FORESIGHT

The International Journal of Applied Forecasting

ARTICLES *by* FORECASTERS *for* FORECASTERS: Q4:2023

How Will Generative AI Influence Forecasting Software?

Join the *Foresight* readership by becoming a member of the International Institute of Forecasters
forecasters.org/foresight/

How Will Generative AI Influence Forecasting Software?

MICHELE TROVERO AND SPIROS POTAMITIS

INTRODUCTION

The recent explosion in the popularity of generative AI, sparked by the release of tools like ChatGPT (Open AI, 2023) and DALL-E (Open AI, 2021), has catapulted the subject into the media limelight and made it an inescapable part of any discussion on the future of analytics, including the future of forecasting software. This article serves as a companion piece to the article by Potamitis and colleagues (pages 50-54 in this issue) and delves into what we believe are specific potential applications of AI, particularly large language models (LLMs) within a forecasting support system (FSS), along with identifying obstacles to the widespread adoption of these technologies.

The ChatGPT Frenzy

Hu (2023) cites a UBS study estimating that ChatGPT reached 100 million monthly active users in January 2023, just two months after its launch, making it the fastest-growing consumer application in history.

Articles and opinions on this matter have spread almost as virally as the use of ChatGPT itself – possibly facilitated by the same generative AI tools at the center of so much controversy. This meteoric rise in attention has fueled a lively debate between proponents of the new technology and those who have expressed concerns. The latter group has reservations about what they see as the false promises of AI (Chomsky, 2023) and the possibility of disastrous consequences from creating advanced AI without adequate safeguards in place (Future of Life Institute, 2023). Advocates of this technology maintain that the advantages inherent in AI will ultimately lead to favorable outcomes for humanity (for example, Andreessen, 2023 and Khan, 2023).

In parallel, advancements in technology have been occurring at an unprecedented rate. The major tech companies, AI startups, and researchers have taken notice and have hastily released their own LLMs and generative AI–based chatbots or search engines. Wasting no time, Microsoft seamlessly integrated the technology from ChatGPT, which has since been upgraded to the new GPT-4 version, into its Bing search engine. Google followed suit with Google Bard, based on the PaLM 2 LLM model (Pichai, 2023), and Meta AI released LLaMA (Meta AI, 2023), an open-source LLM. Hardly a week goes by without the release of a new LLM that promises to be better or more efficient. By the time this article is published, it is likely that the panorama of generative AI and LLMs will have already changed dramatically.

On the wave of this debate, political entities are swiftly moving to introduce legislation to regulate AI, such as the AI Act proposed by the European Commission (2021).

APPLICATIONS OF LLMS IN FORECASTING SUPPORT SYSTEMS

While others argue the philosophical and regulatory concerns that have sprung up, we wish to focus solely on the aspects of these new technologies that may influence the development of forecasting support systems.

There are a number of areas where we expect LLMs to be applied in ways that benefit forecasting software. Here are a few of them.

Code Generation and Increasing Automation

AI methods are already used in forecasting software to automate tasks or aid manual operations. For example, Valsaraj

This article originally appeared in *Foresight*, Issue 71 (*forecasters.org/foresight*) and is made available with permission of *Foresight* and the International Institute of Forecasters.

Key Points

- The recent explosion in the popularity of generative AI, with the release of tools like ChatGPT, has made the use of large language models (LLMs) an inescapable topic in the discussion of the future of forecasting software.
- Forecasting software can benefit from LLMs in enhanced code generation, improved user interfaces, intelligent suggestions, better interpretability, and increased automation.
- The lack of abundant, available, and suitable data for model training can limit the wider applicability of AI models in forecasting support systems (FSSs).
- Possibly driven by regulations, the application of AI in FSSs will require heightened attention to the ethical use of models and data.

and colleagues (2018) and SAS Institute Inc. (2018) used an AI model to guide judgmental overrides of statistical forecasts within a demand-planning application. LLMs have shown great promise in the field of code generation. Applications such as GitHub Copilot, launched in mid-2022 (Dohmke, 2022), are becoming commonplace in helping developers write code quickly and efficiently. We believe this to be the most promising aspect of the current generation of LLMs for software applications in the short term. In the context of forecasting software, LLMs can be applied to automate the generation of code snippets or scripts that facilitate various aspects of the forecasting process. For example, LLMs can help generate code for data preprocessing, feature engineering, model selection, and evaluation. By leveraging the contextual understanding and language-generation capabilities of LLMs, forecasting software can provide users with ready-to-use code templates or even generate complete code solutions tailored to their specific forecasting needs. Automation is a key goal of forecasting software, and LLM-generated flows can aid in increasing the level of automation. LLMs might help to automate repetitive or time-consuming tasks in the forecasting process, such as data cleaning, outlier detection, and model evaluation. They might also help to automate the selection of appropriate forecasting models based on the data characteristics and user requirements. By automating these tasks, LLMs reduce the manual effort required by users, streamline the forecasting workflow, and enable faster and more efficient forecasting processes.

