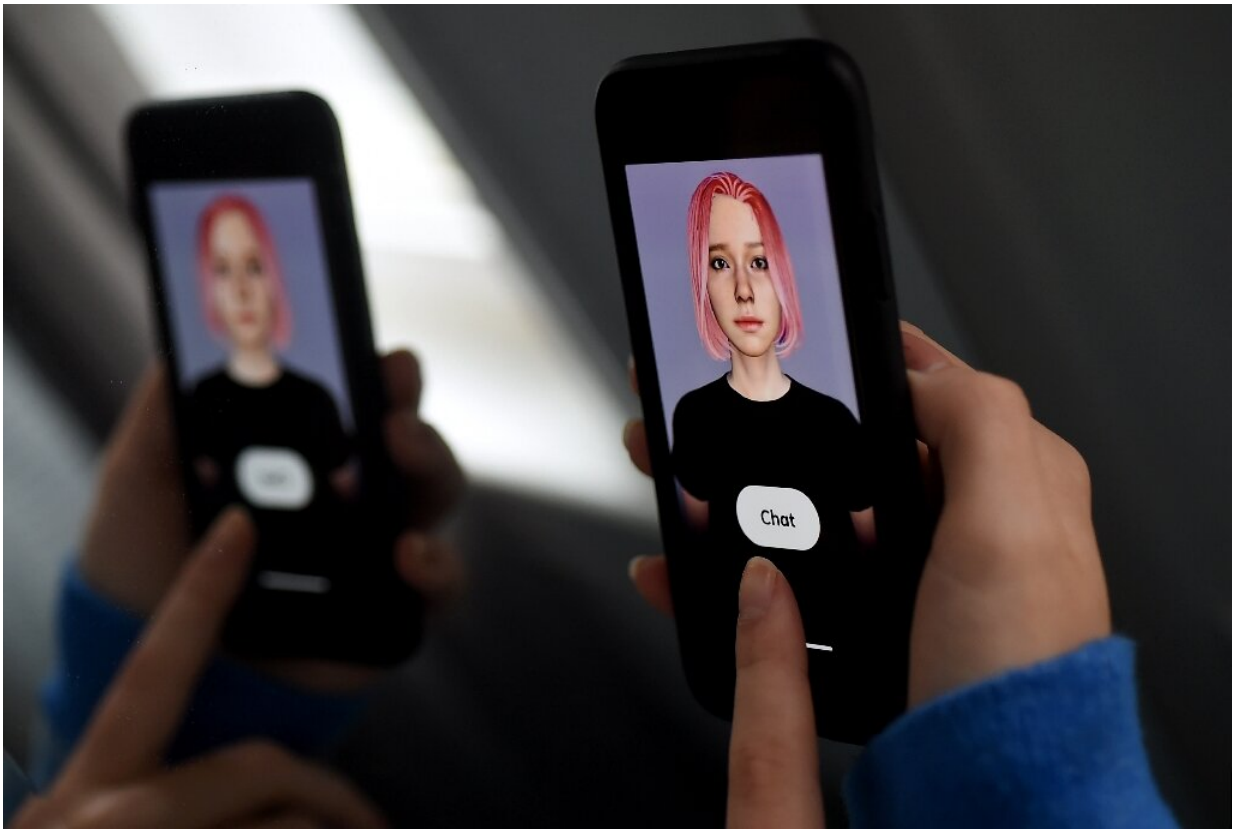
Sexting chatbot ban points to looming battle over AI rules

February 12 2023, by Joseph BOYLE and Laurence Benhamou



Users have complained that the Replika chatbot was coming on too strong with explicit texts and images.

Users of the Replika "virtual companion" just wanted company. Some of them wanted romantic relationships, sex chat, or even racy pictures of

their chatbot.

But late last year users started to complain that the bot was coming on too strong with explicit texts and images—sexual harassment, some alleged.

Regulators in Italy did not like what they saw and last week barred the firm from gathering data after finding breaches of Europe's massive data protection law, the GDPR.

The company behind Replika has not publicly commented and did not reply to AFP's messages.

The General Data Protection Regulation is the bane of big tech firms, whose repeated rule breaches have landed them with billions of dollars in fines, and the Italian decision suggests it could still be a potent foe for the latest generation of chatbots.

Replika was trained on an in-house version of a GPT-3 model borrowed from OpenAI, the company behind the ChatGPT bot, which uses vast troves of data from the internet in algorithms that then generate unique responses to user queries.

These bots and the so-called generative AI that underpins them promise to revolutionise internet search and much more.

But experts warn that there is plenty for regulators to be worried about, particularly when the bots get so good that it becomes impossible to tell them apart from humans.

'High tension'

Right now, the European Union is the centre for discussions on

regulation of these new bots—its AI Act has been grinding through the corridors of power for many months and could be finalised this year.

But the GDPR already obliges firms to justify the way they handle data, and AI models are very much on the radar of Europe's regulators.

"We have seen that ChatGPT can be used to create very convincing phishing messages," Bertrand Pailhes, who runs a dedicated AI team at France's data regulator Cnil, told AFP.

He said generative AI was not necessarily a huge risk, but Cnil was already looking at potential problems including how AI models used personal data.

"At some point we will see high tension between the GDPR and generative AI models," German lawyer Dennis Hillemann, an expert in the field, told AFP.

The latest chatbots, he said, were completely different to the kind of AI algorithms that suggest videos on TikTok or search terms on Google.

"The AI that was created by Google, for example, already has a specific use case—completing your search," he said.

But with generative AI the user can shape the whole purpose of the bot.

"I can say, for example: act as a lawyer or an educator. Or if I'm clever enough to bypass all the safeguards in ChatGPT, I could say, 'Act as a terrorist and make a plan'," he said.

'Change us deeply'

For Hillemann, this raises hugely complex ethical and legal questions

that will only get more acute as the technology develops.

OpenAI's latest model, GPT-4, is scheduled for release soon and is rumoured to be so good that it will be impossible to distinguish from a human.

Given that these bots still make tremendous factual blunders, often show bias and could even spout libellous statements, some are clamouring for them to be tightly controlled.

Jacob Mchangama, author of "Free Speech: A History From Socrates to Social Media", disagrees.

"Even if bots don't have free speech rights, we must be careful about unfettered access for governments to suppress even synthetic speech," he said.

Mchangama is among those who reckon a softer regime of labelling could be the way forward.

"From a regulatory point of view, the safest option for now would be to establish transparency obligations regarding whether we are engaging with a human individual or an AI application in a certain context," he said.

Hillemann agrees that transparency is vital.

He envisages AI bots in the next few years that will be able to generate hundreds of new Elvis songs, or an endless series of Game of Thrones tailored to an individual's desires.

"If we don't regulate that, we will get into a world where we can differentiate between what has been made by people and what has been

made by AI," he said.

"And that will change us deeply as a society."