

Lies vs Truth and Hate vs Love: **Exploring recent trends in fake news and hate speech research**

Md. Shad Akhtar, Tanmoy Chakraborty, IIIT Delhi, India

The emergence of digital media platforms has made accessibility to information extremely efficient and just a few clicks/touches away. However, every coin has two sides. On one side, we have a huge amount of information without any geographical constraints; On the other hand, not every information on the web is trustworthy, verified, or accurate. Hostile posts such as fake news, hate speeches, offensive posts, etc., are one such example that are unarguably detrimental for societal good.

Fake news are, in general, intentionally created posts to spread malicious and false narratives among the public. The creator leverages the chaos/misinformation to gain political, financial, or regional advantages in a quick time, which otherwise would have been difficult to obtain in a legal and unbiased environment. Fake news detection is a non-trivial problem, and it often requires domain expertise or real world knowledge to detect fake news. Digital media platforms are bombarded with new information continuously. It is a challenging task to verify each and every post on the internet to identify whether or not the information is fake or real. However, we can simplify the process if we can filter out the posts that are worth exploring for the fake news detection. In general, a (real/fake) news projects a claim, and the veracity of the claim discriminates between a fake or real news. A claim is an assertion that may or may not be true, and is a worthy candidate to establish the veracity of the assertion. In this tutorial, we will explore the claim detection and veracity prediction tasks in detail.

Another dimension of hostile posts is hate speech. It is a statement that expresses hatred towards a specific group based on their regional, religious, political, or minority status. Some hate speeches are so intense that they can incite violence or unrest in society and cause losses to lives and properties. Curbing hate speeches is extremely important and many organizations including law and enforcement agencies (LEAs) resort to various measures for tackling it. However, most of these measures are retroactive and often initiated after the damage has been done. Second possibility is the proactive measure -where the users are restricted to write hate speeches-, but a complete restriction could be a disaster for freedom-of-expression. Hate speech normalization is another counter measure that ensures freedom-of-expression in a controlled environment. It targets high intensive hate speeches and converts them into weak hate speeches without restricting a users' opinion or view. In the second half of the tutorial, we will explore hate speech normalization and it's subtasks.

The proposed tutorial will cover recent research and future directions in fake news and hate speeches. In particular, following topics will be covered.

1. Fake news and Claims
 - a. Claim Detection
 - b. Claim Veracity Prediction
 - c. Datasets
2. Hate speech normalization
 - a. Measuring degree of hate
 - b. Identification of hateful phrases
 - c. Paraphrasing to reduce hate

Dr. Tanmoy Chakraborty is an Assistant Professor and a Ramanujan Fellow at the Dept. of Computer Science and Engineering, IIIT Delhi, India. Prior to this, he was a postdoctoral researcher at University of Maryland, College Park, USA. He completed his Ph.D as a Google India PhD scholar from IIT Kharagpur, India in 2015. His primary research interests include Social Computing and Natural Language Processing. He has received several awards including Google India Faculty Award, DST Early Career Research Award, DAAD Faculty award, Best PhD thesis award by Xerox Research, IBM Research and Indian National Academy of Engineering (INAE).

Dr. Md. Shad Akhtar is currently an Assistant Professor in the Department of Computer Science and Engineering, IIIT-Delhi. Prior to joining IIIT-Delhi in January, 2020, he served IIIT Dharward as an Assistant Professor. He completed his PhD from IIT Patna in Nov 2019, and holds M.Tech and BEng degrees in computer science from IIT(ISM)-Dhanbad and JMI, New Delhi, respectively. He also has industry experience of more than two years with HCL Technologies Ltd. His main area of research is Natural Language Processing with a focus on affective analysis, conversation dialog systems, low-resource and code-mixed languages.