Ministry of Communications and Information
An Engaged and Connected Singapore

OPENING ADDRESS BY MRS JOSEPHINE TEO, MINISTER FOR COMMUNICATIONS AND INFORMATION, AT THE ONLINE TRUST AND SAFETY FORUM

ON 15 MAY 2024, 11.05AM

Driving Singapore's Progress Towards a Safer Internet

Dr Yang Yinping, Director for the Centre for Advanced Technologies in Online Safety

Colleagues and friends

1. Thank you for inviting me to speak at today's Forum.

2. This time last week, I was on a train from Washington DC to New York. I had been on a panel at an AI Expo organised by the Special Competitive Studies Project and was due to speak on the first-ever Global Roundtable on Capacity Building on ICT Security under the auspices of the United Nations.

3. At international conferences on all aspects of digital development, there is often an underlying question, of how our economies can benefit from widespread use of technologies, while our societies are shielded from their worst excesses.

4. Concerns around risks of AI have become especially prominent, even as we are still grappling with the ever-growing cyber risks.

5. A giant of sci-fi, Arthur C Clarke: wrote in his most famous book <2001: A Space Odyssey> that "Any sufficiently advanced technology is indistinguishable from magic".

6. Can present-day experiences with digital technologies be described as magic? I think the answer is yes, in many ways. But ask any scams victim, the parent of online child sexual abuse, communities targeted by hate speech or voters besieged by disinformation. It's hard to feel magical anymore. The sense that we sometimes get is one of bewilderment, anger and potentially despair.

7. Even if we have not personally encountered the worst online harms, their prevalence must surely take a toll on our collective sense of well-being. A sampling of

news that have become commonplace to the point that we may even have become immune to it:

a. In 2021, a 25-year-old influencer from China, who was suffering from depression, was egged on by her followers to drink pesticide during a live-streamed video. She died the next day.

b. In December 2023, a child in Texas allegedly died by suicide following cyberbullying. When he passed away, he appeared to be online and was still wearing a gaming headset. The suspected perpetrator was another boy in another state, Michigan, i.e. more than 1000 miles away.

c. In Singapore, many of us have come across deepfake videos that feature our political leaders supposedly advocating investment schemes with guaranteed returns.

d. My image too has been used in scams, so it is not unusual if you feel like you have seen me before, when I asked you to 'buy something' or 'believe in something'.

8. It was therefore not surprising that when MCI conducted an Online Safety Poll in 2023, about two-thirds of Singapore users said they had encountered harmful online content, with the most common being cyberbullying (29%) and sexual content (28%).

9. On a global level, The World Economic Forum's Global Risk Report 2024 has identified AI-driven misinformation and disinformation as the top global risk over the next two years.

10. These findings underscore the pressing need for concerted action to address online harms and protect our communities. There is one risk in particular that we must fight, and that is the growing sense of helplessness and simultaneously, pushback against reasonable attempts to tame the beast.

11. This has sometimes been seen through blanket rejection of new laws and regulations, especially in regard to misinformation and disinformation, through distorted

arguments about their chilling effect on free speech, for example, or insinuations that they are instruments for silencing critics.

12. These accusations can be a formidable barrier against meaningful action. There is little governments can do to except to build trust with citizens, through serious efforts to activate stakeholders to collectively identify counter-measures that work, and engage the public to seek understanding and support. This has in fact been Singapore's approach.

Singapore’s approach towards building online trust

13. Singapore has consistently adopted an accretive and multi-pronged approach to build trust and safety in our digital spaces.

14. Instead of relying on big-bang legislation, we have taken steps to develop guardrails and build foundations as we go along.

15. As early as 2012, we introduced the Personal Data Protection Act - a set of rules on how personal information should be handled. The PDPA exemplifies our pro-innovation data regime, protecting personal data but also promoting legitimate uses.

16. We passed the Cybersecurity Act in 2018, and was one of the first jurisdictions in the world to introduce such legislations to protect our critical information infrastructure (CII) against cyber threats.

17. Just last week, we amended the Act to stay ahead of newly emerging cybersecurity challenges. The amended Act requires CII owners to report more types of incidents including those that happen in their supply chains, and allows CSA to manage entities beyond CIIs.

18. In 2019, we introduced the Protection from Online Falsehoods and Manipulation Act (POFMA).

19. We then introduced the Online Safety (Miscellaneous Amendments) Act in 2022 to better protect Singapore users from harmful online content on services such as social media platforms, and passed the Online Criminal Harms Act last year, to deal more effectively with online activities that are criminal in nature.

