four research gaps in the literature (as mentioned in Chapter 3 & 4) and devised solutions for each. Table 4.1 provides a summary of the PhD work proposed to date.

Table 4.1: Overview of the proposed work in the PhD Timeline so far. We identified four gaps and proposed solutions for each of them.

Gaps in the Literature	**Proposed Solution**
Design Multimodal Detection Baselines	1. SpotFake: A Multi-modal Framework for Fake News Detection 2. SpotFake+: A Multimodal Framework for Fake News Detection viaTransfer Learning
Identifying the role of Multiple Images	Inter-modality Discordance for Multimodal Fake News Detection
Extracting Intra ad Inter Modality Relationship	Leveraging Intra and Inter Modality Relationship for Multimodal Fake News Detection
Resource Creation for Indic languages	FactDrill: A Data Repository of Fact-Checked Social Media Content to Study Fake News Incidents in India

Chapter 5

PhD Thesis Timeline

PART I: EXPOSING FAKE NEWS
Status: Completed

1. SpotFake: A Multi-modal Framework for Fake News Detection
2. SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning
3. Inter-modality Discordance for Multimodal Fake News Detection
4. Leveraging Intra and Inter Modality Relationship for Multimodal Fake News Detection

PART II: EXAMINING FAKE NEWS
Status: Ongoing (June 2022-October 2022)

1. FactDrill: A Data Repository of Fact-Checked Social Media Content to Study Fake News Incidents in India (completed)
2. Characterization of Fake News in India: Ongoing (June 2022- September 2022)
3. Validating the performance of Fake News Detection Models (July 2022-October 2022)

PART IV: INTERVENTION
Status: Yet to be started (Start month: October 2022)

Publications

1. Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. "Spotfake: A multi-modal framework for fake news detection." In 2019 IEEE fifth international conference on multimedia big data (BigMM), pp. 39-47. IEEE, 2019.

2. Shivangi Singhal, Anubha Kabra, Mohit Sharma, Rajiv Ratn Shah, Tanmoy Chakraborty, and Ponnurangam Kumaraguru. "Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract)." In Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 10, pp. 13915-13916. 2020.

3. Shivangi Singhal, Mudit Dhawan, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. "Inter-modality Discordance for Multimodal Fake News Detection." In ACM Multimedia Asia, pp. 1-7. 2021.

4. Shivangi Singhal, Tanisha Pandey, Saksham Mrig, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. "Leveraging Intra and Inter Modality Relationship for Multimodal Fake News Detection." (2022).

5. Shivangi Singhal, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. "FactDrill: A Data Repository of Fact-Checked Social Media Content to Study Fake News Incidents in India." In Proceedings of the International AAAI Conference on Web and Social Media, vol. 16, pp. 1322-1331. 2022.

6. Shivangi Singhal, Rishabh Kaushal, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. "Fake News in India: Scale, Diversity, Solution, and Opportunities." In the Proceedings of the Communications of the ACM Regional Section on India Region, 2022.

Acknowledgements

1. I want to thank the Precog Research Lab, IIIT Hyderabad members, for providing insightful comments on the work.

2. I want to thank Anubha Kabra, Mohit Sharma, Mudit Dhawan, Tanisha Pandey, Saksham Mrig and Rishabh Kaushal for being amazing co-authors in the work produced so far.

3. I want to thank TCS Research for proving the PhD Fellowship.

Bibliography

[1] Sadia Afroz, Michael Brennan, and Rachel Greenstadt. Detecting hoaxes, frauds, and deception in writing style online. In *2012 IEEE Symposium on Security and Privacy*, pages 461–475. IEEE, 2012.

[2] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, pages 127–138. Springer, 2017.

[3] Syeda Zainab Akbar, Ankur Sharma, Himani Negi, Anmol Panda, and Joyojeet Pal. Anatomy of a rumour: Social media and the suicide of sushant singh rajput. *arXiv preprint arXiv:2009.11744*, 2020.

[4] Ahmed Al-Rawi, Derrick O[U+02BC]Keefe, Oumar Kane, and Aimé-Jules Bizimana. Twitter's fake news discourses around climate change and global warming. 2021.

[5] Firoj Alam, Fahim Dalvi, Shaden Shaar, Nadir Durrani, Hamdy Mubarak, Alex Nikolov, Giovanni Da San Martino, Ahmed Abdelali, Hassan Sajjad, Kareem Darwish, et al. Fighting the covid-19 infodemic in social media: A holistic perspective and a call to arms. In *ICWSM*, pages 913–922, 2021.

[6] Tariq Alhindi, Savvas Petridis, and Smaranda Muresan. Where is your evidence: Improving fact-checking by justification modeling. In *Proceedings of the first workshop on fact extraction and verification (FEVER)*, pages 85–90, 2018.

[7] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.

