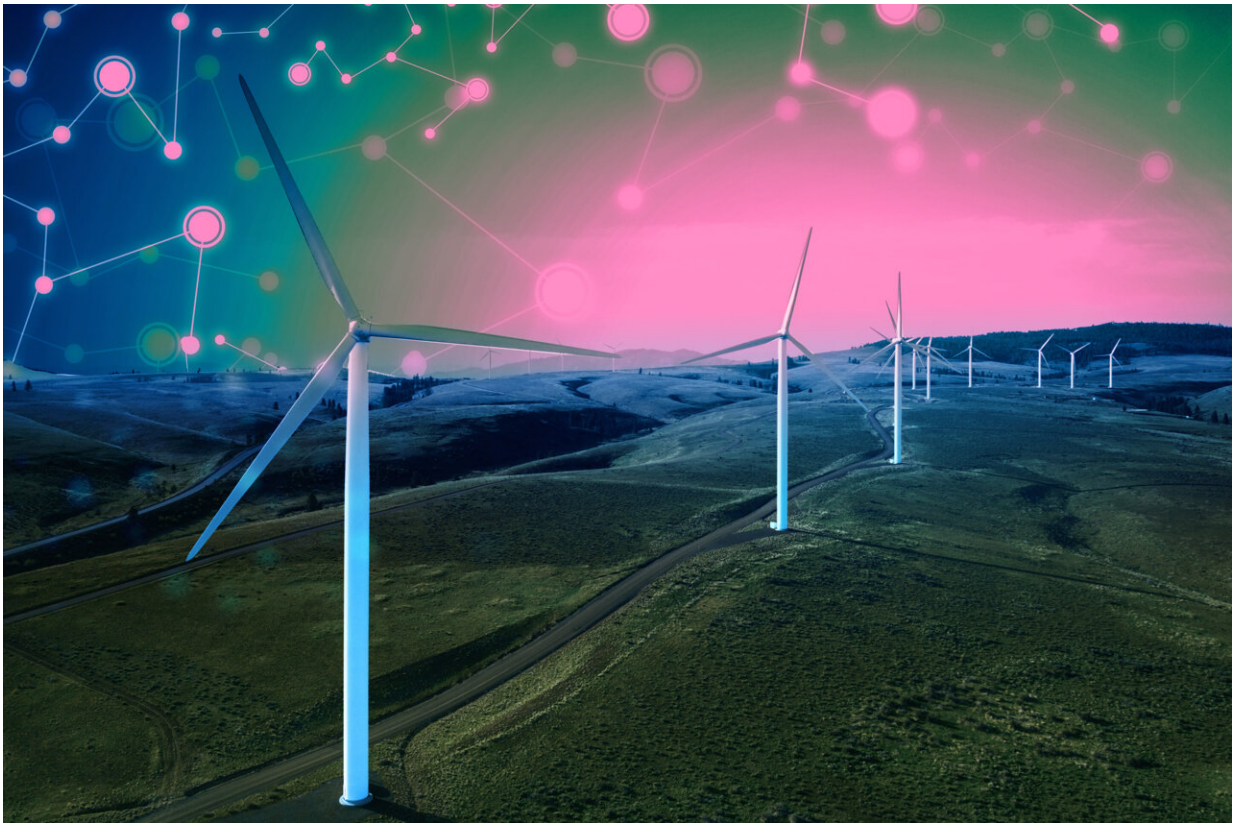


Researchers use large language models to flag problems in complex systems

August 13 2024, by Adam Zewe



The new method could someday help alert technicians to potential problems in equipment like wind turbines or satellites. Credit: MIT News

Identifying one faulty turbine in a wind farm, which can involve looking at hundreds of signals and millions of data points, is akin to finding a

needle in a haystack.

Engineers often streamline this complex problem using deep-learning models that can detect anomalies in measurements taken repeatedly over time by each turbine, known as time-series data.

But with hundreds of [wind turbines](#) recording dozens of signals each hour, training a deep-learning model to analyze time-series data is costly and cumbersome. This is compounded by the fact that the model may need to be retrained after deployment, and wind farm operators may lack the necessary machine-learning expertise.

In a new study, MIT researchers found that [large language models](#) (LLMs) hold the potential to be more efficient [anomaly](#) detectors for time-series data. Importantly, these pretrained models can be deployed right out of the box.

The researchers developed a framework, called SigLLM, which includes a component that converts time-series data into text-based inputs that an LLM can process. A user can feed these prepared data to the model and ask it to start identifying anomalies. The LLM can also be used to forecast future time-series [data points](#) as part of an anomaly detection pipeline.