20. In addition to legislative measures, Singapore has prioritised empowering individuals with the knowledge and skills to navigate the digital landscape safely.

21. Initiatives such as the National Library Board's S.U.R.E program and the Digital Skills for Life framework that was launched earlier this year help to equip our citizens with tools to critically evaluate information and protect themselves from online threats.

Launch of CATOS

22. The Government's consistent belief is that Singapore's approach to building trust and safety online must be grounded in knowledge and understanding of the transmission mechanisms, and how the design of systems and services can lead to unintended consequences.

23. We must have the ability to detect these risks as well as to research the effectiveness of mitigating measures. We must also recognise the effects of local context, where observations in Singapore may not mirror those elsewhere.

24. These considerations have led us to invest in growing our own research capabilities. In 2022, we launched the Digital Trust Centre (DTC) to promote research, translation, and talent development in trust technologies.

25. DTC focuses on trusted data sharing and computation, digital identity, and algorithms to evaluate the trustworthiness of systems. It aims to strengthen Singapore's role as a trusted hub in the digital economy.

26. An equally important and complementary effort will be through the Centre for Advanced Technologies in Online Safety, or CATOS, which I'm happy to launch today.

27. A centre hosted by A*STAR, CATOS will be another useful tool in our arsenal in the battle against online harms, and will inform our efforts to make the Internet a safer place for our citizens.

28. CATOS, which has been about a year in the making, demonstrates:

a. Singapore’s belief that technology can do good; and that we can leverage technology to overcome human limitations; and fight the harms of technology with technology.

b. CATOS also demonstrates our confidence in our research community; and that by gathering the top minds here and around the world, we can develop innovative approaches to address the challenges of our time.

c. Many of the technological capabilities developed in CATOS will help detect online harms in our local and regional contexts.

d. For example, recent tests have shown that the CATOS-developed analysis engines were able to accurately detect intense emotions, such as fear, anger and sadness, and the presence of hate speech in online expressions collected from popular social media and online platforms. You may ask, we as humans can detect that too, so what would be the value of this? I think it is because the tools can potentially address the problem of pervasiveness of harms. These tools can potentially help to flag polarising exchanges before they cause escalation in tensions in places that we least expect.

e. It is likely that the tools we use can also be adapted for other contexts, and allow us to use Tech for the Public Good, not just in Singapore, but to benefit societies elsewhere.

29. Therefore, the Government has committed a total of $50 million dollars to fund CATOS over five years. The mission of CATOS is to develop cutting-edge technologies that prevent, detect, and mitigate online harms.

30. This ranges from misinformation, online hate and discrimination, harm to health and well-being, and threats to personal and community safety – all of which are crucial to ensuring our sense of well-being when engaging online. I’m told we will be able to see demonstrations of CATOS’ work in the showcase outside.

31. Collaboration is also high on CATOS’ agenda. It will work with partners to develop an Online Trust and Safety sandbox for stakeholders to access and trial new tools for detecting various forms of online harms.

32. Our aim is to accelerate the growth of the online trust and safety ecosystem here in Singapore, through agile and timely collaboration, joint experimentation and fine-tuning of solutions.

a. Let me share an example of how this is already happening. CheckMate is a fully home-grown, volunteer-driven initiative to counter misinformation. It allows anyone who wishes to check or report a dubious message to send it to a WhatsApp number.

b. Behind the scenes, CheckMate uses a combination of machine learning and crowdsourcing to assess the information and respond to the query accordingly.

c. CATOS will be partnering CheckMate to explore technologies and tools that can help CheckMate and its group of volunteers to fact-check more quickly, effectively, and at scale. In other words, it should be able to deal with not just a few queries, it needs to be able to follow up on multiple queries at the same time at scale and accurately, in order for people to feel that they can rely on the tools.

d. If and when they succeed, we as WhatsApp users can have an added means of checking suspicious messages, anytime, anywhere.

Conclusion

33. Let me conclude my remarks.

34. Colleagues and friends, the scale and complexity of online harms demand a concerted and coherent response from all stakeholders.

35. CATOS will be a platform for constructive partnerships between key players of the online trust and safety ecosystem in Singapore.

36. Regardless of which organisation you represent – whether it is a government agency, social media platform, trust and safety service provider, academic institution, or

NGO – I encourage you to be part of this platform and help make the Internet a safer place for all.

37. Thank you very much for being here.

+++