Source: *The Straits Times*, pB3
Date: 28 October 2025

Chatbots could be first responder for mental health issues, but with limits

Any tool designed for mental health should state in plain terms what it does and what it does not do. A chatbot can teach micro skills, explain common problems in everyday language and encourage first steps. It should not be left to manage acute risk or decide whether someone has a disorder, says the writer. ST FILE PHOTO

Singapore needs clear standards to ensure safe and effective use of chatbots, with a hybrid approach that includes humans who can step in when needed.

Nur Hani Zainal

On the MRT after a long day, a commuter opens a mental health chatbot on her phone. She is not looking for a diagnosis. She wants to get through the evening without the knot in her chest taking over.

The bot checks in, offers a short breathing exercise, and suggests the next step if the tightness does not ease in an hour. It is not therapy. It is a well-timed nudge.

Scenes like this are now common. AI chatbots sit in pockets, on laptops and in workplace portals. They are fast, always on and often free. For many, they are becoming the first stop for worry, low mood or insomnia.

Some users find them helpful. Others are disappointed. A minority may be harmed, especially when the system overpromises or fails to recognise a moment when a trained human should step in. Overseas incidents underscore the stakes.

The National Eating Disorders Association in the US suspended its "Tessa" chatbot after it gave advice that could worsen eating disorders. In Belgium, a family reported that prolonged conversations with a consumer chatbot preceded a man's death by suicide.

More recently, the parents of 16-year-old Adam Raine in California filed a lawsuit alleging that the teen's prolonged interactions with ChatGPT validated his self-destructive thoughts before he took his own life. The case has renewed calls for stronger safety checks in AI systems. These cases prompted providers to review how their systems detect distress and escalate risk.

In Singapore, the Government's mindline.sg shows both the opportunity and the challenge. It is an anonymous digital mental health resource with an AI-enabled chatbot that delivers brief therapeutic exercises.

Independent analyses of usage patterns have found that its dialogue-based exercises are among the platform's most used features, suggesting that many people prefer guided, bite-sized help.

The question is not whether people will use mental health chatbots. They already do. The question is how to make these tools safe and useful, and how to help people understand their limits.

October is a month when many places spotlight mental health. It is a good moment to insist that access and safety grow in tandem.

Chatbots will not replace therapists. They can, however, become reliable first responders for brief skills and timely nudges, provided they operate within clear lines and hand over to humans when a person's needs exceed the bot's scope.

THE PROS AND CONS OF BOTS

Chatbots excel at micro skills that reduce friction when emotions surge. A one-minute grounding prompt before a presentation. Two minutes of paced breathing when panic rises.

When designed well, such just-in-time support lowers the opportunity costs that keep many from seeking care, including shift workers, caregivers, students with tight schedules, and those who would not otherwise walk into a clinic.

However, chatbots can confuse confidence with competence and speak too smoothly about things they do not really know.

They can offer reassurances that sound caring but are wrong for the situation. They can fail to detect risks hidden in subtle phrasing, misread sarcasm or miss code-switching.

These are ordinary failure modes of systems that predict the next word rather than understand a life in context. These risks are not abstract; the recent overseas incidents show how overly fluent systems can miss red flags with real-world consequences.

Public sector chatbots have had misfires here at home, reminding us that generic systems can feel ill-matched to local needs. In 2021, for example, the "Ask Jamie" bot on a government site was taken down after widely shared gaffes.

Certain groups are more vulnerable to these pitfalls. Young people who take language literally or who struggle with self-regulation can overtrust a fluent system and miss its boundaries.

People under acute stress can interpret polite suggestions as prescriptions and feel worse when they cannot act on them. In multilingual settings, even slight shifts in phrasing can carry different emotional weight across languages.

THE HUMAN FACTOR STILL MATTERS

There is a straightforward fix. Pair the bot with a human.

Across large pools of trials of internet-delivered cognitive behavioural therapy, programmes that included human guidance showed stronger and more durable benefits than those that left people to navigate on their own.

This was also a pattern observed across multi-country randomised trials over the past two decades, including work in Asia, Europe and North America. Guidance amplifies outcomes.

A trained coach or clinician can transform a generic breathing script into a plan that fits a person's schedule, can spot stall points and can redirect to care when an app's tips are no longer enough.

In clinics, a bot can reinforce skills between therapy sessions and collect short check-ins. The clinician then adjusts intensity. In workplaces and schools, a coach can follow up on a bot's prompt, offer a brief human exchange, and escalate quickly when risk appears.

Singapore's institutions are already exploring digital tools. The next step is to move from pilots to a clear, shared standard for mental health chatbots. At present, Singapore regulates software as a medical device when it claims clinical functions, and it has national model frameworks for responsible AI.

These are important, yet there is no chatbot-specific standard for mental health that spells out required guardrails, escalation pathways and transparency metrics. That gap is new but fixable.

ESSENTIALS FOR SAFE USE

Any tool designed for mental health should state in plain terms what it does and what it does not do. A chatbot can teach micro skills, explain common problems in everyday language and encourage first steps. It should not be left to manage acute risk or decide whether someone has a disorder.

This means no diagnosis, no crisis handling, no replacement of clinical judgment. When risk arises, the bot should detect and flag distress, then quickly hand off to a trained human.

Institutions that deploy chatbots should ensure that escalation pathways reach a person in minutes, not days. The first responder need not always be a specialist. Coaches who are trained, supervised and embedded in a service can cover a great deal safely and at scale.

In clinical services, escalation should reach clinicians who can assess and treat. In schools and workplaces, escalations should be routed to a designated support team with clear links to community resources.

Privacy deserves equal attention. Tools should collect only what they need to deliver the service. They should explain what is stored on the device and what is in the cloud, who has access and for how long.

Content should be co-created with people who have lived experience of anxiety and depression, and with caregivers. These are not nice-to-haves. They are what make tools credible and humane.

Finally, publish a few safety and engagement metrics in a dashboard that users and families can understand. How often does the bot miss a risk that later requires help? How quickly do escalations reach a human? How many users drop off after the first week? Do outcomes differ by age, language or other factors that signal inequity?

These are practical steps any large employer, school or healthcare provider can take.

Regulators, payers and professional bodies also have a role. Singapore has a strong record of setting technology guardrails while enabling innovation. The same approach can work here through a compact between public agencies, providers and community groups.

Other jurisdictions are beginning to legislate transparency for health-related chatbots, including requirements that AI systems disclose their identity and that operators report how they identify and respond to self-harm risk. We can learn from these moves and calibrate them to our context.

Some will ask whether any of this is necessary, given that chatbots help a bit and harm rarely. The answer lies in trust, which is hard to build and easy to lose. Mental health care depends on it.

• Nur Hani Zainal is a presidential young professor in clinical psychology at the Faculty of Arts and Social Sciences, National University of Singapore, and director of the Optimizing Wellness Lab. Her research focuses on digital mental health interventions.

Source: The Straits Times © SPH Media Limited. Permission required for reproduction.