[8] Rami Aly, Zhijiang Guo, Michael Sejr Schlichtkrull, James Thorne, Andreas Vlachos, Christos Christodoulopoulos, Oana Cocarascu, and Arpit Mittal. FEVEROUS: Fact extraction and VERification over unstructured and structured information. 2021.

[9] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, and L. Zhang. Bottom-up and top-down attention for image captioning and visual question answering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6077–6086, 2018.

[10] Sanjeev Arora, Yingyu Liang, and Tengyu Ma. A simple but tough-to-beat baseline for sentence embeddings. In *International conference on learning representations*, 2017.

[11] Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4685–4697, Hong Kong, China, November 2019. Association for Computational Linguistics.

[12] Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, et al. Overview of checkthat! 2020: Automatic identification and verification of claims in social media. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 215–236. Springer, 2020.

[13] Christina Boididou, Symeon Papadopoulos, Markos Zampoglou, Lazaros Apostolidis, Olga Papadopoulou, and Yiannis Kompatsiaris. Detection and visualization of misleading content on twitter. *International Journal of Multimedia Information Retrieval*, 7(1):71–86, 2018.

[14] Christina Boididou, Symeon Papadopoulos, Markos Zampoglou, Lazaros Apostolidis, Olga Papadopoulou, and Yiannis Kompatsiaris. Detection and visualization of misleading content on twitter. *International Journal of Multimedia Information Retrieval*, 7(1):71–86, 2018.

[15] Gary D Bond, Rebecka D Holman, Jamie-Ann L Eggert, Lassiter F Speller, Olivia N Garcia, Sasha C Mejia, Kohlby W Mcinnes, Eleny C Ceniceros, and Rebecca Rustige. 'lyin'ted','crooked hillary', and 'deceptive donald': Language of lies in the 2016 us presidential debates. *Applied Cognitive Psychology*, 31(6):668–677, 2017.

[16] Mihaela Bošnjak and Mladen Karan. Data set for stance and sentiment analysis from user comments on Croatian news. In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 50–55, Florence, Italy, August 2019. Association for Computational Linguistics.

[17] Alexandre Bovet and Hernán A Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):1–14, 2019.

[18] Petter Bae Brandtzaeg, Marika Lüders, Jochen Spangenberg, Linda Rath-Wiggins, and Asbjørn Følstad. Emerging journalistic verification practices concerning social media. *Journalism Practice*, 10(3):323–342, 2016.

[19] Adam Breuer, Roee Eilat, and Udi Weinsberg. Friend or faux: graph-based early detection of fake accounts on social networks. In *Proceedings of The Web Conference 2020*, pages 1287–1297, 2020.

[20] David A Broniatowski, Amelia M Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C Quinn, and Mark Dredze. Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American journal of public health*, 108(10):1378–1384, 2018.

[21] Juan Cao, Peng Qi, Qiang Sheng, Tianyun Yang, Junbo Guo, and Jintao Li. Exploring the role of visual content in fake news detection. *Disinformation, Misinformation, and Fake News in Social Media*, pages 141–161, 2020.

[22] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684, 2011.

[23] Mingxi Cheng, Songli Wang, Xiaofeng Yan, Tianqi Yang, Wenshuo Wang, Zehao Huang, Xiongye Xiao, Shahin Nazarian, and Paul Bogdan. A covid-19 rumor dataset. *Frontiers in Psychology*, 12:644801, 2021.

[24] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[25] Rajdipa Chowdhury, Sriram Srinivasan, and Lise Getoor. Joint estimation of user and publisher credibility for fake news detection. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1993–1996, 2020.

[26] Joon Son Chung and Andrew Zisserman. Out of time: Automated lip sync in the wild. pages 251–263, 03 2017.

[27] Nadia K Conroy, Victoria L Rubin, and Yimin Chen. Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, 52(1):1–4, 2015.

[28] John Cook, Peter Ellerton, and David Kinkead. Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, 13(2):024018, 2018.

[29] Jan Christian Blaise Cruz, Julianne Agatha Tan, and Charibeth Cheng. Localization of fake news detection via multitask transfer learning. *arXiv preprint arXiv:1910.09295*, 2019.

[30] Limeng Cui and Dongwon Lee. Coaid: Covid-19 healthcare misinformation dataset. *arXiv preprint arXiv:2006.00885*, 2020.

[31] Limeng Cui, Haeseung Seo, Maryam Tabar, Fenglong Ma, Suhang Wang, and Dongwon Lee. Deterrent: Knowledge guided graph attention network for detecting healthcare misinformation. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 492–502, 2020.

[32] Limeng Cui, Suhang Wang, and Dongwon Lee. Same: Sentiment-aware multi-modal embedding for detecting fake news. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, page 41–48, 2019.

[33] Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barrón-Cedeno, and Preslav Nakov. Prta: A system to support the analysis of propaganda techniques in the news. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 287–293, 2020.

