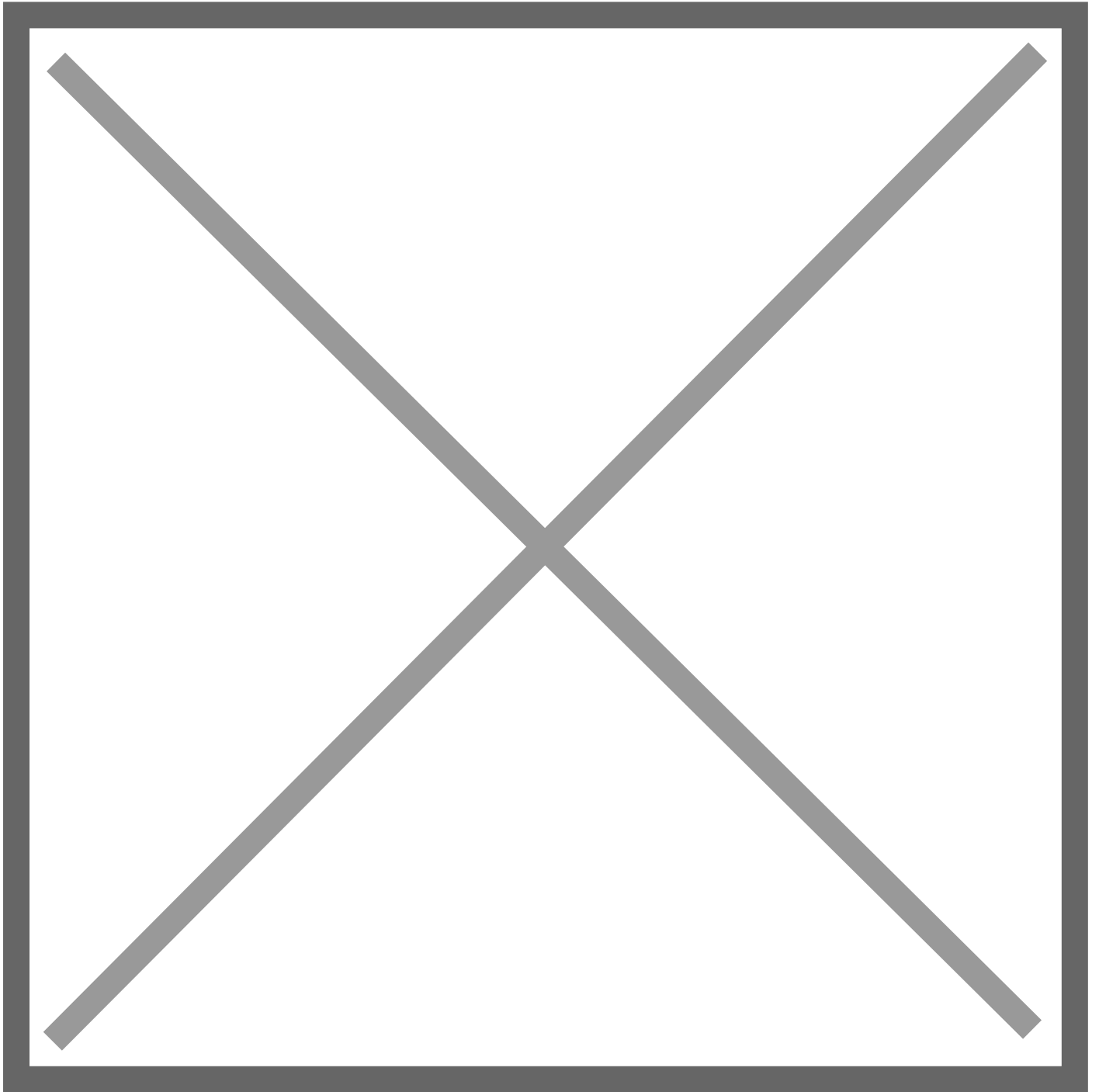


Pfizer Logo

From N^o H^o to Better Potential Medicines: How Language Translation Technology Is Being Applied to Drug Design

Tuesday, December 3, 2019



This is the first in a two-part series.

If you've ever used Google Translate, you've seen how the app can effortlessly translate between two very different languages, such as going from English to Chinese. Now, the same technology is being applied to a new challenge: building better medicines.

This technology, known as sequence-to-sequence (Seq2Seq), is a type of [machine learning](#) framework behind many of the language-processing apps we use today — from Siri and Alexa to customer service chatbots. It

works by taking an input from one domain — say, a sentence in English — and producing an output in another domain, like a sentence in Chinese. Because the model looks at sentences in their contexts, accounting for grammar and word order, it's able to produce more natural sounding translations.

Scientists are now applying this machine learning framework to help design better drugs. And similar to the challenges of translating between languages, the model can be trained to potentially help scientists build more active and potent molecules. “I think of molecule optimization as an editing or translation problem,” says Farhan Damani, a Ph.D. student in machine learning at Princeton who, as a Pfizer intern, worked to develop the technology. “We’re not building new molecules from scratch, rather we’re using an editing approach to tweak and optimize certain properties of existing ones,” he says.

By [studying large data sets](#) of molecules, the model can “translate” an existing compound to a new compound with improved biochemical properties such as potency or fat solubility. “Humans can only reason in two to three dimensions, and these models can optimize for a multidimensional set of criteria,” says Damani.

A Hybrid Approach to AI

Applying [artificial intelligence \(AI\)](#) to drug discovery has been a rapidly growing area, but there have been challenges thus far. Traditional machine learning models can be used to create entirely new molecules, but these computer-derived molecules don't necessarily have the properties to make them potential therapeutics. For example, they may be too toxic or simply wouldn't bind with a target. “From a practical perspective and from initial attempts, we’ve had chemists come back to us and say, 'There’s no way we’re going to use the compounds that have been generated from these other frameworks,’” says Stephen Ra, Machine Learning Lead, Medicinal Sciences.

In order to realize the potential of AI for [drug design](#), improved machine learning models are needed. The Seq2Seq approach is one of a handful of new and different machine learning models that are being developed and evaluated in Pfizer's Medicine Design group. While most traditional machine learning models attempt to learn how data was generated in order to generate more examples like it, the Seq2Seq and similar models take a different approach by starting from a candidate compound with certain characteristics and “translating” (or generating) these compounds to have, for example, improved Quantitative Estimate of Druglikeness (QED), a score that tries to capture a compound's overall quality based on its physiochemical properties.

It's like taking the first draft of a novel and having an editor refine its style, structure or pacing, while keeping the overall integrity of the story. “It's a nice hybrid of relying on our experts' intuition and experimental results, and then using machine learning tools to automate certain tasks”, says Damani. “And having well-informed chemists involved with this project has really helped us develop approaches that are relevant to building potential new medicines.”

A Moment of Affirmation

Ultimately, Damani was encouraged that the Seq2Seq approach might have real-world utility when he saw the results of a particularly difficult experiment where they were able to further improve upon an already optimized set of compounds. “That was my moment of affirmation. But there's still a lot of work to be done,” he says. , This is the first in a two-part series. If you've ever used Google Translate, you've seen how the app can effortlessly translate between two very different languages, such as going from English to Chinese. Now, the same technology is being applied to a new challenge: building better medicines.

Originally published, Tuesday, December 3, 2019