Spotlight on BlueDot

There is a nuance to LLM usage that doesn't tend to make it into the marketing material. Everyone is excited, and everyone is using it everywhere. Right?

This all might be a bit 2023, but two years ago a company called BlueDot held a webinar where they invited their community to discuss whether LLMs were ready for prime time in predicting the spread of infectious diseases. And that's the latest-most-detailed messaging I can find. Even the Open WebUI instance living on the server in my basement can't find anything newer. In any case, they were pro LLM.

The first signs of the next pandemic tend to be hidden away in small local newspapers halfway across the world, in a language you may not speak, and if you're an epidemiologist or a healthcare worker, you probably want to know about it. Even after that first signal is detected, how can you be sure that it'll spread to your community? It'd be great to know what medical supplies you or your organization need to stockpile. Maybe you could figure it out if you had some air travel data. This is where BlueDot comes in. The idea is that the faster you know an outbreak is happening, the faster you can mitigate its effects. And it seems to work. Well, the detecting bit. The mitigating bit is largely in the hands of their users. They have accurately predicted the spread of Ebola in 2014, a Zika outbreak in Florida in 2016, and alerted their customers to COVID-19 five days ahead of the World Health Organization.

Ok, so pretty impressive, but they accomplished all of this before LLMs burst onto the scene. So what are they doing now? When being wrong can have catastrophic consequences, how do you integrate the biggest technological trend the world has ever seen?

The answer is carefully. They have trust issues, which is pretty understandable, so they're not using off the shelf models trained on the open web. The devil's in the details here, they may still be using off the shelf *models*, that part wasn't explicitly clear, but with an educated guess and some deductive reasoning, we can assume they don't quite have the resources to build their own models from scratch. They are, however, definitely doing their own training, on their own curated datasets. Meticulously curated datasets.

It's a great reminder that your model is only as good as your training data, and LLMs haven't changed this tenet of AI/ML. BlueDot works with well trained subject matter experts from various fields including surveillance analysts,

epidemiologists, clinicians, and even veterinarians; research publications are a must if you want to be one of these experts. They make calls based on data availability, frequency, and quality, and would rather not generate any insights at all than generate one based on bad data.

BlueDot is also placing particular emphasis on collating information, as opposed to generating it. LLMs are used to understand what their users are asking, but the answer supplied comes from their vast amounts of data, all accessible to the LLM through handy APIs. The worst they can do is misunderstand your question, and give you a pretty chart you didn't ask for. Graphs, charts, maps, and tables are actually a huge win for their users. Of course, they're generated using BlueDot's impeccable data, and are completely customizable by the user. Why generate a text response when it is so much easier to convince your boss to start the pandemic protocol with a pretty graph? Everyone loves visuals.

So the next time you can't afford to be wrong, just remember to focus on the quality of your data, collaborate with some experts to make it bullet-proof, then use it to generate some awesome visuals.