[34] Enyan Dai, Yiwei Sun, and Suhang Wang. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 853–862, 2020.

[35] Anupam Das and Ralph Schroeder. Online disinformation in the run-up to the indian 2019 election. *Information, Communication & Society*, 24(12):1762–1778, 2021.

[36] Marco Del Tredici and Raquel Fernández. Words are the window to the soul: Language-based user representations for fake news detection. 2020.

[37] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016.

[38] Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. SemEval-2017 task 8: RumourEval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 69–76, Vancouver, Canada, August 2017. Association for Computational Linguistics.

[39] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[40] Arkin Dharawat, Ismini Lourentzou, Alex Morales, and ChengXiang Zhai. Drink bleach or do what now? covid-hera: A dataset for risk-informed health decision making in the presence of covid19 misinformation. *arXiv preprint arXiv:2010.08743*, 2020.

[41] Apoorva Dhawan, Malvika Bhalla, Arora Deeksha, Kaushal Rishabh, and Ponnurangam Kumaraguru. Fakenewsindia: A benchmark dataset of fake news incidents in india, collection methodology and impact assessment in social media. *Journal of Computer Communications*, 2022.

[42] Ding Ding, Edward W Maibach, Xiaoquan Zhao, Connie Roser-Renouf, and Anthony Leiserowitz. Support for climate policy and societal action are linked to perceptions about scientific agreement. *Nature Climate Change*, 1(9):462–466, 2011.

[43] Francesco Ducci, Mathias Kraus, and Stefan Feuerriegel. Cascade-lstm: A tree-structured neural classifier for detecting misinformation cascades. In *proceedings of the 26th ACM SIGKDD international conference on Knowledge Discovery & Data Mining*, pages 2666–2676, 2020.

[44] Mohamed K Elhadad, Kin Fun Li, and Fayez Gebali. Covid-19-fakes: A twitter (arabic/english) dataset for detecting misleading information on covid-19. In *International Conference on Intelligent Networking and Collaborative Systems*, pages 256–268. Springer, 2020.

[45] Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeno, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, and Pepa Atanasova. Overview of the clef-2019 checkthat! lab: automatic identification and verification of claims. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 301–321. Springer, 2019.

[46] Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 171–175, 2012.

[47] Pasquale Ferrara, Tiziano Bianchi, Alessia De Rosa, and Alessandro Piva. Image forgery localization via fine-grained analysis of cfa artifacts. *IEEE Transactions on Information Forensics and Security*, 7(5):1566–1577, 2012.

[48] William Ferreira and Andreas Vlachos. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*. ACL, 2016.

[49] Christie M Fuller, David P Biros, and Rick L Wilson. Decision support for determining veracity via linguistic-based cues. *Decision Support Systems*, 46(3):695–703, 2009.

[50] Yi Fung, Christopher Thomas, Revanth Gangi Reddy, Sandeep Polisetty, Heng Ji, Shih-Fu Chang, Kathleen McKeown, Mohit Bansal, and Avirup Sil. Infosurgeon: Cross-media fine-grained information consistency checking for fake news detection.

In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1683–1698, 2021.

[51] Elizabeth A Gage-Bouchard, Susan LaValley, Molli Warunek, Lynda Kwon Beaupin, and Michelle Mollica. Is cancer information exchanged on social media scientifically accurate? *Journal of Cancer Education*, 33(6):1328–1332, 2018.

[52] Anastasia Giachanou, Guobiao Zhang, and Paolo Rosso. Multimodal multi-image fake news detection. In *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 647–654. IEEE, 2020.

[53] Miroslav Goljan, Jessica Fridrich, and Mo Chen. Defending against fingerprint-copy attack in sensor-based camera identification. *IEEE Transactions on Information Forensics and Security*, 6(1):227–236, 2010.

[54] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[55] Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. Semeval-2019 task 7: Rumoureval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, 2019.

[56] Jesse Graham, Jonathan Haidt, and Brian A Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of personality and social psychology*, 96(5):1029, 2009.

[57] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378, 2019.

[58] Gisel Bastidas Guacho, Sara Abdali, Neil Shah, and Evangelos E Papalexakis. Semi-supervised content-based detection of misinformation via tensor embeddings. In *2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM)*, pages 322–325. IEEE, 2018.

[59] Maike Guderlei and Matthias Aßenmacher. Evaluating unsupervised representation learning for detecting stances of fake news. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6339–6349, 2020.

[60] Chuan Guo, Juan Cao, Xueyao Zhang, Kai Shu, and Miao Yu. Exploiting emotions for fake news detection on social media. *ArXiv*, abs/1903.01728, 2019.

[61] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. Rumor detection with hierarchical social attention network. In *Proceedings of the 27th ACM international conference on information and knowledge management*, pages 943–951, 2018.

[62] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. Tweetcred: Real-time credibility assessment of content on twitter. In *International conference on social informatics*, pages 228–243. Springer, 2014.

