How Effective is POFMA in Battling Online Falsehoods?

By Benjamin Ang and Xue Zhang

SYNOPSIS

A study conducted in 2023 found that the Protection from Online Falsehoods and Manipulation Act (POFMA) has been effective in mitigating the risks arising from the spread of online falsehoods, which can undermine social cohesion. Media companies and the public must also play their part in this effort.

COMMENTARY

In his speech at the recent launch of Smart Nation 2.0, Prime Minister Lawrence Wong alluded to real-world risks from information and communication technology (ICT) advancements, such as the undermining of social cohesion through disseminating falsehoods. The Southport riots of July 2024 is a case in point. A flood of online falsehoods that falsely described the killer of three girls as being a Muslim immigrant triggered one of the UK's most severe riots in 13 years. Another example is the deliberate dissemination of online falsehoods in Bangladesh in August, which inflamed tensions between Hindus and Muslims.

Grappling With Online Falsehoods

Although many countries have enacted laws against the creation and dissemination of online misinformation and disinformation, there are challenges to be overcome. For instance, while the UK government relied on the Online Safety Act 2023 to prosecute those involved in spreading falsehoods online during the UK riots, difficulties were faced in proving that the sender knowingly sent a false message, especially in social media cases, where the nuances and context in language can easily change the meaning of a message.

POFMA's Continuing Relevance

In October 2019, the government enacted the Protection from Online Falsehoods and Manipulation Act ([POFMA](#)), which aims to prevent the electronic communication of misinformation and disinformation and to safeguard against the use of online platforms to communicate falsehoods and information manipulation.

From the time POFMA was proposed and throughout its enactment and implementation, critics have perceived it negatively as the government's way of curtailing [freedom of speech](#) and [undermining independent thought](#). In practice, it was more frequently used during the COVID-19 pandemic to correct online posts that could have caused panic and impeded public health measures. As of 30 June 2024, the POFMA office has handled [66 cases](#) and issued 114 correction directions.

In the leadup to POFMA's fifth anniversary, a team of researchers from the Centre of Excellence for National Security (CENS) at the S. Rajaratnam School of International Studies (RSIS) carried out a study in 2023 (publication forthcoming) to explore how the public perceived POFMA. A total of 1,004 responses (including those that had skipped certain questions) were analysed.

In a nutshell, POFMA was found to be effective in preventing the spread of online falsehoods, with over half of the respondents agreeing or strongly agreeing on this. As to whether POFMA was effective in stopping falsehood creation, slightly more than one-third of the respondents agreed or strongly agreed.

The study also aimed to determine the respondents' views regarding the believability of POFMA clarifications. In each of the ten use cases, more than 80 per cent of the respondents responded that they believed in the POFMA clarification rather than the original post.

These findings suggest that POFMA is "fit for purpose" in combating online falsehoods. They also show that respondents have a high level of trust in the fact checks provided with POFMA correction directions, which indicates that such fact checks help separate fact from falsehood. At the same time, the findings show that there is a not insignificant segment of the population that is susceptible to falsehoods. The study suggests that more has to be done to prevent a further erosion of the infrastructure of fact that supports our democracy.

AI-Enabled Disinformation

The risks of real-world harms resulting from online falsehoods have been exacerbated by AI-enabled content, particularly deepfakes. Deepfakes use AI to create hyper-realistic images, audio, or video clips of notables, which can manipulate public perceptions of critical social and political issues, exacerbate social tensions, inflame social conflict, and even undermine a nation during times of crisis.

In 2022, for example, a deepfake of Ukrainian President [Volodymyr Zelensky](#) telling his soldiers to lay down their weapons was circulated online, presumably to "sow panic and confusion". In November 2023, a deepfake of London Mayor [Sadiq Khan's](#) voice allegedly making inflammatory comments could have sparked serious public

disorder. And, in June this year, then Prime Minister Lee Hsien Loong cautioned that a deepfake video of him purportedly commenting on foreign leaders and international relations was circulating online. He added that the deepfake was a malicious attempt to create the impression that the views were supported by him and/or endorsed by the Singapore government, which is "dangerous and potentially harmful to…[Singapore's]…national interests".

The impact of AI-enabled disinformation has been a concern in elections held in the UK, India and Indonesia this year. It will also be a concern in the US presidential election this November. While legislative and other measures to address the risks of AI-enabled disinformation have been proposed, such as Singapore's Elections (Integrity of Online Advertising) (Amendment) Bill, how to minimise the spreading of online falsehoods remains challenging.

Conclusion

Addressing this challenge requires a holistic approach. In addition to legislative tools like POFMA, with correction directions as the key means to counter online falsehoods, media companies and the general public also have roles to play.

Media platforms, for instance, should further enhance their review mechanisms to enable them to respond faster in identifying and acting against misinformation.

Members of the public need to exercise critical thinking and responsibility in sharing information they receive or encounter online. Fact checks by trusted sources can help evaluate posts read online, especially those that evoke emotional responses. Individuals should avoid sharing such posts unless they are sure of the veracity of the content. The effectiveness of POFMA correction directions and fact-checking will be limited if individuals persist in spreading online falsehoods even after they have been flagged and debunked.

Benjamin Ang is a Senior Fellow and Head of the Centre of Excellence for National Security (CENS) at S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University, Singapore. He is also the Head of Digital Impact Research at RSIS and leads the Future Issues in Technology programme. Xue Zhang is a Research Fellow at CENS. Both authors wish to acknowledge Dymples Leong and Sean Tan's contributions to the survey.