

Fake news law: Calibrated approach, but nimbleness needed

Consider the reach of false information in deciding on course of action, and require tech firms to use AI to flag dubious accounts

Carol Soon and Shawn Goh

For *The Straits Times*

On Monday, Minister for Law and Home Affairs K. Shanmugam tabled the Protection from Online Falsehoods and Manipulation Bill in Parliament. The Bill was almost two years in the making, since the Government first mentioned in 2017 that it would legislate against "fake news".

The tabling of the Bill comes after regulatory moves around the world such as in Germany, France and, more recently, Australia. Even Facebook chief executive Mark Zuckerberg has called for governments to be more involved in regulating the Internet. However, underscoring the Bill is an acknowledgement that legislation on its own is insufficient and a whole-of-society approach is required to increase people's competency to tackle the scourge.

These components of the Bill stand out – interventions for falsehoods that are published online, and the prevention of the publication and dissemination of falsehoods through a proposed Code of Practice for tech companies. The Bill provides a calibrated response to the problem of online falsehoods. Nevertheless, more can be done to ensure nimbleness in intervention and prevention.

STEMMING THE SPREAD

One clear merit of the Bill is that it allows for calibrated responses to counter online falsehoods.

First, the Bill targets only "false or misleading statements of facts", and not opinions, criticisms, satire or parody. It also provides a right of appeal to the High Court and judicial review to minimise potential censorship. This is

important as one key concern flagged by many written representations to the Select Committee, including the authors', is that legislation should not stifle free speech.

Second, take-down directions are reserved for extreme situations where falsehoods pose egregious harm to society. The take-down order, which is part of the Stop Communication Direction, would require the person who communicated the false statement to remove it from an online location within a specified time period.

Besides considering the falsity of a piece of information, a take-down direction will also consider if certain thresholds of harm have been breached and whether public interest has been undermined.

This aims to safeguard public interest in areas including security, election outcomes, public safety and public health. One possible example would be to target anti-vaccination misinformation that may potentially put Singapore at risk of an epidemic similar to that in the United States.

However, an over-reliance on take-down may risk creating a Streisand Effect – named after entertainer Barbra Streisand, who attracted more public attention to her residence when she tried to prevent the release of its photographs – where attempts to suppress a falsehood result in an unintended consequence of drawing more attention to the falsehood instead. Thus, on top of considering the criteria of falsity and public interest, the potential reach of a falsehood should also be considered.

As part of a 5Cs framework to guide legislative action in our proposal to the Select Committee last year, we had recommended that "consequence" – in terms of the extent and magnitude of impact of a falsehood – should be considered.

Some falsehoods die a natural



The recent exhibition, *Our Next Battlefield*, curated by the Singapore Discovery Centre to mark Singapore's Total Defence Day, which falls on Feb 15, focused on fake news, disinformation and cyberthreats. ST FILE PHOTO

death because they gain little or no traction with audiences, or are perceived to be too implausible. Such small flames should not be given the oxygen to grow bigger.

It is thus heartening to note that, third, the Bill recognises corrections as the primary remedy against online falsehoods. To protect public interest, the Government can request relevant actors to make either targeted corrections that are published alongside a falsehood, or general corrections that are amplified on multiple platforms to inform the public of the existence of a falsehood and its corresponding correction.

This has value in cultivating an informed citizenry in two ways. First, by not taking down a falsehood, the room for debate and deliberation remains for Singaporeans to see all sides of an issue and make their own sound assessments. Second, the existence

of a falsehood could play an important part in educating the public on the misconceptions surrounding the issue at hand. One good example is the Central Provident Fund WhatsApp rumour that went viral last year, where corrections provided timely clarifications on Singaporeans' CPF withdrawals.

PREVENTION IS BETTER THAN CURE

However, the devil is in the details when it comes to optimising the effectiveness of corrections. This is where governments suffer from information asymmetry because platform companies have the technological expertise and access to user data to know what works better.

For instance, after a brief trial in 2017, Facebook replaced its Disputed Flags feature with Related Articles after realising that providing more context to a story is more helpful for users to identify

falsehoods than to flag them with warning labels, and more effective in reducing the number of shares of a falsehood. In other words, it is essential for the Government to work closely with technology companies and leverage their intelligence in this fight.

Thus, it is a welcome move that the Bill also provides for a Code of Practice that ensures platform companies have adequate systems and processes in place to prevent the abuse of their platforms by malicious actors and to enhance the integrity of information online.

The proposed Code echoes the European Commission's Code of Practice on Disinformation announced in September last year, which aims to provide safeguards against the abuse of inauthentic accounts and bots, prevent the abuse of digital advertising tools to spread falsehoods, and de-prioritise the visibility of falsehoods on platforms.

Such upstream measures stress the importance of a proactive approach to curb falsehoods at the point of publication before they can gain any traction or before corrections or take-down directions need to be issued by the Government.

However, while the proposed Code of Practice covers essential areas, it could go a step further by calling for greater efforts in detection. Just as algorithms are deployed to predict and personalise information for users based on their online behaviours and engagement with content, the same can be done to detect if certain users or groups are prone to producing and sharing online falsehoods.

A recent *New York Times* article reported that Russians have tried to circumvent Facebook's existing safeguards and undermine Ukraine's upcoming presidential election by paying Ukrainian citizens to give Russian agents access to their Facebook accounts. In such cases, artificial intelligence may be useful in identifying abrupt changes in users' online behaviours or the content they share, and monitoring for potential falsehoods. Perhaps some may think such strategies bear Orwellian undertones, but detection with ethics by platform companies will play an increasing part in the changing techscape.

Technology will continue to evolve and user behaviours will adapt to new technology. Thus, laws will always have to play catch-up with technology.

The good fight rests with all of us who consume, produce and share content. While new legislation is taking shape, Singaporeans should not be misled into thinking that we can now rest on our laurels in the fight against online falsehoods.

stopinion@sph.com.sg

• Carol Soon is senior research fellow and Shawn Goh is research assistant at the Institute of Policy Studies. They had presented a written submission to the Select Committee on Deliberate Online Falsehoods last year.