Publication: The Straits Times, pg A25
Date: 20 June 2019
Headline: Can fake news law counter AI challenge?

Can fake news law counter AI challenge?

The rise of deepfakes shows how AI is a game changer. The Protection from Online Falsehoods and Manipulation Act is a starting point to fight it. But there are other tools, too.

Benjamin Tham and Josephine Seah

For The Straits Times

No one disputes that online falsehoods are a scourge on modern society. Since 2016, when "post-truth" was dubbed the word of the year, disinformation – the spreading of false information with intent – and misinformation – the spreading of false information without intent – have captured everyone's attention.

Countries from Germany to India to Australia have all either passed or tabled legislation meant to combat the harms brought by both of these.

These countries include Singapore. Last month, Parliament passed the Protection from Online Falsehoods and Manipulation Bill after a marathon debate over two days. The Protection from Online Falsehoods and Manipulation Act (Pofma) joins these efforts to address the challenges brought about by the dissemination of online falsehoods – an old phenomenon that has been repackaged in this age of social networks and artificial intelligence (AI). Singapore is, and will be, far from immune from the possibility of online falsehoods spreading and sowing discord, distrust and confusion.

But questions arise: Where does Pofma stand in relation to the rise of AI, and is it sufficiently equipped?

CHALLENGE OF DEEPFAKES AND THE LIKE

Among the various debates surrounding Pofma, one oft-forgotten aspect in the regulation of this difficult area is the impact of technology in the context of AI and the digital age.

The technology for the manipulation of texts, voices, pictures and videos is rapidly improving, and the pace of such growth is unlikely to slow down in the immediate future.

With AI increasingly easing the path for image and voice manipulation and the generation of fake videos – "deepfakes" – and images, we will soon have to grapple with an increasing number of legal and ethical conundrums arising from even more sophisticated technologies that will soon be more widely available. These include, for example, AI-generated texts, as well as deepfakes made with greater precision and finesse. Just recently, there have been deepfakes of Mona Lisa talking and of Facebook founder Mark Zuckerberg.

In the United States last week, a faked LinkedIn profile that used a generative adversarial network-generated photo – that is, it was created with the help of AI – was exposed with alleged connections to state espionage.

Already, the world is struggling with a cascade of information. Every minute, 400 videos are uploaded onto YouTube, 500,000 comments are posted on Facebook, 450,000 new tweets are added on Twitter, and 47,000 new posts appear on Instagram.

AI can change the rules of this game entirely. With it, one could easily create large volumes of articles called "neural fake news", cloak oneself under a convincingly fake persona, and disseminate these around the Internet at unprecedented speeds through fake accounts and bots.

There are currently developments along the lines of "fighting fire with fire": building AI models to sniff out synthetic photos, texts and videos.

For example, researchers are currently developing tools to combat deepfakes ahead of the 2020 US election by analysing speech patterns, head movements and facial expressions of the current primary candidates. While promising, many of these models are in their early stages of development. They could be some years away from an effective technological solution.

Is Pofma therefore sufficiently equipped to handle volumes of information, the speed at which information travels over the Web and the speed at which AI is developing?

SOME SUGGESTIONS

In this regard, Pofma is chasing a moving target, constrained by current understanding of how and why disinformation and misinformation spread.

For example, a correction direction or even a disabling direction will do little to stem the sharing of a screenshot of a text post, shared as an image rather than a text file, spread online through multiple uploads on a variety of social media platforms.

First, in the upcoming subsidiary legislation, we suggest implementing procedural measures to enable the targeting of AI-generated falsehoods efficiently and effectively. For example, provisions should be made to allow for the speedy identification and determination of online falsehoods generated and/or propagated by AI and also for effective service of any account restriction directions.

Second, we suggest provisions be made that would require the measures provided to be reviewed every 24 months. Regular and constant calibration of Pofma's weaponry is essential for a right balance to be struck between ensuring fundamental rights are not compromised, and that Pofma is well equipped to combat the scourge of online falsehoods. This is imperative in the regulation of a constantly evolving piece of technology. If Pofma's weaponry is found to be insufficient to cope with the technological advances that keep emerging, then it ought to be strengthened, and if, conversely, it is found to be too powerful for its aims and purposes, then it ought to be toned down.

In this regard, we further suggest that a select committee can similarly be convened for the purpose of such review, allowing various interest groups and stakeholders to make representations.

MUTUAL TRUST AND VERIFICATION

AI is poised to challenge information landscapes. But the problem is not intractable.

Rather than give in to fears or worst-case scenarios, we should not forget that these tools are – for now – still simply tools. Regulation cannot be our first line of defence. But Pofma also helpfully lays out stakeholders who can rise to the challenge of online falsehoods.

Media literacy efforts can be enhanced to cultivate a discerning public. Internet intermediaries can develop their own tools to identify fake news and accounts. Digital advertising intermediaries can reveal if ads are being bought by political campaigns.

Pofma is best viewed as a starting point that compels us to expect more of ourselves and the society in which we live.

As we move ahead, it is not just ministerial wielding of Pofma that should be debated, but also efforts to build trust among ourselves. After all, an ounce of prevention is always worth a pound of cure.

There is work for all here: The responsibility to confront fake news, as the Pofma debates remind us, is societal.

* Benjamin Tham and Josephine Seah are research associates at the Centre for AI and Data Governance, School of Law, Singapore Management University.