Social media for news consumption is a double-edged sword. On the one hand, its low cost, easy access, and rapid dissemination of information allow users to consume and share the news. On the other hand, it can make viral “fake news”, i.e., low-quality news with intentionally false information. **(https://www.kdnuggets.com/2017/10/guide-fake-news-detection-social-media.html)**

In the recent years, the reliability of information on the Internet has emerged as a crucial issue of modern society. Social network sites (SNSs) have revolutionized the way in which information is spread by allowing users to freely share content. As a consequence, SNSs are also increasingly used as vectors for the diffusion of misinformation and hoaxes. **(**[**https://www.semanticscholar.org/paper/Some-Like-it-Hoax%3A-Automated-Fake-News-Detection-in-Tacchini-Ballarin/35d395a89d5859a23e40666e0e90cdd5e0e48da9**](https://www.semanticscholar.org/paper/Some-Like-it-Hoax%3A-Automated-Fake-News-Detection-in-Tacchini-Ballarin/35d395a89d5859a23e40666e0e90cdd5e0e48da9)**)**

When the Internet was first made accessible for public use in the 1990s, its main purpose was for the seeking and accessing of information.[[113]](https://en.wikipedia.org/wiki/Fake_news#cite_note-113) As fake news was introduced to the Internet, this made it difficult for some people to find truthful information. The impact of fake news has become a worldwide phenomenon.[[114]](https://en.wikipedia.org/wiki/Fake_news#cite_note-114) Fake news is often spread through the use of [fake news websites](https://en.wikipedia.org/wiki/Fake_news_website), which, in order to gain credibility, specialize in creating attention-grabbing news, which often [impersonate](https://en.wikipedia.org/wiki/Website_spoofing) well-known news sources.[[115]](https://en.wikipedia.org/wiki/Fake_news#cite_note-115)[[116]](https://en.wikipedia.org/wiki/Fake_news#cite_note-snopesfieldguide-116)[[117]](https://en.wikipedia.org/wiki/Fake_news#cite_note-bengilbert-117) Jestin Coler, who said he does it for "fun",[[17]](https://en.wikipedia.org/wiki/Fake_news" \l "cite_note-60Overtime-17) also said he earned US$10,000 per month from advertising on his fake news websites.[[95]](https://en.wikipedia.org/wiki/Fake_news#cite_note-60Minutes-95) In 2017, the inventor of the [World Wide Web](https://en.wikipedia.org/wiki/World_Wide_Web), [Tim Berners-Lee](https://en.wikipedia.org/wiki/Tim_Berners-Lee) claimed that fake news was one of the three most significant new disturbing Internet trends that must first be resolved, if the Internet is to be capable of truly "serving humanity." ([**https://en.wikipedia.org/wiki/Fake\_news#On\_the\_Internet**](https://en.wikipedia.org/wiki/Fake_news#On_the_Internet)**)**

**Fake news** in the words of Donald Trump is:

**misinformation**consisting of either**completely fabricated elements**or in a subtler manner,**the association of truth with lies**in an effort of**attracting attention on an issue or a cause**by feeding on**people’s gullibility.**

For [example](https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook?utm&utm_term=.xcGkVBNoxk#.wwwqW6rpmq), the most popular fake news was more widely spread on Facebook than the most popular authentic mainstream news during the U.S. 2016 president election. Therefore, fake news detection on social media has attracted increasing attention from researchers to politicians.

Therefore, fake news detection on social media has recently become an emerging research that is attracting tremendous attention.

# Fake News: Lies spread faster on social media than truth does

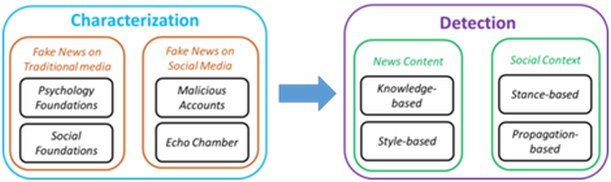
People are quicker to repeat something that's wrong than something that's true

Fake news detection on social media has unique characteristics and presents new challenges. First, fake news is intentionally written to mislead readers to believe false information, which makes it difficult to detect based on news content. Thus, we need to include auxiliary information, such as user social engagements on social media, to help differentiate it from the true news. Second, exploiting this auxiliary information is nontrivial in and of itself as users’ social engagements with fake news produce data that is big, incomplete, unstructured, and noisy. This quick guide is based on a recent [survey](http://dl.acm.org/citation.cfm?id=3137600) [1] that presents issues of fake news detection on social media, state-of-the-art research findings, datasets, and further directions. Next, we will highlight the major perspectives for this survey.

**Characterization and Detection**

Figure 1 is an overview of detecting fake news on social media, including two phases: characterization and detection. Fake news itself is not a new problem, and the media ecology has been changing over time from newsprint to radio/television, and recently online news and social media. The impact of fake news on traditional media can be described from the perspective of psychology and social theories.  For example, two major psychology factors make consumers naturally vulnerable to the fake news: (i) Naïve Realism: consumers tend to believe that their perceptions of reality are the only accurate views. (ii) Confirmation Bias: consumers prefer to receive information that confirms their existing views. As another example, social identity theory and normative influence theory describe that preference for social acceptance is essential to a person’s identity, making people choose “socially safe” option for consuming news, even the news being shared is fake news.

Fake news on social media has its unique characteristics. For example, malicious accounts can be easily and quickly created to boost the spread of fake news, such as social bots, cyborg users, or trolls. In addition, users are selectively exposed to certain types of news because of the way news feed appear on the homepage in social media. Therefore, users on social media tend to form groups containing like-minded people where they are likely to polarize their opinions, resulting in an echo chamber effect.