Friendly and Interactive Natural Language Processing Interfaces

Natural language processing (NLP) interfaces play a crucial role in enhancing the user's ease of interaction with forecasting systems and enable greater access to people of all abilities, such as by allowing the use of text-to-speech for a person with limited vision. Current-generation analytical software, such as SAS® Model Studio (SAS Institute Inc., 2023), can already be driven by a rule-based chatbot (SAS Institute Inc., 2020). Rule-based chatbots provide assurances in terms of the expected responses, but they come with limitations because of the need to create an infrastructure specific to the application at hand. One attractive feature of LLMs is that they can be employed to easily develop generic, user-friendly NLP interfaces that let users communicate with the software through natural language queries or commands. The most popular example is, of course, ChatGPT, which combines a powerful LLM with chatbot capabilities. These interfaces can understand user inputs, extract relevant information in the form of tokens that can be passed to an API for processing, and provide meaningful responses. LLMs can enable more conversational and intuitive interactions within an FSS, making it easier for users to communicate their forecasting requirements, obtain insights, and receive feedback or suggestions. These suggestions can be related to various aspects of the forecasting process, such as selecting appropriate models, tuning hyperparameters, handling missing data, or dealing with outliers. LLMs can analyze the context and user inputs, understand the goals or challenges in forecasting, and generate suggestions or recommendations based on best practices or previous successes. By offering suggestions, LLMs can assist users in making informed decisions and potentially improve the accuracy and efficiency of their forecasting tasks. While this is an attractive scenario for improving FSSs, it requires two conditions in order to operate effectively:

1. *Availability of resources to train or fine-tune the LLMs in the context of the FSS*. Training a current-generation LLM requires a large volume of relevant textual data and, even by conservative estimates, a very large budget (Li, 2020). To begin with, acquiring substantial amounts of documentation is indispensable for satisfying the bottomless appetite of your data-driven model. For an FSS, gathering the necessary volume of examples and possible user interactions remains a formidable challenge. (See the later paragraph on pretrained models for more details.) Additionally, assuming that a sufficient corpus of documents can be amassed, the next step entails allocating considerable financial resources to continually fine-tune and maintain the model (Charush, 2023). The cost of this process would inevitably weigh on the overall cost of maintaining the FSS. On the positive side, the quality of the NLP interaction should improve as the model is progressively fine-tuned.
2. *Effective ways to control LLMs' "hallucinations" and biases*. We all have read stories of people receiving bizarre responses from generative AI-based chatbots, such as ChatGPT, or responses containing errors presented with unwavering certainty. Developers refer to these responses as "hallucinations." To guarantee that AI operates effectively, there should be rules in place to ensure appropriate responses to situations beyond its capacity or deemed inappropriate for the context of the FSS. This is still an active area of research that has not found a satisfactory answer. Similarly, responses that are derived from biased data sources can lead to dangerous outcomes for some populations. Because of the intricate and opaque nature of LLMs, meticulous care must be taken in using data and generating suggestions to uphold ethical standards. This entails thoroughly examining data pools for bias, comprehending the origins and authenticity of data sources, verifying data accuracy, and safeguarding individual privacy. Evaluating the application for unforeseen consumer interactions has to become an integral component of FSS development. The number of scenarios that must be tested and controlled for in an FSS can grow exponentially.

Improving Interpretability and Trust in the Forecasts

Trust in the results is a critical aspect of FSSs, particularly when black-box techniques are involved. Intelligibility, or interpretability, is a key component that FSSs need in order to build trustworthy forecasts (Spavound and Kourentzes, 2022). In conjunction with model interpretability methodologies, LLMs could be used to generate human-readable explanations or summaries of the underlying forecasting models, highlighting the key factors or features driving the predictions. LLMs can also assist in visualizing the model's attention weights, which indicate the relative importance of various elements in the data. This increased approachability can help users understand how the forecasting models work, build trust in the results, and gain insights into the factors influencing the forecasts. On the other hand, if not mitigated, the non-deterministic nature of LLM output can itself generate confusion and become an obstacle to perceiving the FSSs as trustworthy (Engelbrecht, 2023).