[63] Jonathan Haidt and Jesse Graham. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1):98–116, 2007.

[64] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. A retrospective analysis of the fake news challenge stance-detection task. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1859–1874, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[65] Andreas Hanselowski, Christian Stab, Claudia Schulz, Zile Li, and Iryna Gurevych. A richly annotated corpus for different tasks in automated fact-checking. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 493–503, Hong Kong, China, November 2019. Association for Computational Linguistics.

[66] Fatima Haouari, Maram Hasanain, Reem Suwaileh, and Tamer Elsayed. Arcov19-rumors: Arabic covid-19 twitter dataset for misinformation detection. *arXiv preprint arXiv:2010.08768*, 2020.

[67] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. Claimbuster: The first-ever end-to-end fact-checking system. *Proc. VLDB Endow.*, 10(12):1945–1948, aug 2017.

[68] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[69] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[70] Benjamin D Horne and Sibel Adali. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *Eleventh international AAAI conference on web and social media*, 2017.

[71] Tamanna Hossain, Robert L. LoganIV, Arjuna Ugarte, Yoshitomo Matsubara, Sean Young, and Sameer Singh. COVIDLies: Detecting COVID-19 misinformation on social media. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online, December 2020. Association for Computational Linguistics.

[72] Seyedmehdi Hosseinimotlagh and Evangelos E Papalexakis. Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In *Proceedings of the Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*, 2018.

[73] Linmei Hu, Tianchi Yang, Luhao Zhang, Wanjun Zhong, Duyu Tang, Chuan Shi, Nan Duan, and Ming Zhou. Compare to the knowledge: Graph neural fake news detection with external knowledge. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 754–763, 2021.

[74] G Huang, Z Liu, L Van Der Maaten, and KQ Weinberger. Densely connected convolutional networks in: Proceedings of the ieee conference on computer vision and pattern recognition, 4700–4708, 2017.

[75] Yu Huang, Chenzhuang Du, Zihui Xue, Xuanyao Chen, Hang Zhao, and Longbo Huang. What makes multi-modal learning better than single (provably). *Advances in Neural Information Processing Systems*, 34, 2021.

[76] Minyoung Huh, Andrew Liu, Andrew Owens, and Alexei A Efros. Fighting fake news: Image splice detection via learned self-consistency. In *Proceedings of the European conference on computer vision (ECCV)*, pages 101–117, 2018.

[77] Shan Jiang and Christo Wilson. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–23, 2018.

[78] Yichen Jiang, Shikha Bordia, Zheng Zhong, Charles Dognin, Maneesh Singh, and Mohit Bansal. HoVer: A dataset for many-hop fact extraction and claim verification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3441–3460, Online, November 2020. Association for Computational Linguistics.

[79] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 795–816, 2017.

[80] Zhiwei Jin, Juan Cao, Jiebo Luo, and Yongdong Zhang. Image credibility analysis with effective domain transferred deep networks. *arXiv preprint arXiv:1611.05328*, 2016.

[81] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.

[82] Zhiwei Jin, Juan Cao, Yongdong Zhang, Jianshe Zhou, and Qi Tian. Novel visual and statistical image features for microblogs news verification. *IEEE transactions on multimedia*, 19(3):598–608, 2016.

[83] Marcia K Johnson and Carol L Raye. Reality monitoring. *Psychological review*, 88(1), 1981.

[84] Zhezhou Kang, Yanan Cao, Yanmin Shang, Tao Liang, Hengzhu Tang, and Lingling Tong. Fake news detection with heterogenous deep graph convolutional network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 408–420, 2021.

[85] Debanjana Kar, Mohit Bhardwaj, Suranjana Samanta, and Amar Prakash Azad. No rumours please! a multi-indic-lingual approach for covid fake-tweet detection. In *2021 Grace Hopper Celebration India (GHCI)*, pages 1–5. IEEE, 2020.

[86] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. Multi-source multi-class fake news detection. In *Proceedings of the 27th international conference on computational linguistics*, pages 1546–1557, 2018.

[87] Hamid Karimi and Jiliang Tang. Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3432–3442, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[88] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. Mvae: Multimodal variational autoencoder for fake news detection. In *The world wide web conference*, pages 2915–2921, 2019.

[89] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 18661–18673. Curran Associates, Inc., 2020.

[90] Jude Khouja. Stance prediction and claim verification: An Arabic perspective. In *Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER)*, pages 8–17, Online, July 2020. Association for Computational Linguistics.

[91] Urja Khurana and Bachelor Opleiding Kunstmatige Intelligentie. The linguistic features of fake news headlines and statements. *Diss. Master's thesis, University of Amsterdam*, 2017.

[92] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October 2014. Association for Computational Linguistics.

[93] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3402–3413, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[94] Neema Kotonya and Francesca Toni. Explainable automated fact-checking for public health claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7740–7754, Online, November 2020. Association for Computational Linguistics.