**Figure 1. Fake news detection on social media: from characterization to detection**

The aforementioned theories are valuable in guiding research of fake news detection. Existing algorithms for fake news detection can be generally categorized as (i) News Content Based and (ii) Social Context Based.

* News content based approaches focus on extracting various features in fake news content, including knowledge-based and style-based. Since fake news attempts to spread false claims, knowledge-based approaches aim to using external sources to fact-check the truthfulness of the claims in news content. In addition, fake news publishers often have malicious intents to spread distorted and misleading, requiring particular writing styles to appeal to and persuade a wide scope of consumers that are not seen in true news articles. Style-based approaches try to detect fake news by capturing the manipulators in the writing style.
* Social context based approaches aim to utilize user social engagements as auxiliary information to help detect fake news. Stance-based approaches utilize users’ viewpoints from relevant post contents to infer the veracity of original news articles. In addition, propagation-based approaches reason about the relations of relevant social media posts to guide the learning of credibility scores by propagating credibility values between users, posts, and news. The veracity of a news piece is aggregated by the credibility values of relevant social media posts.

**Datasets**

Even though online news can be collected from different sources, manually determining the veracity of news is a challenging task, usually requiring annotators with domain expertise who performs a careful analysis of claims and additional evidence, context, and reports from authoritative sources. Existing public datasets of fake news are rather limited due to these challenges. To facilitate the research for fake news detection, this survey [1] provides a usable dataset, named [FakeNewsNet](https://github.com/KaiDMML/FakeNewsNet" \t "_blank), which includes news content and social context features with reliable ground truth fake news labels.

**Promising Future Research**

Fake news detection on social media is a newly emerging research area. The survey [1] discusses related research areas, open problems, and future research directions from a data mining perspective. As shown in Figure 2, research directions are outlined in four perspectives: Data-oriented, Feature-oriented, Model-oriented, and Application-oriented.



**Figure 2. Future directions and open issues for fake news detection on social media**

* Data-oriented: it focuses on different aspects of fake news data, such as benchmark data collection, psychological validation of fake news, and early fake news detection.
* Feature-oriented: it aims to explore effective features for detecting fake news from multiple data sources, such as news content and social context.
* Model-oriented: it opens the door to build more practical and effective models for fake news detection, including supervised, semi-supervised and unsupervised models.
* Application-oriented: it encompasses research that goes beyond fake news detection, such as fake new diffusion and intervention.

[1] Shu, K., Sliva, A., Wang, S., Tang, J. and Liu, H., 2017. [Fake News Detection on Social Media: A Data Mining Perspective. ACM SIGKDD Explorations Newsletter, 19(1), pp.22-36.](http://dl.acm.org/citation.cfm?id=3137600)

**Bio:**[**Kai Shu**](http://www.public.asu.edu/~skai2/)is a Graduate Research Assistant at Arizona State University, His research interests include social media mining, especially information credibility, fake news, and machine learning. [Huan Liu](http://www.public.asu.edu/~huanliu/" \t "_blank) is a professor in School of Computing, Informatics, and Decision Systems Engineering, Ira A. Fulton Schools of Engineering, Arizona State University.

This Project comes up with the applications of NLP (Natural Language Processing) techniques for detecting the 'fake news', that is, misleading news stories that comes from the non-reputable sources. Only by building a model based on a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) tfidf  matrix, (word tallies relative to how often they’re used in other articles in your dataset) can only get you so far. But these models do not consider the important qualities like word ordering and context. It is very possible that two articles that are similar in their word count will be completely different in their meaning. The data science community has responded by taking actions against the problem. There is a Kaggle competition called as the “Fake News Challenge” and Facebook is employing AI to filter fake news stories out of users’ feeds. Combatting the fake news is a classic text classification project with a straight forward proposition. Is it possible for you to build a model that can differentiate between “Real “news and “Fake” news? So a proposed work on assembling a dataset of both fake and real news and employ a Naive Bayes classifier in order to create a model to classify an article into fake or real based on its words and phrases.

<https://www.pantechsolutions.net/fake-news-detection-using-machine-learning>

### FAKE NEWS CHALLENGE STAGE 1 (FNC-I):  STANCE DETECTION

**Exploring how artificial intelligence technologies could be leveraged to combat fake news.**

The goal of the Fake News Challenge is to explore how artificial intelligence technologies, particularly machine learning and natural language processing, might be leveraged to combat the fake news problem. We believe that these AI technologies hold promise for significantly automating parts of the procedure human fact checkers use today to determine if a story is real or a hoax.  
  
Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts. Fortunately, the process can be broken down into steps or stages. A helpful first step towards identifying fake news is to understand what other news organizations are saying about the topic. We believe automating this process, called Stance Detection, could serve as a useful building block in an AI-assisted fact-checking pipeline. So stage #1 of the Fake News Challenge (FNC-1)focuses on the task of Stance Detection.

Stance Detection involves estimating the relative perspective (or stance) of two pieces of text relative to a topic, claim or issue. The version of Stance Detection we have selected for FNC-1 extends the work of Ferreira & Vlachos. For FNC-1 we have chosen the task of estimating the stance of a body text from a news article relative to a headline. Specifically, the body text may agree, disagree, discuss or be unrelated to the headline.

For additional details, see [http://www.fakenewschallenge.org](http://www.fakenewschallenge.org/)

For competition related announcements, sign up at https://groups.google.com/forum/#!search/fnc-1-compete