AI MODELS IN FORECASTING

Most of the current generation of LLMs are based on a variation of the transformer neural network architecture introduced by Vaswani and colleagues (2017).

One question is whether the same typology of models that are behind LLMs can be employed for forecasting. Some versions of recurrent neural networks (RNNs), such as long short-term memory models (LSTMs), have been proven to be valuable for forecasting. This debate is still wide open and ripe for research. While vanilla transformer models appear to be outperformed by simpler linear models (Zeng and colleagues, 2022) for a variety of data sets, there are at least some successful implementations of transformer models in the industry, such as at Zalando SE (Kunz and colleagues, 2023). Research is continuing to make the transformer architecture more effective for time series forecasting (see Nie and colleagues, 2022, Serravalle Reis Rodrigues and colleagues, 2023, and references therein).

In Potamitis and colleagues we covered the influence of new models on FSSs, including AI-based methodologies, and stressed the importance for FSSs to be open and extensible to accommodate the fast pace of model evolution. Modern forecasting software already lets users extend the built-in algorithms with custom algorithms using heterogeneous languages. Should transformer models prove to be successful at forecasting, an open and extensible architecture will allow for fast adoption and experimentation, at least from the technological point of view. A different question is whether the complexity and cost of training such models will make them more effective than simpler, more economical models (Petropoulos and colleagues, 2022). It is an important consideration for users of an FSS that further validates the need for the FSS to be open and allow experimentation with a wide variety of methodologies to find the one that best balances the unique needs and goals of each user.

PRETRAINED MODELS AND DATA AVAILABILITY – OR LACK THEREOF

LLM models such as the GPT models behind ChatGPT are pretrained. Sites like HuggingFace offer thousands of pretrained machine learning models and data sets. Yet at the time this article was written, only five of them are listed as related to forecasting. You might wonder whether pretrained models will be available for forecasting in the near future.

The abundance of text and image data available for training large language models with generative AI has made the success of such models possible. Text and image data are abundant and publicly accessible, cover diverse topics and sources, and are readily available from sources such as books, articles, and the internet. This enables the training of models with billions of parameters that exhibit remarkable language understanding and generation capabilities. Even with an abundance of available data, the proper selection and labeling of a training set is a very costly effort (Charush, 2023). Apart from cost considerations, we believe that the lack of publicly available, comprehensive, and sufficiently varied data represents the biggest challenge to providing effective pretrained models that are general enough to be incorporated in forecasting software. Here are some of the challenges specific to time series data:

- *Limited publicly available and diverse time series data*. Publicly available time series data sets that cover a wide range of domains and include relevant features can be scarce. While some well-known repositories such as Kaggle and platforms like Google Dataset Search do offer time series data, they often have limitations in terms of data volume, diversity, and domain specificity.
- *Data privacy and ownership*. Time series data often contain sensitive information, such as financial data, customer records, or proprietary business data. Companies and organizations may be reluctant to share their time series data publicly out of concerns about privacy or competition.
- *Customized and domain-specific context*. Forecasting tasks are highly domain-specific, and the success of a forecasting model depends on its ability to capture the nuances and patterns unique to a particular domain. Generic pretrained forecasting models might not be suitable for all domains because of the

need for specialized domain knowledge and contextual understanding. Lack of access to diverse and domain-specific time series data further exacerbates the challenge of providing pretrained models that are universally applicable across various forecasting scenarios.

Addressing the lack of a comprehensive public time series data set will require efforts in data collection, sharing, and collaboration between organizations, research institutions, software vendors, and the public sector. Initiatives such as the Monash Time Series Forecasting Archive (Godahewa and colleagues, 2021) that promote the creation of standardized and openly accessible time series data sets can help bridge the gap and foster the development of pretrained forecasting models suitable for a wide range of domains. We believe that such efforts should be a priority for the advancement of forecasting research and practice.

Addressing the lack of a comprehensive public time series data set will require efforts in data collection, sharing, and collaboration between organizations, research institutions, software vendors, and the public sector.

