DM23-1034

Script: Large Language Models for Social and Economic Good: 4 Case Studies
SME(s): *Matthew Walsh and Dominic Ross*
Interviewer/Facilitator: *Thomas Scanlon*
Interview Conducted*: Monday September 18, 2023 at 10:30 a.m. ET*

<Canned Intro>

Suzanne: Welcome to the SEI Podcast Series. My name is Thomas Scanlon, and I am a technical manager in the SEI's CERT Division.

Today I am joined by Matthew Walsh, a senior data scientist on my team, and Dominic Ross, multi-media design team lead. Today we are here to talk about large language models and four case studies that we developed on their potential uses.

Matthew/Dominic: *Thank you.*

1. **Thomas**:
   Let's start by having each of you tell our audience about yourself and what brought each of you to the SEI. Matthew, you are new to our podcasts. Let's have you go first.

   Matthew responds.

Thomas: Dominic, you previously recorded a podcast with us on deepfakes. We will include a link to that in our transcript, but for those members of our audience who didn't hear that podcast, tell us a little bit about yourself, what brought you to the SEI, and the work that you do here.
Dominic responds.

2. **Thomas**: The SEI has several projects and publications underway related to generative AI and large language models. For our audience members who aren't familiar with these topics, can we start with a brief overview of large language models, their inception, and their sudden popularity. These tools have been around, but since the release of ChatGPT late last year this topic has been dominating news headlines.

MMW

- Brief history of language research in psychology
- Brief history of language research in AI
- Little bit about LLMs, and what has made GPT so successful

3. **Thomas**: As we know, with all new technological advancements, there are benefits and there are drawbacks.

We have all read about innovative uses and scary outcomes of tools such as ChatGPT. In a recent blog post and white paper, you explore the capabilities and limitations of LLMs. Can you walk us through the benefits and challenges of using LLMs?

MMW

- Value proposition
    - Higher quality content (depth of knowledge; breadth of knowledge)
    - In less time
- Risks
    - ChatGPT has blind spots;
    - ChatGPT can hallucinate;
    - Not AGI--ChatGPT can't handle multi-part tasks;
    - So impressive that it's easy to overestimate its abilities

4. **Thomas**: One potential drawback that I would like to dig deeper on is the common misconception of how LLMs are trained. Many think that they are trained and released into the wild, and that's it. Can you talk about the continual evolution of tools built on top of LLMs, such as ChatGPT?

MMW

- Get into distinction between foundation models, supervised fine-tuning, and instructional tuning

5. **Thomas:** Your work also detailed four case studies exploring the potential outcomes of LLMs. Can you please walk us through those case studies, how they were developed, and your findings?

MMW.

Briefly touch on 4 use cases

- Code generation
- Education
- Research
- Strategic foresight

In all of these cases, we see ChatGPT working alongside humans rather than instead of them

6. **Thomas**: One aspect that we like to emphasize on our podcasts is transition. For our audience members who want to take what you've learned and apply it to their own work with LLMs, where do they start?

Dominic?

7. **Thomas**: What is next for you both? What else are you both working on that we can bring you back to talk about?

MMW

- Delineating range of tasks that LLMs may be applied to, and developing test harnesses and other means to provide assurances about LLM behavior in those tasks
- Apply LLMs to software security, both in terms of generating code with fewer vulnerabilities; and using LLMs to detect and remediate vulnerabilities already present in code.
- Leveraging LLMs along with other forms of generative AI, like synthetic speech and video, to create training assets

Thomas: Matthew and Dominic, thank you for talking with us today. We will include links in the transcript to resources mentioned during this podcast.

Finally, a reminder to our audience that our podcasts are available on Soundcloud, Stitcher, Apple Podcasts, and Google Podcasts as well as the SEI's YouTube Channel. If you like what you see and hear today, give us a thumbs up.

Thanks again for joining us.

<Canned Outro>