New tool uses AI to flag fake news for media fact-checkers

December 16 2019

Credit: University of Waterloo

A new artificial intelligence (AI) tool could help social media networks and news organizations weed out false stories.

The [tool](#), developed by researchers at the University of Waterloo, uses deep-learning AI algorithms to determine if claims made in posts or stories are supported by other posts and stories on the same subject.

"If they are, great, it's probably a real [story](#)," said Alexander Wong, a professor of systems design engineering at Waterloo. "But if most of the other material isn't supportive, it's a strong indication you're dealing with fake [news](#)."

Researchers were motivated to develop the tool by the proliferation of online posts and [news stories](#) that are fabricated to deceive or mislead readers, typically for political or economic gain.

Their system advances ongoing efforts to develop fully automated technology capable of detecting [fake news](#) by achieving 90 per cent accuracy in a key area of research known as stance detection.

Given a claim in one post or story and other posts and stories on the same subject that have been collected for comparison, the system can correctly determine if they support it or not nine out of 10 times.

That is a new benchmark for accuracy by researchers using a large dataset created for a 2017 scientific competition called the Fake News Challenge.

While scientists around the world continue to work towards a fully automated system, the Waterloo technology could be used as a screening tool by human fact-checkers at social media and [news organizations](#).

"It augments their capabilities and flags information that doesn't look quite right for verification," said Wong, a founding member of the Waterloo Artificial Intelligence Institute. "It isn't designed to replace people, but to help them fact-check faster and more reliably."

AI algorithms at the heart of the system were shown tens of thousands of claims paired with stories that either supported or didn't support them. Over time, the system learned to determine support or non-support itself when shown new claim-story pairs.

"We need to empower journalists to uncover truth and keep us informed," said Chris Dulhanty, a graduate student who led the project. "This represents one effort in a larger body of work to mitigate the spread of disinformation."

A paper on their work, "Taking a Stance on Fake News: Towards Automatic Disinformation Assessment via Deep Bidirectional Transformer Language Models for Stance Detection," was presented this month at the Conference on Neural Information Processing Systems in Vancouver.

  More information: Chris Dulhanty et al. Taking a Stance on Fake News: Towards Automatic Disinformation Assessment via Deep Bidirectional Transformer Language Models for Stance Detection. arXiv:1911.11951 [cs.CL]. arxiv.org/abs/1911.11951

Provided by University of Waterloo