IEEE
CTSoc™
World Newsletter

How to Lie & Confuse with ChatGPT (& Other LLMs) in Consumer Technology

INTRODUCTION

ChatGPT and similar large language models (LLMs) revolutionized human interaction with machines with a significant promise for consumer technology in a number of ways. These promises include help in idea generation by developers and implementers, improved user experience, conversational interfaces, customer support, and training. These improvements might result in more user-friendly environments and in cost savings for businesses and consumers.

Although ChatGPT and similar AI models offer many potential benefits, they come with several potential disadvantages and challenges. For example, lack of understanding, biases, inappropriate content, misinformation and inconsistency. As we are getting into the AIdriven era, understanding the strengths and limitations of such models becomes paramount for both developers and end-users. The AI-generated inaccuracies and hallucinations have ethical implications, and they need to be thoroughly investigated from all imaginable aspects, in particular, because the easiness of every day use of LLM models may result in over-reliance.

Hence, the big question is how do we minimize, eliminate, or prevent the disadvantages and overcome the challenges.

CASE STUDIES – INTERVIEWING CHATGPT

For example, when one of us (NG) once asked ChatGPT: “Who is Nahum Gershon?” he received an answer that he is well known for his work on AI and that he published books and articles in the field of AI. The truth: he has never worked in AI! When he asked the same question again a number of times, he received a plethora of answers like:

“He is best known for his work in the development of software engineering methodologies and tools, particularly in the areas of requirements engineering, software testing, and software process improvement”. Not true!

Or, “Gershon received his Ph.D. in computer science from the Technion – Israel Institute of Technology in 1985. He then worked as a researcher at the IBM T.J. Watson Research Center in New York, where he developed the Requirements Definition and Analysis (RDA) methodology for software requirements engineering”. All fake …

The answer to: “Who is Gordana Velikic from Serbia” was: “I’m sorry, but I don’t have access to specific information about private individuals, especially those who may not be widely known public figures or who have limited online presence”.

On the Internet, however, it is easy to find out that Gordana was, among other things, a Director for Science Programs for RT-RK in Serbia, and General Chair of IEEE ICCEBerlin several times. A search on Google has resulted in no less than 16 of her talks that appear on YouTube. All that information is published before September 2021, a date which is set as a cut-off date for data used in ChatGPT training.

STRATEGIES FOR VALIDATION

The answers to these questions made by ChatGPT illustrate some problems with hallucinations, consistency and insufficiency in finding information.

The conclusion is that information generated by ChatGPT needs to be checked and validated before being used. Some strategies could include:

- Asking the same question from ChatGPT more than once and possibly asking the same question from different LLM systems.
- Doing a search on the internet (e.g., Google) on the question and on the various details described in ChatGPT answers. Validate the individual items presented as facts.
- Asking other people to look for answers to the same questions and discussing the findings together.
- Understanding how the models generate answers, and what the potential hallucination pitfalls are.

FUTURE ENDEAVORS

In some future meetings of the IEEE Consumer Technology Society, we plan to run sessions on “How to Lie and Confuse with ChatGPT (and other LLMs) in Consumer Technology. By suggesting and discussing ways to lie and confuse with ChatGPT (and other LLMs), we expect to learn about the potential pitfalls of ChatGPT and develop ways to avoid them. We are planning to conduct an online discussion group on these matters as well.

AN EXAMPLE FROM A DIFFERENT TECHNOLOGY AREA

In 1992, NG started a series of sessions on How and Confuse with Visualization at the IEEE Visualization Conference with a similar set of goals. He ran them for a good number of years and these sessions are still running at the IEEE Visualization Conference.

By **Nahum Gershon** & **Gordana Velikic**