BIZCOMMUNITY

Is ChatGPT yet another hurdle for data privacy?

By Ahmore Burger-Smidt

1 Mar 2023

ChatGPT is an OpenAI-developed artificial intelligence (AI) chatbot which has been programmed to have advanced conversational capabilities. This means that the chatbot can answer questions and assist users with composing essays, emails and even writing code in response to certain inputs.

Image source: sebastien decoret – 123RF.com

There has been significant buzz about ChatGPT over the past couple of months and quite remarkably so considering it was only launched on 30 November 2022.

Perhaps of great significance to ChatGPT's meteoric rise has been the fact that it is open for public usage free of charge. This is mainly because the service is still in its research phase, and although that may seem attractive for users, it also means that ChatGPT is collecting a significant amount of data and personal information.

ChatGPT is enabled by its language model. A language model is a probability distribution over a sequence of words. One example of this is predictive text which allows your device to predict the next word or even complete a sentence for you when typing a message or email. This function allows language models to learn from text and gain the capability to produce original text, predict words or sentences, and recognise speech.

In order to function effectively and improve, the language model requires a large volume of data. The more data the AI is fed the better it becomes at detecting patterns, anticipating text and speech, and creating new text and speech, amongst other functions.

OpenAI supplied ChatGPT with over 300 billion words mined from the internet. Sources include articles, blogposts, online reviews and comments. To give an example of the scale of data required for AI models, the entirety of English Wikipedia (about 6 million articles) made up only 0.6% of the data required to "train" ChatGPT. Further, user prompts or inputs to the service along with other background data such as IP address, browser type and settings, and behavioural data, also contribute towards improving ChatGPT.

ChatGPT: Deepfake and copyright concerns
Stefaans Gerber and Alisha Muller 8 Feb 2023

This raises issues around personal information and data protection. In light of the Protection of Personal Information Act 4 of 2013 (PoPIA), some of the main concerns that ChatGPT raises include:

- Further processing - OpenAI uses data, including personal information, to "train" ChatGPT and improve its functionality. This begs the question whether data subjects, especially where their information is not publicly available (i.e. scraped from the internet), have consented to such actions.
- Data retention - ChatGPT records and stores every input or message a user provides to it without an option of erasure/deletion.
- Information quality - ChatGPT has been alleged to spread misinformation by responding to factual questions in misleading or inaccurate ways.
- Data subject access rights - there is currently no process available to data subjects to exercise their participation rights as envisioned in PoPIA.

The advent of ChatGPT and indeed other AI chatbots calls for data subjects to be more prudent with the information they choose to share online. This may include making one's social media account private. However, where providing information onto a public platform is unavoidable (eg. online reviews, comments, etc.) this will call for policymakers and lawmakers to start developing frameworks for AI use that incorporate the right to privacy.

AI is only going to become more consumer facing. Therefore, its impact on data subjects and consumers more generally must be controlled by effective regulation. Without sufficient protection to data subjects who are subject to AI, policymakers, lawmakers and regulators will run the risk of falling further behind Big Tech and AI developers when it comes to data protection.

ABOUT AHMORE BURGER-SMIDT

Director at Werksmans Attorneys. Main Practice areas: • Africa • Competition Focus: • Competition Law and Data privacy • Appearances before Competition Commission, Competition Tribunal and Competition Appeal Court • Obtaining approval for mergers and takeovers from the competition authorities • Former Deputy Commissioner of the South African Competition Commission • Named as a recommended lawyer in Competition by Chambers Global (2017 and 2018). Education: • BCom (University of North West) • LLB (University of North West) • MBL (UNISA) Received Old Mutual medal

- Is ChatGPT yet another hurdle for data privacy? - 1 Mar 2023
- Hands up! Icasa wants your biometrics - 28 Mar 2022
- What does Facebook's data breach mean for personal information and businesses in SA? - 14 Jun 2018

View my profile and articles...