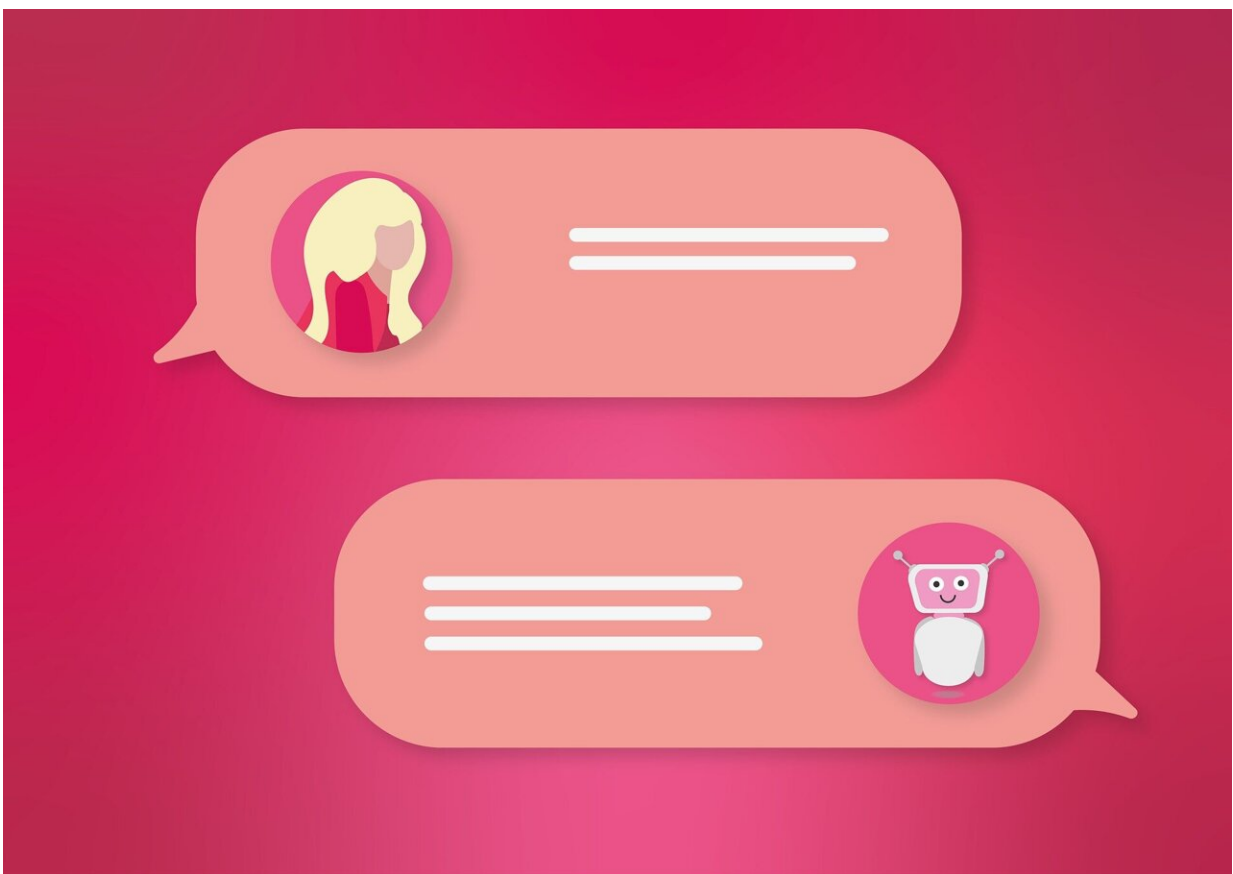


Q&A: Professor discusses ChatGPT-inspired large language model built for the finance industry

June 1 2023, by Lisa Ercolano



Credit: Pixabay/CC0 Public Domain

First there was ChatGPT, an artificial intelligence model with a

seemingly uncanny ability to mimic human language. Now there is the Bloomberg-created BloombergGPT, the first large language model built specifically for the finance industry.

