

Generative AI for Regulatory Intelligence

Introduction

The world has been abuzz since the release of ChatGPT by OpenAI in 2023. ChatGPT was among the first widely touted tools of a set called Generative AI, or Gen AI for short. Gen AI has become popular jargon for businesses to claim their products are superior to the competition, and for job-seekers to claim a state-of-the-art skillset.

One basic way of using Gen AI tools such as ChatGPT is to find detailed answers to questions and then pursue the sources cited by the search results. This is more sophisticated than traditional keyword searches provided by Google, Bing, and others. In addition to the search results, LLMs can include relevant information, supporting references, and even anecdotes. The downside to this approach is that models usually lag by a few days. So if a draft of a guidance document has since changed, the professional may find answers to questions from a previous draft and may be misled. A second challenge is that documents published by the FDA, CMS, and other regulatory agencies, follow certain formatting guidelines. Extracting information from these is not straightforward and requires custom training of the models. ChatGPT and Gemini do not treat these documents as special documents so answers extracted may not yield helpful responses. Nevertheless, using Gen AI tools for search can be an extremely valuable and productive exercise.

Speaking to several regulatory professionals, I see two myths that I would like to dispel here. First, publicly available solutions such as ChatGPT can solve ALL challenges. Second, that AI will make regulatory professionals obsolete very soon. I will dispel both these myths by the end of this article.

What is Gen AI?

While Gen AI is a straightforward concept, the underlying technology has reached its current state following decades of research and development. Innovation in this field is advancing at a blistering pace. What we are seeing today, with tools such as ChatGPT and Gemini (by Google), is only scratching the surface of the possibilities ahead.

So what exactly is Gen AI? It is software that takes in text as its input and generates other text as its output. Let us look at a simple example. I entered the word “*Hello*” into Gemini. The output was “*Hello! How can I help you today?*”. How did Gemini generate this response and not some other text?

To understand this better, we must explore what goes into building programs such as Gemini, ChatGPT, or any other Gen AI tool. The GPT in ChatGPT stands for Generalized Pretrained Transformer. A transformer is a stack of neural networks, which in turn are a stack of mathematical functions. In short, a transformer is a collection of millions of mathematical functions working together to process information. When transformers are used solely for language, they are called Large Language Models or LLMs. However, they may be used for image generation, like in Open AI's DALL-E, video creation, or sound generation.

Pretrained means that the transformer, the LLM, is fed with millions of documents, from sources such as Wikipedia, newspaper articles, research papers, books, census data, and pretty much everything that exists in the public domain. The result is that the transformer creates a sophisticated probabilistic model that will predict the next word given an input word based on the documents that it has processed during pretraining.

There are thousands of LLMs as well as transformers available today, though only a handful have gained broad popularity. For instance, ChatGPT offers GPT 3.5 and GPT 4.0, two different models with different capabilities.

Making Inferences

Models are today trained for a wide range of tasks, often running into 100s of activities that humans typically perform. These include inference, translation from one language to another, answering questions, understanding images, comprehending charts, performing mathematical operations, and much more. LLMs would be good at certain types of inference. For example, I typed the following into Gemini:

“What conclusions can we draw from the Inflation Reduction Act”.

The output was remarkably lucid, included references, and even stated that the act was significant with some major environmental and economic impact. In the following example, the results were disappointing:

“What are the implications of recently proposed rules by the FDA for medical devices”.

The response was very generic, and there were no specifics included.

Improving Outputs with Prompt Engineering

A recent technique that gained enormous popularity is Prompt Engineering. Prompts are text paragraphs that set a context and scenario for the LLM to respond to. Here is an example of a prompt:

“Assume I am a regulatory affairs professional working in the pharmaceutical industry. What should I know about the Inflation Reduction Act.”

This prompt was entered into ChatGPT 3.5 and into Gemini. Both produced responses about a page long highlighting key provisions of the law, which I found immensely educational and useful. Here is another prompt:

“Assume I am a regulatory affairs professional working in the pharmaceutical industry. Explain the Inflation Reduction Act in 200 words or less.”

Both models produced outputs but in different formats.

The key pitfalls with prompts are that they may require some trial and error and may result in different responses from different models, or even when the prompt is slightly changed for the same model. For instance, I changed the context for the prompt from pharmaceutical industry to transportation, and again to energy. The output produced by ChatGPT 3.5 was identical in all three instances. Gemini produced more precise outputs, with specific provisions that impacted each sector. Our goal here is not to compare models but to illustrate that outputs can vary even though the inputs are identical. Designing the correct prompts can be time-consuming with no guarantee of helpful results. It is indeed a technique one must use with caution.

Regulatory Writing

Some professionals are exploring whether they could use Gen AI for regulatory and medical writing towards drug and device approvals. To understand whether this is feasible, we must note that LLMs such as ChatGPT and Gemini, are trained largely on public data. It is estimated that only 10% of data worldwide is public and 90% is private.

For instance, an LLM such as ChatGPT, during training, will never encounter any drug or device research that is kept private. So no public model can correctly produce outputs (writing in this case) based on private information. It is possible to train models on private data for such writing by an enterprise that purchases an OpenAI or Gemini enterprise license. The cost of such training is enormous, and without adequate input examples, the outputs would likely be far below expectations and underperform human writing by a huge margin. A common condition users experience when models are asked to generate text from sources never encountered during training, is a hallucination. Hallucinations result when the model begins to make up answers using its completion and probabilistic models, especially when its knowledge base is limited.

In my opinion, Gen AI will not be able to accurately perform medical writing tasks in the foreseeable future.

Reinforcement Learning and Human Feedback (RLHF)

Another popular technique that is gaining in popularity is the RLHF. This technique uses human experts to perform verifications of the outputs during the training phases so the models become much “smarter”. RLHF is an expensive and time-consuming process with no guaranteed results due to the inherently probabilistic nature of Gen AI. Certain tasks such as identifying animal images, may lend themselves well to RLHF, while tasks such as inferring from new legislation may not.

Conclusion

Gen AI can be a valuable tool for regulatory professionals. Trying free and publicly available solutions has its benefits but also several pitfalls. Gen AI technology is advancing rapidly, and becoming familiar with available solutions will provide a huge benefit to regulatory professionals. There is no threat from AI to regulatory professionals’ careers in the next decade or two. AI will make regulatory professionals more productive just like how the internet did.

Lexim AI offers an end-to-end solution that cuts regulatory analysis from months to minutes. It brings all relevant regulatory documents and data into one location. It has successfully integrated Gen AI into its platform to help regulatory professionals accelerate time to market, improve quality, and reduce risk.