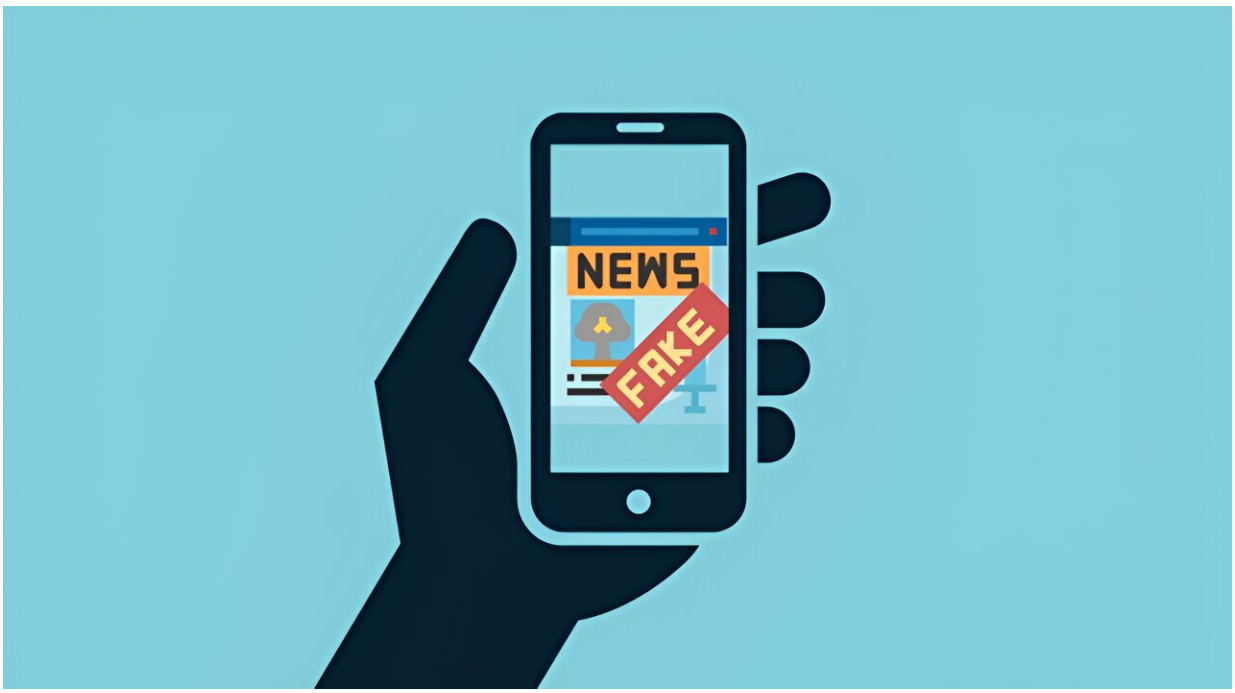
New algorithm enhances disinformation detection on social media

May 31 2024



Credit: IMDEA Networks

Disinformation is a growing phenomenon on digital platforms, significantly impacting social, political, and economic events. It has long posed a threat to freedom and democracy. However, it is now even more pressing due to the speed at which campaigns spread through digital media.

Researchers from IMDEA Networks, Cyprus University of Technology, and LSTECH ESPAÑA SL have developed the HyperGraphDis algorithm, which enables the detection of disinformation on social media , helping to combat the proliferation of fake news.

Dr. Marius Paraschiv, Senior Researcher at IMDEA Networks and one of the authors of the study, explains, "Our study 'HyperGraphDis: Leveraging Hypergraphs for Contextual and Social-Based Disinformation Detection' proposes a detection method that considers the complex social structures between users, as well as relational and semantic elements, to determine the nature of their generated content."

Dr. Nikolaos Laoutaris, Research Professor at IMDEA Networks, adds, "With ever-increasing volumes of social media data, achieving high accuracy in detecting fake news is not enough; detection algorithms need to be scalable and fast to handle large volumes of data in near real-time. Our new HyperGraphDis algorithm not only enhances detection accuracy but also significantly reduces execution time, making it much more practical than competing methods."

Specifically, the scientists evaluated four Twitter/X datasets on the 2016 US presidential elections and the COVID-19 pandemic. HyperGraphDis outperformed existing methods in both accuracy and computational efficiency, underscoring its effectiveness and scalability in tackling the challenges posed by the spread of disinformation.

The study reveals that disinformation is not always directly verifiable and depends on context. Therefore, Paraschiv emphasizes that it is crucial to "consider the background and analyze the relationships and environment of those disseminating the information, the communities to which they belong, or their relationship to known disinformation sources.

"Additionally, these sources are not necessarily the primary generators of a disinformation element; they may simply be propagators or amplifiers, serving as individuals who introduce this disinformation into a community where it is then amplified by other members."

How HyperGraphDis works

HyperGraphDis combines advanced techniques such as hypergraph neural networks, graph clustering for community detection, and natural language processing for text understanding. This allows for more efficient and accurate detection of disinformation.

The researchers focused on Twitter/X due to the availability of complex datasets, although HyperGraphDis can be adapted to other social platforms. Another advantage is that it offers platform owners an effective way to mitigate the effects of disinformation, providing a better understanding of how it spreads and how to combat it effectively by facilitating fact-verified and contextually tailored responses.

Challenges and future developments

The project faced numerous challenges, from collecting up-to-date data from Twitter/X (messages get deleted, accounts get removed, message content can be edited and not reflect the original meaning, etc.) to constructing complex hypergraphs. However, the researchers are already looking to the future: multimodal disinformation detection using advanced models like GPT-4.

"It may be the next logical step, but there are significant obstacles to overcome, including scaling issues and information aggregation from multiple sources," concludes Paraschiv.