Like ChatGPT and other recently introduced popular language models, this new AI system can write human-quality text, answer questions, and complete a range of tasks, enabling it to support a diverse set of natural language processing tasks unique to the finance industry.

Mark Dredze, an associate professor of computer science at Johns Hopkins University's Whiting School of Engineering and visiting researcher at Bloomberg, was part of the team that created it. Dredze is also the inaugural director of research (Foundations of AI) in the new AI-X Foundry at Johns Hopkins.

The Hub spoke with Dredze about BloombergGPT and its broader implications for AI research at Johns Hopkins.

What were the goals of the BloombergGPT project?

Many people have seen ChatGPT and other large language models, which are impressive new artificial intelligence technologies with tremendous capabilities for processing language and responding to people's requests. The potential for these models to transform society is clear. To date, most models are focused on general-purpose use cases. However, we also need [domain](#)-specific models that understand the complexities and nuances of a particular domain. While ChatGPT is impressive for many uses, we need specialized models for medicine, science, and many other domains. It's not clear what the best strategy is for building these models.

In collaboration with Bloomberg, we explored this question by building an English language model for the financial domain. We took a novel

approach and built a massive data set of financial-related text and combined it with an equally large data set of general-purpose text. The resulting data set was about 700 billion tokens, which is about 30 times the size of all the text in Wikipedia.

We trained a new model on this combined data set and tested it across a range of language tasks on finance documents. We found that BloombergGPT outperforms—by large margins—existing models of a similar size on financial tasks. Surprisingly, the model still performed on par on general-purpose benchmarks, even though we had aimed to build a domain-specific model.

Why does finance need its own language model?

While recent advances in AI models have demonstrated exciting new applications for many domains, the complexity and unique terminology of the financial domain warrant a domain-specific model. It's not unlike other specialized domains, like medicine, which contain vocabulary you don't see in general-purpose text. A [finance](#)-specific model will be able to improve existing financial NLP tasks, such as sentiment analysis, named entity recognition, news classification, and question answering, among others. However, we also expect that domain-specific models will unlock new opportunities.

For example, we envision BloombergGPT transforming natural language queries from financial professionals into valid Bloomberg Query Language, or BQL, an incredibly powerful tool that enables financial professionals to quickly pinpoint and interact with data about different classes of securities. So if the user asks, "Get me the last price and market cap for Apple," the system will return `get(px_last,cur_mkt_cap) for(['AAPL US Equity'])`. This string of code will enable them to import the resulting data quickly and easily into data science and portfolio management tools.

What did you learn while building the new model?

Building these models isn't easy, and there are a tremendous number of details you need to get right to make them work. We learned a lot from reading papers from other research groups who built language models. To contribute back to the community, we wrote a paper with over 70 pages detailing how we built our [data set](#), the choices that went into the model architecture, how we trained the model, and an extensive evaluation of the resulting model. We also released detailed "training chronicles" that contains a narrative description of the model-training process. Our goal is to be as open as possible about how we built the model to support other research groups who may be seeking to build their own models.

What was your role?

This work was a collaboration between Bloomberg's AI Engineering team and the ML Product and Research group in the company's chief technology office, where I am a visiting researcher. This was an intensive effort, during which we regularly discussed data and model decisions, and conducted detailed evaluations of the model. Together we read all the papers we could find on this topic to gain insights from other groups, and we made frequent decisions together.

The experience of watching the model train over weeks is intense, as we examined multiple metrics of the model to best understand if the model training was working. Assembling the extensive evaluation and the paper itself was a massive team effort. I feel privileged to have been part of this fantastic group.

Provided by Johns Hopkins University

Citation: Q&A: Professor discusses ChatGPT-inspired large language model built for the finance industry (2023, June 1) retrieved 23 April 2024 from <https://techxplore.com/news/2023-06-qa-professor-discusses-chatgpt-inspired-large.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.