[95] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[96] Srijan Kumar, Robert West, and Jure Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *Proceedings of the 25th international conference on World Wide Web*, pages 591–602, 2016.

[97] Peter J Lang. A bio-informational theory of emotional imagery. *Psychophysiology*, 16(6):495–512, 1979.

[98] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *International conference on machine learning*, pages 1188–1196. PMLR, 2014.

[99] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[100] Nayeon Lee, Yejin Bang, Andrea Madotto, and Pascale Fung. Misinformation has high perplexity. *arXiv preprint arXiv:2006.04666*, 2020.

[101] W Howard Levie and Richard Lentz. Effects of text illustrations: A review of research. *Ectj*, 30(4):195–232, 1982.

[102] Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. *arXiv preprint arXiv:2011.04088*, 2020.

[103] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In *2018 IEEE International workshop on information forensics and security (WIFS)*, pages 1–7. IEEE, 2018.

[104] Yuezun Li and Siwei Lyu. Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*, 2018.

[105] Yunyao Li, Tyrone Grandison, Patricia Silveyra, Ali Douraghy, Xinyu Guan, Thomas Kieselbach, Chengkai Li, and Haiqi Zhang. Jennifer for COVID-19: An NLP-powered chatbot built for the people and by the people to combat misinformation. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, Online, July 2020. Association for Computational Linguistics.

[106] Anders Edelbo Lillie, Emil Refsgaard Middelboe, and Leon Derczynski. Joint rumour stance and veracity prediction. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 208–221, Turku, Finland, September–October 2019. Linköping University Electronic Press.

[107] Nankai Lina, Sihui Fua, and Shengyi Jianga. Fake news detection in the urdu language using charcnn-roberta. *Health*, 100:100, 2020.

[108] Kuan Liu, Yanen Li, Ning Xu, and Prem Natarajan. Learn to combine modalities in multimodal deep learning. *arXiv preprint*, 2018.

[109] Yang Liu and Yi-Fang Wu. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[110] Antonio Lopez and Jeff Share. Fake climate news: How denying climate change is the ultimate in fake news. *Education*, 2010, 2010.

[111] Yi-Ju Lu and Cheng-Te Li. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. *arXiv preprint arXiv:2004.11648*, 2020.

[112] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. 2016.

[113] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 1751–1754, 2015.

[114] Jing Ma, Wei Gao, and Kam-Fai Wong. Detect rumors in microblog posts using propagation structure via kernel learning. Association for Computational Linguistics, 2017.

[115] Jing Ma, Wei Gao, and Kam-Fai Wong. Detect rumors on twitter by promoting information campaigns with generative adversarial learning. In *The World Wide Web Conference*, WWW '19, page 3049–3055, New York, NY, USA, 2019. Association for Computing Machinery.

[116] Babak Mahdian and Stanislav Saic. Using noise inconsistencies for blind image forensics. *Image and Vision Computing*, 27(10):1497–1503, 2009.

[117] Steven A McCornack, Kelly Morrison, Jihyun Esther Paik, Amy M Wisner, and Xun Zhu. Information manipulation theory 2: A propositional theory of deceptive discourse production. *Journal of Language and Social Psychology*, 33(4):348–377, 2014.

[118] Michail Mersinias, Stergos Afantenos, and Georgios Chalkiadakis. CLFD: A novel vectorization technique and its application in fake news detection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3475–3483, Marseille, France, May 2020. European Language Resources Association.

[119] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[120] Ahmadreza Mosallanezhad, Mansooreh Karami, Kai Shu, Michelle V Mancenido, and Huan Liu. Domain adaptive fake news detection via reinforcement learning. In *Proceedings of the ACM Web Conference 2022*, pages 3632–3640, 2022.

[121] Kai Nakamura, Sharon Levy, and William Yang Wang. Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6149–6157, Marseille, France, May 2020. European Language Resources Association.

[122] Preslav Nakov, Alberto Barrón-Cedeno, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Wajdi Zaghouani, Pepa Atanasova, Spas Kyuchukov, and Giovanni Da San Martino. Overview of the clef-2018 checkthat! lab on automatic identification and verification of political claims. In *International conference of the cross-language evaluation forum for european languages*, pages 372–387. Springer, 2018.

[123] Lakshmanan Nataraj, Tajuddin Manhar Mohammed, BS Manjunath, Shivkumar Chandrasekaran, Arjuna Flenner, Jawadul H Bappy, and Amit K Roy-Chowdhury. Detecting gan generated fake images using co-occurrence matrices. *Electronic Imaging*, 2019(5):532–1, 2019.

[124] Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. Fang: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 1165–1174, 2020.

[125] Jeppe Nørregaard and Leon Derczynski. DanFEVER: claim verification dataset for Danish. In *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 422–428, Reykjavik, Iceland (Online), May 31–2 June 2021. Linköping University Electronic Press, Sweden.