Additionally, the nondeterministic nature of AI models can itself be an impediment to the adoption of FSSs based solely on pretrained models. The lack of stability mentioned by Spavound and Kourentzes, and difficulty in validating and replicating the results, can undermine trust in the results of the FSS and can lead to confusion and frustration both for the business that provides the FSSs and for its customers.

In summary, while we are eager to be proven wrong, we are pessimistic that the use of pretrained models will become commonplace in FSSs, at least in the short-to-medium term.

CONCLUSIONS

The leaps that generative AI models, and LLMs in particular, have made in recent years, coupled with the release of the easy chatbot interface provided by ChatGPT in late 2022, have propelled the discussion of generative AI from the exclusive circles of specialized researchers into the general public.

Alongside the debate, a shift is happening in the dynamic between technology teams and business units regarding AI. Rather than business leaders having to be convinced of the benefits of using AI, there is now significant demand from within the enterprise itself. CIOs are pressuring their companies to urgently develop AI strategies (*MIT Technology Review* Insights, 2023). As a result, technology teams are being asked to find novel areas of AI application. We expect this demand to trickle down to forecasting software as well.

In this article we discussed how, by applying LLMs, forecasting software can benefit from enhanced code generation, improved user interfaces, intelligent suggestions, better interpretability, and increased automation. These enhancements harness the capabilities of LLMs in comprehending natural language, generating code and text, and enabling context-aware decisions. They can contribute to the development of forecasting systems that are more streamlined, user-friendly, and precise. We also outlined our belief that the lack of abundant, available, and suitable data for model training can limit the wider applicability of AI models in FSSs.

Coupled with the exciting opportunities they bring, we believe AI models will also present new challenges and the need to pay closer attention to the ethical use of AI and data, which is only in its infancy in FSSs. Possibly driven by regulatory changes as well, FSSs will have to address data privacy, bias accountability, transparency, model governance, and many other aspects at the forefront of the discussion on the future of AI.

Overall, we want to reiterate the conclusions of Potamitis and colleagues. These are exciting times indeed for those of us who develop forecasting software. The demand for analytical software has been steadily increasing, driven both by advances in technology and by severely disruptive events such as the COVID-19 pandemic. Furthermore, the long-standing divide between the two cultures of statistics described by Leo Breiman (2001) over two decades ago appears to finally be closing. The availability and sophistication of forecasting models are increasing at a pace never seen before. The prospects for further developments in this domain seem almost boundless. With the rapid progress and convergence of these factors, the future of forecasting software looks extremely promising.

REFERENCES

Andreessen, M. (2023). Why AI Will Save the World. **https://a16z.com/2023/06/06/ai-will-save-the-world/**

Anonymous. (n.d.). HuggingFace. **https://huggingface.co/docs/hub/index**

Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author), *Statistical Science*, 16, 199-231. **https://www.jstor.org/stable/2676681**

Charush (2023). Large Language Models, Small Budget: How Businesses Can Make It Work. **https://blog.accubits.com/large-language-models-small-budget-how-businesses-can-make-it-work/** [Accessed 11 July 2023].

Chomsky, N. (2023). The False Promise of ChatGPT. **https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html**

Dohmke, T. (2022). GitHub Copilot Is Generally Available to All Developers. **https://github.blog/2022-06-21-github-copilot-is-generally-available-to-all-developers/**[Accessed 11 July 2023].

Engelbrecht, S. (2023). Discover the Challenges of Non-deterministic Language Models and Strategies to Get More Consistent Results. **https://www.sitation.com/non-determinism-in-ai-llm-output/**

European Commission (2021). AI Act. **https://artificialintelligenceact.eu/the-act/**

Future of Life Institute (2023). Pause Giant AI Experiments: An Open Letter. **https://futureoflife.org/open-letter/pause-giant-ai-experiments/**

Godahewa, R., et al. (2021). Monash Time Series Forecasting Archive, *Neural Information Processing Systems Track on Datasets and Benchmarks*, 33. **https://forecastingdata.org/** [Accessed 2023].

Hu, K. (2023). ChatGPT Sets Record for Fastest-Growing User Base – Analyst Note. **https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/**

Khan, J. (2023). Meta's Chief A.I. Scientific Calls A.I. Doomers Preposterous and Predicts LLMs Are Just a Passing Fad. **https://fortune.com/2023/06/14/metas-chief-a-i-scientist-calls-a-i-doomers-preposterous-and-predicts-llms-are-just-a-passing-fad/**

Kunz, M., et al. (2023). Deep Learning Based Forecasting: A Case Study from the Online Fashion Industry. **https://arxiv.org/abs/2305.14406** [Accessed 11 July 2023].

