

ChatGPT: Yet Another Hurdle for Data Privacy?

1 March 2023 4 min read

Data Privacy & Cybercrime

ChatGPT is an OpenAI[1] 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.

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[2] has been the fact that it is open for public usage free of charge.[3] 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: magic bullet or the beginning of the end?

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 (e.g. 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.

Footnotes

[1] OpenAI is a research and development company which conducts artificial intelligence research and develops AI enabled services.

[2] ChatGPT reached 100 million active users within the first two months of its release.

[3] ChatGPT recently launched a paid subscription version called ChatGPT Plus on 1 February 2023.



Ahmore Burger-Smidt

Head of Regulatory

[View Profile](#)