[126] Nelleke Oostdijk, Hans van Halteren, Erkan Başar, and Martha Larson. The connection between the text and images of news articles: New insights for multimedia analysis. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4343–4351, Marseille, France, May 2020. European Language Resources Association.

[127] Naomi Oreskes and Erik M Conway. *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Publishing USA, 2011.

[128] Sunday Oluwafemi Oyeyemi, Elia Gabarron, and Rolf Wynn. Ebola, twitter, and misinformation: a dangerous combination? *Bmj*, 349, 2014.

[129] Archita Pathak and Rohini K Srihari. Breaking! presenting fake news corpus for automated fact checking. In *Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop*, pages 357–362, 2019.

[130] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. Fighting an infodemic: Covid-19 fake news dataset. In *International Workshop on Combating On line Ho st ile Posts in Regional Languages dur ing Emerge ncy Si tuation*, pages 21–29. Springer, 2021.

[131] Kellin Pelrine, Jacob Danovitch, and Reihaneh Rabbany. *The Surprising Performance of Simple Baselines for Misinformation Detection*, page 3432–3441. Association for Computing Machinery, New York, NY, USA, 2021.

[132] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[133] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3391–3401, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[134] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 2173–2178, 2016.

[135] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1003–1012, 2017.

[136] Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. Credeye: A credibility lens for analyzing and explaining misinformation. In *Companion Proceedings of the The Web Conference 2018*, WWW '18, page 155–158, Republic and Canton of Geneva, CHE, 2018. International World Wide Web Conferences Steering Committee.

[137] Kashyap Popat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. Declare: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint arXiv:1809.06416*, 2018.

[138] Juan-Pablo Posadas-Durán, Helena Gómez-Adorno, Grigori Sidorov, and Jesús Jaime Moreno Escobar. Detection of fake news in a new corpus for the spanish language. *Journal of Intelligent & Fuzzy Systems*, 36(5):4869–4876, 2019.

[139] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*, 2017.

[140] Piotr Przybyla. Capturing the style of fake news. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 490–497, 2020.

[141] Vahed Qazvinian, Emily Rosengren, Dragomir Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the 2011 conference on empirical methods in natural language processing*, pages 1589–1599, 2011.

[142] Peng Qi, Juan Cao, Tianyun Yang, Junbo Guo, and Jintao Li. Exploiting multi-domain visual information for fake news detection. In *2019 IEEE international conference on data mining (ICDM)*, pages 518–527. IEEE, 2019.

[143] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937, 2017.

[144] Bhavtosh Rath, Xavier Morales, and Jaideep Srivastava. SCARLET: explainable attention based graph neural network for fake news spreader prediction. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2021.

[145] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st*

Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1650–1659, 2013.

[146] Julio CS Reis, Philipe Melo, Kiran Garimella, Jussara M Almeida, Dean Eckles, and Fabrício Benevenuto. A dataset of fact-checked images shared on whatsapp during the brazilian and indian elections. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 903–908, 2020.

[147] Victoria L Rubin and Tatiana Lukoianova. Truth and deception at the rhetorical structure level. *Journal of the Association for Information Science and Technology*, 66(5):905–917, 2015.

[148] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806, 2017.

[149] Fatima K Abu Salem, Roaa Al Feel, Shady Elbassuoni, Mohamad Jaber, and May Farah. Fa-kes: A fake news dataset around the syrian war. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 573–582, 2019.

[150] Giovanni C Santia and Jake Ryland Williams. Buzzface: A news veracity dataset with facebook user commentary and egos. In *Twelfth international AAAI conference on web and social media*, 2018.

[151] Roney Santos, Gabriela Pedro, Sidney Leal, Oto Vale, Thiago Pardo, Kalina Bontcheva, and Carolina Scarton. Measuring the impact of readability features in fake news detection. In *Proceedings of the 12th language resources and evaluation conference*, pages 1404–1413, 2020.

[152] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681, 1997.

[153] Juan Carlos Medina Serrano, Orestis Papakyriakopoulos, and Simon Hegelich. Nlp-based feature extraction for the detection of covid-19 misinformation videos on youtube. In *Proceedings of the 1st Workshop on NLP for COVID-19 at ACL 2020*, 2020.

[154] Gautam Kishore Shahi and Durgesh Nandini. Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. *arXiv preprint arXiv:2006.11343*, 2020.

[155] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, 96:104, 2017.

[156] Dilip Kumar Sharma and Sonal Garg. Ifnd: a benchmark dataset for fake news detection. *Complex & Intelligent Systems*, pages 1–21, 2021.

[157] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big data*, 8(3):171–188, 2020.

[158] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

[159] Kai Shu, Suhang Wang, and Huan Liu. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the twelfth ACM international conference on web search and data mining*, pages 312–320, 2019.

