

# Fake News in India: Scale, Diversity, Solution, and Opportunities

BY SHIVANGI SINGHAL, RISHABH KAUSHAL,  
RAJIV RATN SHAH, AND PONNURANGAM KUMARAGURU

INDIA WITNESSED A 214% rise in cases relating to fake news in the pandemic year of 2019.<sup>a</sup> There were numerous events across the country that were affected. For instance, during the 2016 Indian banknote demonetization, multiple fake news reports about spying technology added to the banknotes went viral.<sup>b</sup> Around 5,000 social media handles were suspended by Indian security and intelligence agencies during the CAA protests.<sup>c</sup> The dissemination of fake content via WhatsApp was prevalent during India's 2019 general election.<sup>d</sup> The proliferation of fake news in India is massive, and there is a dire need to consider solutions explicitly catering to this region.

- a <https://bit.ly/3u8sYtC>  
b Chilwa, E. Innocent, A., Samoilenko, S. *Handbook of Research on Deception, Fake News, and Misinformation Online*. IGI Global, June 28, 2019.  
c Around 5,000 Pak social media handles spread fake news on CAA. *Outlook India*. (Dec. 16, 2019).  
d Perrigo, B. How Whatsapp Is Fueling Fake News Ahead of India's Elections. *Time* (Jan. 25, 2019).

India is a nation that realizes unity in diversity. Indians follow different religions, practice different customs and traditions, and speak diverse languages.<sup>e</sup> These factors and others make it difficult to detect fake news in India. More specifically, we observe the following:

► **Multilinguality.** Mother tongue of Indians is diverse. There are 22 official languages and only 10.67% of the population converse in English.<sup>f</sup> The current fake news detection solutions are most effective for English and might fail to identify and process information in other languages.

► **Instant messaging platform.** We cannot undermine WhatsApp's role in forming and mobilizing the online public. Since WhatsApp is end-to-end encrypted, identifying and quashing false stories is possible only with support of the users.

► **Digital illiteracy.** In India, the surge in the Internet pen-

- e <https://www.ethnologue.com/guides/countries-most-languages>  
f [https://en.wikipedia.org/wiki/Languages\\_of\\_India](https://en.wikipedia.org/wiki/Languages_of_India)

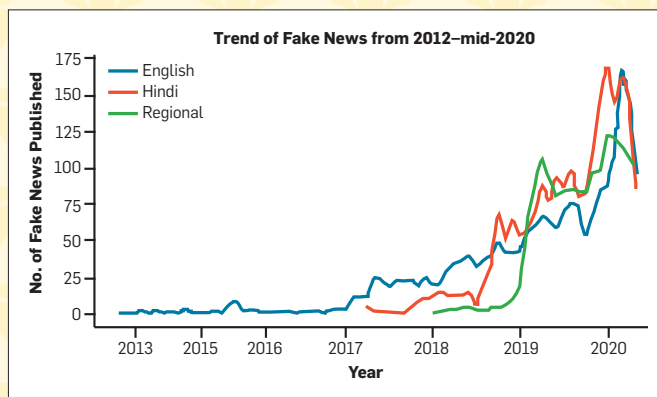


Figure 1. The circulation of fake news over the years in India with an average increase in the production of fake stories from 2018.

etration accompanied by digital illiteracy has resulted in the rise of fake news online.<sup>4</sup> Internet penetration in India has risen from 137 million Internet users in 2012 to over 600 million in 2019.<sup>5</sup> Though audiences have access to information, their inability to comprehend the nuances and implications of fake news has resulted in the growth of fake news dissemination in the online world.

Owing to such diversities, Agarwal et al.<sup>1</sup> recently examined the cross-lingual

dynamics on Sharechat during the 2019 Indian general election across 14 languages. The study highlights the role of image-based content in communicating across language barriers. However, the fake news phenomenon manifests in complex forms and throws unprecedented challenges for computational solutions. One possible pathway would be to design intervention policies to measure the respondents' ability to identify misinformation.<sup>2</sup> The other way to tackle India's situation would be to introduce mediums for the accessibility and utilization of fact-checkers efforts.



Figure 2. Popular fake news events circulated in English, Hindi, and regional languages respectively.

Fact-checking is a method that aims to debunk the viral fake news stories circulated over the Web and provide the supporting claims to prove the veracity. The International Fact-Checking Network<sup>h</sup> (IFCN) is an organization that provides licenses (signatory) to fact-checkers to practice.

## Data Utilization

To initiate the research for the Indian region, we introduced two resources: FakeNewsIndia<sup>3</sup> in 2021 and FactDrill,<sup>5</sup> in 2022. FakeNewsIndia is a subset of FactDrill—a data repository of 22,435 fact-checked social media stories to study fake news incidents in India. Its data spans over seven years, from 2013 to 2020.<sup>i</sup> The samples are from the 11 Indian fact-checking websites certified by IFCN comprising 13 different languages. Fourteen varied attributes associated with each sample are categorized under meta-features, textual features, media features, social features, and event features.

