



Digital Bulletin – Snapchat AI

AI is being spoken about more and more. The obvious questions are about academic work and intellectual property; however, are there other considerations? Are there any safety concerns involved in having content automatically generated by artificial intelligence?

What is AI?

Artificial intelligence combines computer science and datasets, to enable problem-solving. There are three types of AI: Artificial narrow intelligence (ANI), which has a narrow range of abilities, Artificial general intelligence (AGI), which is comparable with human capabilities, or Artificial superintelligence (ASI), which is more capable than a human. AI systems work by processing huge amounts of data, analysing it for patterns, and using these patterns to make predictions about future states. In this way, a chatbot that is fed examples of text can learn to generate lifelike exchanges with people, or an image recognition tool can learn to identify and describe objects in images by reviewing millions of examples.

Snapchat AI

Snapchat users will have seen a new addition to their chat feature called 'My AI'. This automatically goes to the top of the window (above pinned conversations) and cannot be deleted. The aim is to engage with Snapchat users and help them to engage with various features and functions of the platform. It can also provide suggestions when asked questions on any topic. In researching for this bulletin, I spent some time interacting with the Snapchat AI. I found it to be very much information seeking (asking many questions), which is only to be expected as the quality of the reply is only as good as the information it has. Questions about topics such as music drew responses that referenced the most famous or successful works. There are plans for My AI to be able to have a more realistic conversation with users, rather than simply responding to prompts; however, we are not at that point yet.

What are the risks?

As the AI responses are automatic, rather than a considered reply from a human, there is a degree of trust, in the appropriateness of the content. One user decided to test this by setting up an account as a thirteen-year-old girl. He found that My AI did not take the age of the user into account and also offered advice that could be potentially harmful (screenshots of this can be seen [here](#)).

Users are advised to treat My AI as a stranger amid concerns about privacy. For example, any criminal activity that a user 'discusses' with My AI could be linked online as a result. The flow of information (in both directions) between My AI and other online sources has the potential to put users at risk. Anecdotally, there have been reports of My AI referring to content from a phone's photo library, that had not been mentioned in any AI conversation, posing further questions about where My AI gains information.

In summary...

Snapchat say that safety is important and that it adheres to guidelines to ensure that content is not harmful. Their full response can be seen [here](#). Snapchat also provides the facility to report inappropriate content to reduce the likelihood of repeated instances. The feature, in the main part provides an entertaining and fascinating user experience and in the large part does not a significant threat. However, as with many areas of online activity, there is the potential for inappropriate content to be accessed. Advice, or views that are presented to users are not considered, rational, human responses.

If there are any topics that you would like more information about, or if you have any concerns regarding online safety, please feel free to contact me.