Li, C. (2020). OpenAI's GPT-3 Language Model: A Technical Overview. **https://lambdalabs.com/blog/demystifying-gpt-3** [Accessed 11 July 2023].

Meta AI. (2023). Introducing LLaMA: A Foundational, 65-Billion-Parameter Large Language Model. **https://ai.facebook.com/blog/large-language-model-llama-meta-ai/**

MIT Technology Review Insights (2023). The Great Acceleration: CIO Perspectives on Generative AI, *MIT Technology Review*, 18 July. **https://www.technologyreview.com/2023/07/18/1076423/the-great-acceleration-cio-perspectives-on-generative-ai/**

Nie, Y., Nguyen, N.H., Sinthong, P. & Kalagnanam, J. (2022). A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. **https://arxiv.org/abs/2211.14730**

Open AI. (2021). DALL·E 2. **https://openai.com/dall-e-2**

Open AI. (2023). ChatGPT. **https://openai.com/blog/chatgpt**

Petropoulos, F., Grushka-Cockayne, Y., Siemsen, E. & Spiliotis, E. (2022). Wielding Occam's Razor: Fast and Frugal Retail Forecasting. **https://arxiv.org/abs/2102.13209**

Pichai, S. (2023). An Important Next Step on Our AI Journey. **https://blog.google/technology/ai/bard-google-ai-search-updates/**

Potamitis, S., Trovero, M. & Katz, J.A. (2023). A Glimpse into the Future of Forecasting Software, *Foresight*, Issue 71, 50-54.

SAS Institute Inc. (2018). Assisted Demand Planning Using Machine Learning for CPG and Retail. **https://www.sas.com/en/whitepapers/assisted-demand-planning-109971.html** [Accessed 11 June 2023].

SAS Institute Inc. (2020). Giving Your Model a Voice: SAS Conversation Designer and SAS Visual Data Mining and Machine Learning. Video. **https://www.youtube.com/watch?v=zH66556VLUU** [Accessed 11 June 2023].

SAS Institute Inc. (2023). SAS Conversation Designer. **https://www.sas.com/en_us/software/conversation-designer.html** [Accessed 11 June 2023].

SAS Institute Inc. (2023). SAS Model Studio. **https://support.sas.com/en/software/model-studio-support.html** [Accessed 11 June 2023].

Serravalle Reis Rodrigues, V. H., de Melo Barros Junior, P. R., dos Santos Marinho, E. B. & de Jesus Silva, J. L. L. (2023). Wavelet Gated Multiformer for Groundwater Time Series Forecasting, *Scientific Reports*, August, Issue 13. **https://www.nature.com/articles/s41598-023-39688-0**

Spavound, S. & Kourentzes, N. (2022). Making Forecasts More Trustworthy, *Foresight*, Issue 66, 21-25.

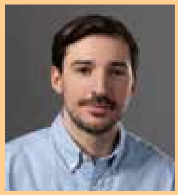
Valsaraj, V., et al. (2018). US Patent No. US10255085B1.

Vaswani, A., et al. (2017). Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA. **https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf**

Zeng, A., Chen, M., Zhang, L. & Xu, Q. (2022). Are Transformers Effective for Time Series Forecasting? Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23). **https://ojs.aaai.org/index.php/AAAI/article/view/26317/26089**

Michele Trovero leads the Forecasting R&D group at SAS. He has over 20 years of experience in the development of commercial software for analytics. His interests cover a wide variety of topics related to forecasting, time series analysis, and machine learning. He has presented at several international conferences on these topics. He received a PhD in statistics from the University of North Carolina at Chapel Hill and has a degree in economics from Bocconi University in Milan, Italy. When he's not thinking about forecasting or driving his children to activities, you can find Michele riding his bikes or running.

michele.trovero@sas.com



Spiros Potamitis is a Data Scientist and Product Marketing Manager at SAS, specializing in forecasting and machine learning. He works with R&D and product management to define SAS products' direction and vision while helping SAS customers apply advanced analytics and AI to drive business value. Spiros holds an MSc in computer engineering and an MSc in information management from the University of Manchester.

spiros.potamitis@sas.com