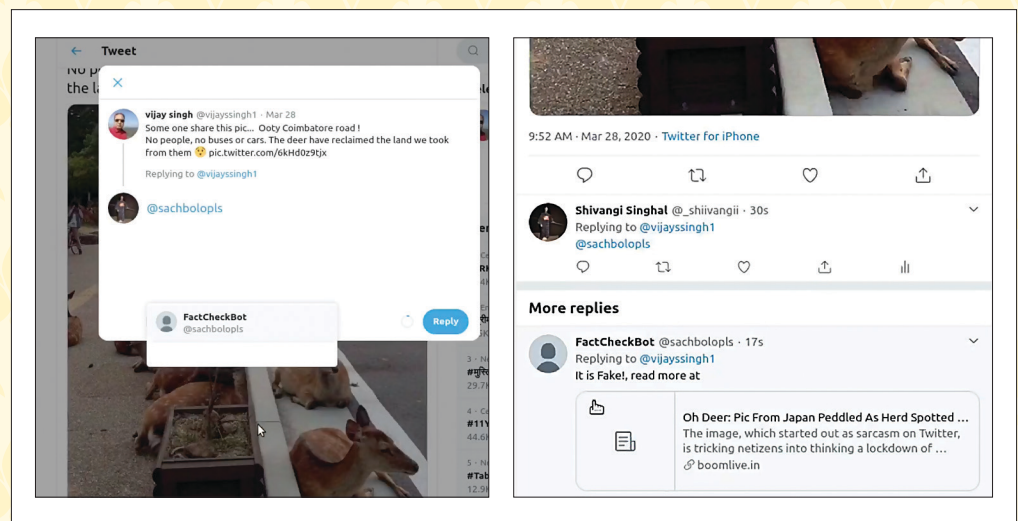
From FactDrill, we noticed an average increase in the production of fake stories with regional languages entering the space in 2018 (see Figure 1). We also concluded from the *domain/tags* attribute present in the FactDrill that political activity is the primary ground for exacerbating the proliferation of fake news in India (see Figure 2).

The FactDrill resource opens the gates for researchers to explore and discover fake news patterns in India. A few possible research directions are:

### 1. Fake news characteriza-

<sup>h</sup> <https://ifencodeofprinciples.poynter.org>

<sup>i</sup> The team is working on the automated resource generation mechanism and intends to release the updated version of the FactDrill soon.



**Figure 3.** A screenshot of the SachBoloPls user interface agent. The system is invoked by a user to check the authenticity of the tweet (left). SachBoloPls response is posted back on the tweet thread (right).

tion. To determine which modality (audio, video, text, image) is most fabricated, what amount of news gets resurfaced on the platform, and how often that happens.

2. *Temporal modeling.* The data spans over seven years enabling the study of the evolution of fake news over the years.

3. *Online-offline interplay.* There might be a direct/indirect effect of an offline event (Pulwama Attack, CAA, NRC, and others) fueling the rage in the online world (propagation of fake news). It might be useful to find such patterns, form solutions, and direct intervention policies.

## Data Accessibility: A System Design

SachBoloPls (means ‘Speak the Truth Please’) is an effort to curb the proliferation of fake news online and make debunk information accessible to the people. Time and again, we have witnessed fake stories resurfacing the online media and the masses falling prey to it, a clear sign of negligence from the Indian audiences toward fact-checking efforts. We believe it might be worthwhile to make audiences aware of such fact-checking organizations and


educate them about the false viral claims.

SachBoloPls is a system that validates news on Twitter in real time. When a user invokes it, the system collects the original news tweet of the thread and matches it with the database. Results are then formed into a tweet and tweeted back to the original line. There are three components of the design that are independent of each other. Thus, extending the working module to other social media platforms like Instagram, WhatsApp, Facebook, and Telegram gives us leverage. SachBoloPls can also incorporate regional languages making it a viable tool to fight against fake news across India.

## Conclusion

The various facets of India make it a challenging ground to study the dissemination of fake content. There is a dire need to devise technological and policy decisions exclusively for India. To initiate the research, we introduced FakeNewsIndia and FactDrill resources and a system designed to impart cognitive abilities to the masses.

**Acknowledgments.** The authors thank Hitkul Jangra,

Ritwik Mishra, Prashant Kodali, and Mudit Dhawan for their insightful comments on an earlier draft. 

## References

- Agarwal, P., Garimella, K., Joglekar, S., Sastry, N. and Tyson, G. Characterising User content on a multi-lingual social network. In *Proceedings of the Intern. AAAI Conf. Web and Social Media* 14, 1 (May 2020), 2–11.
- Badrinathan, S. Educative Interventions to combat misinformation in India. *American Political Science Review* 115, 4 (2021), 1325–1341.
- Dhawan, A., Bhalla, M., Arora, D., Kaushal, R., Kumaraguru, P. FakeNewsIndia: A benchmark dataset of fake news incidents in India, collection methodology and impact assessment in social media. *J. Computer Communications*.
- Paris, B., Reynolds, R., Marcello, G. Disinformation detox: teaching and learning about mis- and disinformation using socio-technical systems research perspectives. *Information and Learning Sciences* 123, 1/2 (2022), 80–110; 10.1108/ILS-09-2021-0083.
- Singhal, S., Shah, R.R., and Kumaraguru, P. FactDrill: A data repository of fact-checked social media content to study fake news incidents in India. In *Proceedings of the 16th Intern. AAAI Conf. Web and Social Media* (Atlanta, GA, USA, June 6–9, 2022).

**Shivangi Singhal** is a Ph.D scholar at the Indraprastha Institute of Information Technology, Delhi, India.

**Rishabh Kaushal** is an assistant professor at the Indira Gandhi Delhi Technical University for Women, Delhi, India.

**Rajiv Ratn Shah** is an assistant professor at the Indraprastha Institute of Information Technology, Delhi, India.

**Ponnuram Kumaraguru** is a professor at The International Institute of Information Technology, Hyderabad, India.