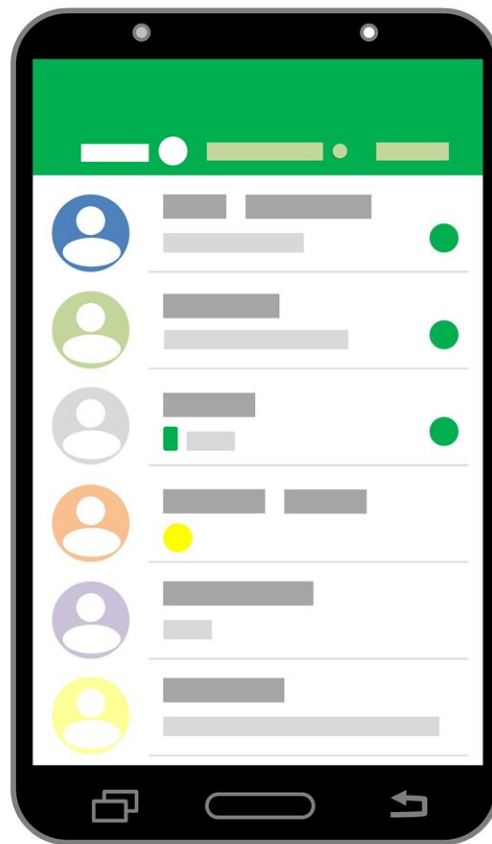# The limitations of AI-generated text

November 23 2021, by Wick Eisenberg

Credit: Pixabay/CC0 Public Domain

Artificial intelligence has reached a point where it can compose text that sounds so human that it dupes most people into thinking it was written by another person. These AI programs—based on what are called

autoregressive models—are being successfully used to create and deliberately spread everything from fake political news to AI-written blog posts that seem authentic to the average person and are published under human-sounding byline.

However, though autoregressive models can successfully fool most humans, their capabilities are always going to be limited, according to research by Chu-Cheng Lin, a Ph.D. candidate in the Whiting School of Engineering's Department of Computer Science.

"Our work reveals that some desired qualities of intelligence—for example, the ability to form consistent arguments without errors—will never emerge with any reasonably sized, reasonably fast autoregressive model," said Lin, a member of the Center for Language and Speech Processing.

Lin's research showed that autoregressive models have a linear thought process that cannot utilize reasoning because they are designed to very quickly predict the next word using previous words. This is an issue because the models are not built to backtrack, edit, or change their work, the way humans do when writing something.

"[Human] professionals in all fields do this. The final product may display spotless work, but it is also likely that the work was not done in a single pass, without editing here and there," Lin said. "But when we train these [AI] models by having them mimic human writing, the models do not observe the multiple rewritings that happened before the final version."

Lin's team also showed that current autoregressive models have another weakness: They do not give the computer enough time to "think" ahead about what it should say after the next word, so there is no guarantee that what it says will not be nonsense.

"Autoregressive models have proven themselves very useful in certain scenarios, but they are not appropriate computational models for reasoning. I also find it interesting that our results suggest certain elements of intelligence do not emerge if all we do is try to get machines to mimic how humans speak," he said.

The result is that the more text that autoregressive models produce, the more obvious their mistakes become, putting the text at risk of being flagged or noticed by another, even less advanced computer programs that require fewer resources to be effective at distinguishing between what was written by an autoregressive models, and what was written by a human.

Because computer programs can decipher what was written by an autoregressive model and what was written by a human, Lin believes that the positives of having AI that can use reasoning far outweigh the negatives, even though a negative could be the spread of misinformation. He says that a process called "text summarization" provides an example of how AI that was capable of using reasoning would be useful.

"These tasks have a computer read a long article, or a table that contains numbers and texts, and then the computer can explain what's going on in a few sentences. For example, summarizing a news article, or a restaurant's ratings on Yelp, using a few sentences," Lin said. "Models that are capable of reasoning can generate texts that are more on the spot, and more factually accurate, too."

Lin has been working on this research, which is part of his thesis, for several years with his adviser, Professor Jason Eisner. He hopes to use these findings to help design a neural network architecture for his thesis research called "Neural Regular Expressions to help AI more effectively understand the meaning of words."

"Among many things, NREs can be used to build a dialog system where machines can deduce unobserved things, such as intent, from conversation with humans, using a rule set predefined by humans. These unobserved things can subsequently be used to shape the machine's response," Lin said.

Provided by Johns Hopkins University