'DeepFake-o-Meter' democratizes deepfake detection

September 11 2024, by Tom Dinki

Lyu works with postdoctoral researcher Shan Jai in the UB Media Forensics Lab. Credit: Meredith Forrest Kulwicki/University at Buffalo

When misleading information spreads online, it can spread fast.

Yet many of the best tools for swiftly debunking viral photographs, videos and audio are only available to researchers, like University at Buffalo deepfake expert Siwei Lyu.

"Everyone from social media users to journalists to law enforcement often has to go through someone like me to figure out if a piece of media shows signs of being generated by artificial intelligence," says Lyu, who routinely obliges such requests. "They can't get an immediate and conclusive analysis when time is of the essence."

That's why Lyu and his team at the UB Media Forensics Lab developed the DeepFake-o-Meter, which combines several state-of-the-art deepfake detection algorithms into one open-source, web-based platform. All users need to do is sign up for a free account and upload a media file. Results typically come back in less than a minute.

Since November, there have been more than 6,300 submissions to the platform. Media outlets used it to analyze various AI-generated content, from a Joe Biden robocall telling New Hampshire residents not to vote to a video of Ukrainian President Volodymyr Zelenskiy surrendering to Russia.

"The goal is to bridge the gap between the public and the research community," says Lu, Ph.D., SUNY Empire Innovation Professor in the Department of Computer Science and Engineering, within the UB School of Engineering and Applied Sciences. "Bringing social media users and researchers together is crucial to solving many of the problems posed by deepfakes."

How it works

Using the DeepFake-o-Meter is straightforward.

Drag and drop an image, video or audio file into the upload box. Then, select detection algorithms based on a variety of listed metrics, including accuracy, running time and the year it was developed.

Each algorithm will then give a percentage of the likelihood the content was AI generated.

"We do not make strong claims about the uploaded content. We simply provide a comprehensive analysis of it from a broad range of methods," says Lyu, who is also co-director of the UB Center for Information Integrity, which combats unreliable and misleading information online. "Users can then use this information to make their own decision about whether they think the content is real."

Transparency

Earlier this year, Poynter analyzed the fake Biden robocall with four free online deepfake detection tools. The DeepFake-o-Meter was most accurate, giving a 69.7% likelihood the audio was AI generated.

Lyu says the other things that set his tool apart are transparency and diversity. The DeepFake-o-Meter is open source, meaning the public has access to the algorithms' source codes, and features algorithms developed by both Lyu and other research groups across the globe, allowing for a broad range of opinions and expertise.

"Other tools' analysis may be accurate, but they do not disclose what algorithms they used to come to that conclusion and the user only sees one response, which could be biased," Lyu says. "We're trying to provide the maximum level of transparency and diversity with open-source codes from many different research groups."

A benefit to researchers, too

Before uploading a piece of media, the site will ask users if they want to share it with researchers.

Lyu and his team mostly train their algorithms on data sets compiled by themselves and other research teams, but he says it's crucial to expose the algorithms to media that's actually circulating online. Nearly 90% of the content uploaded to the DeepFake-o-Meter thus far was suspected of being fake by the user.

"New and more sophisticated deepfakes emerge all the time. The algorithms need to be continuously refined to stay up to the date," Lyu says. "For any research model to have a real-world impact, you need real-world data."

Future of the platform

Lyu hopes to augment the platform's capacity beyond spotting AI-generated content, like identifying the AI tools most likely used to create it in the first place. His group has previously developed such tools.

"This would provide clues to narrow down who is behind it," Lyu says. "Knowing a piece of media is synthetic or manipulated is not always enough. We need to know who is behind it and what is their intention."

Despite the promise of detection algorithms, he cautions that humans still have a large role to play. While algorithms can detect signs of manipulation that the human eye or ear never will, humans have a semantic knowledge of how reality works that algorithms often don't.

"We cannot rely solely on algorithms or humans," Lyu says. "We need

both."

That's why he hopes the DeepFake-O-Meter will eventually foster its own online community, with users communicating with and helping each other suss out AI-generated content.

"I like to think of it as a marketplace for deepfake bounty hunters," he says. "Because it's going to take a collective effort to solve the deepfake problem."

Provided by University at Buffalo

Citation: 'DeepFake-o-Meter' democratizes deepfake detection (2024, September 11) retrieved 11 September 2024 from https://techxplore.com/news/2024-09-deepfake-meter-democratizes.html