While LLMs could not beat state-of-the-art deep learning models at anomaly detection, they did perform as well as some other AI approaches. If researchers can improve the performance of LLMs, this framework could help technicians flag potential problems in equipment like heavy machinery or satellites before they occur, without the need to train an expensive deep-learning model.

"Since this is just the first iteration, we didn't expect to get there from the first go, but these results show that there's an opportunity here to

leverage LLMs for complex anomaly detection tasks," says Sarah Alnegheimish, an [electrical engineering](#) and computer science (EECS) graduate student and lead author of a paper on SigLLM.

Her co-authors include Linh Nguyen, an EECS graduate student; Laure Berti-Equille, a research director at the French National Research Institute for Sustainable Development; and senior author Kalyan Veeramachaneni, a principal research scientist in the Laboratory for Information and Decision Systems. The paper is [published](#) on the *arXiv* preprint server, and the research will be presented at the [IEEE Conference on Data Science and Advanced Analytics](#).

An off-the-shelf solution

Large language models are auto regressive, which means they can understand that the newest values in sequential data depend on previous values. For instance, models like GPT-4 can predict the next word in a sentence using the words that precede it.

Since time-series data are sequential, the researchers thought the auto-regressive nature of LLMs might make them well-suited for detecting anomalies in this type of data.

However, they wanted to develop a technique that avoids fine-tuning, a process in which engineers retrain a general-purpose LLM on a small amount of task-specific data to make it an expert at one task. Instead, the researchers deploy an LLM off the shelf, with no additional training steps.

But before they could deploy it, they had to convert time-series data into text-based inputs the language model could handle.

They accomplished this through a sequence of transformations that

capture the most important parts of the time series while representing data with the fewest number of tokens. Tokens are the basic inputs for an LLM, and more tokens require more computation.

"If you don't handle these steps very carefully, you might end up chopping off some part of your data that does matter, losing that information," Alnegheimish says.

Once they had figured out how to transform time-series data, the researchers developed two anomaly detection approaches.

Approaches for anomaly detection

For the first, which they call Prompter, they feed the prepared data into the model and prompt it to locate anomalous values.

"We had to iterate a number of times to figure out the right prompts for one specific time series. It is not easy to understand how these LLMs ingest and process the data," Alnegheimish adds.

For the second approach, called Detector, they use the LLM as a forecaster to predict the next value from a time series. The researchers compared the predicted value to the actual value. A large discrepancy suggests that the real value is likely an anomaly.

With Detector, the LLM would be part of an anomaly detection pipeline, while Prompter would complete the task on its own. In practice, Detector performed better than Prompter, which generated many false positives.

"I think, with the Prompter approach, we were asking the LLM to jump through too many hoops. We were giving it a harder problem to solve," says Veeramachaneni.

When they compared both approaches to current techniques, Detector outperformed transformer-based AI models on seven of the 11 datasets they evaluated, even though the LLM required no training or fine-tuning.

In the future, an LLM may also be able to provide plain language explanations with its predictions, so an operator could be better able to understand why an LLM identifies a certain data point as anomalous.

However, state-of-the-art deep learning models outperformed LLMs by a wide margin, showing that there is still work to do before an LLM could be used for anomaly detection.

"What will it take to get to the point where it is doing as well as these state-of-the-art models? That is the million-dollar question staring at us right now. An LLM-based anomaly detector needs to be a game-changer for us to justify this sort of effort," Veeramachaneni says.

Moving forward, the researchers want to see if finetuning can improve performance, though that would require additional time, cost, and expertise for training.

Their LLM approaches also take between 30 minutes and two hours to produce results, so increasing the speed is a key area of future work. The researchers also want to probe LLMs to understand how they perform anomaly detection, in the hopes of finding a way to boost their performance.

"When it comes to complex tasks like anomaly detection in time series, LLMs really are a contender. Maybe other complex tasks can be addressed with LLMs, as well?" says Alnegheimish.

More information: Sarah Alnegheimish et al, Large language models can be zero-shot anomaly detectors for time series?, *arXiv* (2024). [DOI:](#)

[10.48550/arxiv.2405.14755](https://arxiv.org/abs/2405.14755)

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Researchers use large language models to flag problems in complex systems (2024, August 13) retrieved 13 August 2024 from <https://techxplore.com/news/2024-08-large-language-flag-problems-complex.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.