[160] Kai Shu, Guoqing Zheng, Yichuan Li, Subhabrata Mukherjee, Ahmed Hassan Awadallah, Scott Ruston, and Huan Liu. Early detection of fake news with multi-source weak social supervision. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 650–666. Springer, 2020.

[161] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[162] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh. Spotfake: A multi-modal framework for fake news detection. In *IEEE International Conference on Multimedia Big Data (BigMM)*, pages 39–47, 2019.

[163] Shivangi Singhal, Anubha Kabra, Mohit Sharma, Rajiv Ratn Shah, Tanmoy Chakraborty, and Ponnurangam Kumaraguru. Spotfake+: A multimodal framework for fake news detection via transfer learning. *Proceedings of the AAAI Conference*, pages 13915–13916, 2020.

[164] Laura C Smith, Kenya J Lucas, and Carl Latkin. Rumor and gossip: Social discourse on hiv and aids. *Anthropology & Medicine*, 6(1):121–131, 1999.

[165] Victor Suarez-Lledo, Javier Alvarez-Galvez, et al. Prevalence of health misinformation on social media: systematic review. *Journal of medical Internet research*, 23(1):e17187, 2021.

[166] Shengyun Sun, Hongyan Liu, Jun He, and Xiaoyong Du. Detecting event rumors on sina weibo automatically. In *Asia-Pacific web conference*, pages 120–131. Springer, 2013.

[167] Cass R Sunstein. *On rumors: How falsehoods spread, why we believe them, and what can be done*. Princeton University Press, 2014.

[168] Wen-Ying Sylvia Chou, Anna Gaysynsky, and Joseph N Cappella. Where we go from here: health misinformation on social media, 2020.

[169] C Szegedy, W Liu, Y Jia, P Sermanet, S Reed, D Anguelov, D Erhan, V Vanhoucke, and A Rabinovich. Going deeper with convolutions. in proceedings of the ieee computer society conference on computer vision and pattern recognition, 2015.

[170] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca De Alfaro. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*, 2017.

[171] Nguyen Thanh Tam, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Nguyen Quoc Viet Hung, and Bela Stantic. From anomaly detection to rumour detection using data streams of social platforms. *Proceedings of the VLDB Endowment*, 12(9):1016–1029, 2019.

[172] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. Defining "fake news" a typology of scholarly definitions. *Digital journalism*, 6(2):137–153, 2018.

[173] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. Claimskg: A knowledge graph of fact-checked claims. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtěch Svátek, Isabel Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web – ISWC 2019*, pages 309–324, Cham, 2019. Springer International Publishing.

[174] Yori Thijssen. Breaking the news: the effects of fake news on political attitudes. Master's thesis, University of Twente, 2017.

[175] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. FEVER: a large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[176] James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. The FEVER2.0 shared task. In *Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER)*, pages 1–6, Hong Kong, China, November 2019. Association for Computational Linguistics.

[177] Fatemeh Torabi Asr and Maite Taboada. Big data and quality data for fake news and misinformation detection. *Big Data & Society*, 6(1):2053951719843310, 2019.

[178] Kathie M d'I Treen, Hywel TP Williams, and Saffron J O'Neill. Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 11(5):e665, 2020.

[179] Maria Tsirintani. Fake news and disinformation in health care-challenges and technology tools. In *Public Health and Informatics*, pages 318–321. IOS Press, 2021.

[180] Udo Undeutsch. Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der psychologie*, 11:26–181, 1967.

[181] Sander Van der Linden, Anthony Leiserowitz, Seth Rosenthal, and Edward Maibach. Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2):1600008, 2017.

[182] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5998–6008, 2017.

[183] Anand Venkatraman, Dhruvika Mukhija, Nilay Kumar, and SJ Nagpal. Zika virus misinformation on the internet. *Travel medicine and infectious disease*, 14(4):421–422, 2016.

[184] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CoRR*, abs/1411.4555, 2014.

[185] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 workshop on language technologies and computational social science*, pages 18–22, 2014.

[186] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA, June 2014. Association for Computational Linguistics.

[187] Nguyen Vo and Kyumin Lee. Where are the facts? searching for fact-checked information to alleviate the spread of fake news. *arXiv preprint arXiv:2010.03159*, 2020.

[188] Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas. Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 647–653, 2017.

[189] Soroush Vosoughi, Mostafa 'Neo' Mohsenvand, and Deb Roy. Rumor gauge: Predicting the veracity of rumors on twitter. *ACM transactions on knowledge discovery from data (TKDD)*, 11(4):1–36, 2017.

[190] David Wadden, Shanchuan Lin, Kyle Lo, Lucy Lu Wang, Madeleine van Zuylen, Arman Cohan, and Hannaneh Hajishirzi. Fact or fiction: Verifying scientific claims. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7534–7550, Online, November 2020. Association for Computational Linguistics.

