RSIS S. RAJARATNAM SCHOOL OF INTERNATIONAL STUDIES Nanyang Technological University

RSIS COMMENTARY

www.rsis.edu.sg

No. 182 – 3 October 2017

*RSIS Commentary is a platform to provide timely and, where appropriate, policy-relevant commentary and analysis of topical issues and contemporary developments. The views of the authors are their own and do not represent the official position of the S. Rajaratnam School of International Studies, NTU. These commentaries may be reproduced electronically or in print with prior permission from RSIS and due recognition to the author(s) and RSIS. Please email: RSISPublications@ntu.edu.sg for feedback to the Editor RSIS Commentary, Yang Razali Kassim.*

# Chatbots: Friend or Fiend?

*By Muhammad Faizal Bin Abdul Rahman and V S Suguna*

## Synopsis

*The world is seeing how Internet technologies can facilitate crime and terrorism for the purposes of victimisation, illicit communication and online radicalisation. As advances in Artificial Intelligence (AI) drive the growing adoption of chatbots, it may be a matter of time before threat actors jump on the bandwagon to stay ahead of law enforcement and security agencies.*

## Commentary

CRIMINALS AND violent extremists have demonstrated swiftness in adopting new technologies and innovative tactics to outsmart security agencies and widen their reach. Since the 9/11 attacks for instance, extremist groups such as ISIS have been particularly adept in leveraging advances in internet technologies – from early websites to social media to messaging apps - for the purposes of online radicalisation and planning terror attacks.

The next frontier of internet technologies – Chatbots – may shape the future chapter of the criminals' and terrorists' playbooks. As advances in AI and Machine Learning (ML) improve the availability and capabilities of Chatbots, the technology may plausibly become a powerful tool for cyber-enabled crimes and online radicalisation.

## Building "Human" Connections

Chatbots are projected to surpass Web 2.0 (social media and messaging apps) as digital assistants on smart devices and desktops that mimic conversations to create more engaging interaction experiences for public, commercial and social applications. Indeed, the social chatbot "Mitsuku" – accessible online -upon query describes itself as "a computer programme designed to talk to you".

Presently, the technology is being explored in numerous industries. Chatbots have been developed by Georgia Institute of Technology as a teaching assistant (Jill Watson) and by the Singapore government to answer public queries (Jamie).

With ML and access to knowledge from a multitude of internet sources and global pool of users, chatbots are envisioned to over time become more human-like by learning from human behaviour and be more succinct in its answers. For instance, the social chatbot “Eugene Goostman” imitated a human teenager online and reportedly fared well in the “Turing test” which judges the indistinguishability of intelligent conversation between machines and humans.

Given the utility, chatbots as a platform for government agencies and businesses to engage with their target audiences, Internet-of-Things (IoT) interface for smart homes, and “virtual friends” would become ubiquitous in the foreseeable future.

“Virtual friendships” that chatbots can potentially offer would be a natural evolution of the current popularity of Web 2.0 particularly among the digital natives; these are currently young people who are inherently accustomed to embracing new technologies to fulfill their practical (e.g. education and communication) and social needs.

The key features of chatbots - human-like and unconstrained by time and space – can help to drive human-machine interactions that would meet the intrinsic social needs for belonging, acceptance and friendship as defined by Maslow. Thus, it is unsurprising that Kaspersky Lab’s blog on “the dangerous future of chatbots” foreshadowed the technology’s ability to learn and influence human behaviour as “a goldmine for social engineering and crime”.

**Virtual Threat Agents**

Similar to human threat actors, chatbots may be instrumental in criminal and extremist enterprises to build rapport with their respective target audiences for the purposes of victimisation (e.g. phishing and other online scams) and online radicalisation. Indeed, the tactic of using software robots that mimic humans for propagating radical views and garnering support is hardly unprecedented given the reported use of pro-Trump twitter bots during the 2016 US presidential campaign.

Similar to how ISIS demonstrated greater proficiency than Al Qaeda in online radicalisation, the use of chatbots may plausibly be one of the tools that would be integral to what comes after the imminent fall of ISIS. Such chatbots can potentially be programmed with a basic database of extremist narratives and responses to manipulate the psychological vulnerabilities and social grievances of people who are seeking answers and support.

Even without explicit programming, the case of social chatbot Tay unexpectedly sprouting neo-Nazi remarks exemplifies the risk that AI can teach itself – through online conversations - to spread radical propaganda.

Hence, chatbots in the wrong hands and like any technology can facilitate security threats. By functioning as robot extremists spreading radical propaganda, chatbots can support or even supplant online (human) extremists who have been detained by security and intelligence agencies during counterterrorism operations.

## Exposing the Wolves

Governments and communities need to begin anticipating the possible risk scenarios that may emanate from the imminent ubiquity of chatbots and plan for strategies to address the risks while the technology is still nascent.

First, security agencies need to collaborate with communities and the industry both in the practical uses of chatbots amid smart cities initiatives and also in developing mechanisms to identify and contain chatbots that may be malicious by design or have gone rogue. This includes deploying chatbots to help counter extremist propaganda coming from malicious chatbots.

Second, existing education on cyber wellness and internet literacy would have to keep pace with technological changes including in promoting the safe use of chatbots and equipping the general population with the knowhow to identify and stay away from malicious chatbots.

Third, security agencies need to develop a framework that helps to decide when certain chatbots should be allowed to remain online for purposes of gathering intelligence on the threat actors and their sponsors, and when a certain threat threshold is crossed thus necessitating the removal of the chatbots.

In sum, the complex challenges that are being faced in confronting contemporary threats facilitated by technology point to the need for a multi-pronged approach in addressing the plausible criminal and terrorist threats that may emanate from the imminent ubiquity of chatbots. This would include a whole of society involvement besides legislative and law enforcement measures.

*Muhammad Faizal bin Abdul Rahman is a Research Fellow with the Homeland Defence Programme and V S Suguna is an Associate Research Fellow at the Centre of Excellence for National Security (CENS), a unit of the S. Rajaratnam School of International Studies (RSIS), Nanyang Technological University, Singapore.*