[191] William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

[192] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. Eann: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining*, pages 849–857, 2018.

[193] Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. Weak supervision for fake news detection via reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 516–523, 2020.

[194] Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 540–547, 2020.

[195] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. Systematic literature review on the spread of health-related misinformation on social media. *Social science & medicine*, 240:112552, 2019.

[196] Claire Wardle and Hossein Derakhshan. Information disorder: Toward an interdisciplinary framework for research and policymaking, 2017.

[197] Maxwell Weinzierl and Sanda Harabagiu. Identifying the adoption or rejection of misinformation targeting covid-19 vaccines in twitter discourse. In *Proceedings of the ACM Web Conference 2022*, pages 3196–3205, 2022.

[198] Ke Wu, Song Yang, and Kenny Q Zhu. False rumors detection on sina weibo by propagation structures. In *2015 IEEE 31st international conference on data engineering*, pages 651–662. IEEE, 2015.

[199] Kun Wu, Xu Yuan, and Yue Ning. Incorporating relational knowledge in explainable fake news detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2021.

[200] Liang Wu and Huan Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*, pages 637–645, 2018.

[201] Yang Wu, Pengwei Zhan, Yunjian Zhang, Liming Wang, and Zhen Xu. Multimodal fusion with co-attention networks for fake news detection. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2560–2569, 2021.

[202] Chen Yang, Xinyi Zhou, and Reza Zafarani. Checked: Chinese covid-19 fake news dataset. *Social Network Analysis and Mining*, 11(1):1–8, 2021.

[203] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD workshop on mining data semantics*, pages 1–7, 2012.

[204] Fan Yang, Shiva K Pentyala, Sina Mohseni, Mengnan Du, Hao Yuan, Rhema Linder, Eric D Ragan, Shuiwang Ji, and Xia Hu. Xfake: Explainable fake news detector with visualizations. In *The World Wide Web Conference*, pages 3600–3604, 2019.

[205] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. Unsupervised fake news detection on social media: A generative approach. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 5644–5651, 2019.

[206] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes using inconsistent head poses. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8261–8265. IEEE, 2019.

[207] Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Zhoujun Li, and Philip S Yu. Ti-cnn: Convolutional neural networks for fake news detection. *arXiv preprint arXiv:1806.00749*, 2018.

[208] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R Salakhutdinov, and Quoc V Le. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32, 2019.

[209] Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, and Songlin Hu. Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 5444–5454, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.

[210] Markos Zampoglou, Symeon Papadopoulos, Yiannis Kompatsiaris, Ruben Bouwmeester, and Jochen Spangenberg. Web and social media image forensics for news professionals. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10, pages 159–166, 2016.

[211] Daniel Yue Zhang, Lanyu Shang, Biao Geng, Shuyue Lai, Ke Li, Hongmin Zhu, Md Tanvir Amin, and Dong Wang. Fauxbuster: A content-free fauxtography detector using social media comments. In *2018 IEEE international conference on big data (big data)*, pages 891–900. IEEE, 2018.

[212] Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. Mining dual emotion for fake news detection. In *Proceedings of the Web Conference 2021*, pages 3465–3476, 2021.

[213] Lina Zhou, Judee K Burgoon, Jay F Nunamaker, and Doug Twitchell. Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications. *Group decision and negotiation*, 13(1):81–106, 2004.

[214] Xing Zhou, Juan Cao, Zhiwei Jin, Fei Xie, Yu Su, Dafeng Chu, Xuehui Cao, and Junqiang Zhang. Real-time news cer tification system on sina weibo. In *Proceedings of the 24th international conference on world wide web*, pages 983–988, 2015.

[215] Xinyi Zhou, Atishay Jain, Vir V Phoha, and Reza Zafarani. Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice*, 1(2):1–25, 2020.

[216] Xinyi Zhou, Kai Shu, Vir V Phoha, Huan Liu, and Reza Zafarani. "this is fake! shared it by mistake": Assessing the intent of fake news spreaders. In *Proceedings of the ACM Web Conference 2022*, pages 3685–3694, 2022.

[217] Xinyi Zhou, Jindi Wu, and Reza Zafarani. Safe: Similarity-aware multi-modal fake news detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2020.

[218] Xinyi Zhou and Reza Zafarani. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40, 2020.

[219] Dimitrina Zlatkova, Preslav Nakov, and Ivan Koychev. Fact-checking meets fauxtography: Verifying claims about images. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2099–2108, Hong Kong, China, November 2019. Association for Computational Linguistics.

[220] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. Exploiting context for rumour detection in social media. In *International conference on social informatics*, pages 109–123. Springer, 2017.

[221] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one*, 11(3):e0150989, 2016.

[222] Miron Zuckerman, Bella M DePaulo, and Robert Rosenthal. Verbal and nonverbal communication of deception. In *Advances in experimental social psychology*, volume 14, pages 1–59. Elsevier, 1981.