

[Interview] This Vancouver-based Startup Plans To Boost Drug Design With AI

Nov. 6, 2019 by Andrii Buvailo

Variational AI is a newly formed artificial intelligence (AI)-driven molecule discovery & drug design startup out of Vancouver, British Columbia, Canada. The company has developed Enki, an AI-powered small molecule discovery service.

The founders of Variational AI are planning to build on top of their state-of-the-art expertise in machine learning, reflected in more than 40 research publications, including those presented at NIPS/NeurIPS, ICML, ICLR, CVPR, ICCV, and other top events in the area of artificial intelligence research.

The organizing principle of Variational AI is that the exponentially growing cost of drug discovery can only be halted if the pharmaceutical industry shifts the paradigm by which it searches the space of molecules. Variational AI has developed a machine learning algorithm that organizes the full space of 10^{60} drug-like molecules based upon their pharmacological properties rather than their chemical structure, enabling state-of-the-art QSAR and transformative multi-property inverse QSAR/QSPR. They use sophisticated deep learning to build a nonlinear ligand-based pharmacophore, which implicitly accounts for induced fit, solvation, entropic effects, and multiple binding domains. Just as a chess grandmaster develops an intuition for good moves by studying millions of games from early childhood, their algorithms learn pharmacological intuition from vast volumes of raw experimental data that would overwhelm any human. Within their property-based search space, they simultaneously optimize pharmacological activity, synthesizability, ADME, and toxicity, directly producing highly novel lead candidates. By integrating multiple assays for each pharmacological property across many datasets, they avoid overfitting to any single approximate or noisy assay and increase the probability that your final drug candidate successfully progresses through clinical testing.

Canada has been emerging as a global hub for artificial intelligence technologies, and it also seems to be a suitable place for building AI-driven drug discovery startups. Recently, I made a brief review of top AI-augmented drug design companies in Toronto, a city of biotech entrepreneurs and technology innovators, and found that 13 top-notch companies are already showing considerable achievements in the pharmaceutical space -- having raised collectively more than \$60 million in early-stage venture funding and grants over the last several years. But Toronto is not the only "sweet spot" in Canada to build a successful company in the pharmaceutical space, especially when it comes to computation and artificial

intelligence. Variational AI is one recent example of a company in Vancouver -- another promising destination for AI entrepreneurs, thanks to its rooted academic background, strength in applied AI, and rising support for the entrepreneurial community by the government of British Columbia. According to the AI Network of British Columbia (AINBC), Vancouver has 150+ AI/machine learning start-ups working across a range of application areas.

I have asked several questions to Handol Kim, Co-Founder, and CEO at Variational AI, to have a glimpse into the plans and aspirations of the newly formed company, and find out how Variational AI is going to compete with a growing number of AI-driven drug discovery vendors out there. Below is a brief interview:



Handol Kim, Co-founder, CEO, Variational AI

Handol, at what stage is your drug design platform? Do you already have some proof of concept and how does it compare to existing industry benchmarks (if any)?

The standard benchmark tasks used to evaluate AI algorithms for molecular property optimization are a maximization of the Crippen octanol-water partition coefficient (logP) and the quantitative estimate of drug-likeness (QED). LogP and QED are useful heuristics for optimizing ADME properties and can be evaluated quickly and reproducibly, so the quality of optimized molecules can be compared consistently between algorithms. On these measures, we produce molecules up to 3.3 times better than competitors. AI molecular property prediction is commonly benchmarked using the NIH's Tox21 dataset, comprising in vitro measurements of twelve toxicity pathways across ~8000 molecules. Again, we achieve the best performance in the world predicting the properties of the held-out test set designated for this benchmark.

Moreover, our algorithm only requires a fixed training set of molecules on which properties, such as binding affinity or toxicity, have been measured. In practice, this training set will include the collected experimental data from a drug development pipeline, such as high-throughput screening, in vitro measurements of ADME and toxicity, and in vivo experiments on a smaller subset of molecules. It is only by leveraging this experimental data that machine learning can avoid the approximations, biases, and false confidence of structure-based drug design.

In contrast, the most successful competing approaches use reinforcement learning, genetic algorithms, or particle swarm optimization, which require property evaluation on hundreds of thousands of novel, previously-unsynthesized molecules progressively selected by the algorithm. While computational approximations like clogP and QED can be evaluated inexpensively on novel molecules, these competing approaches will encounter severe difficulties with experimental data, for which evaluation of a novel molecule requires onerous synthesis and testing. As a result, these competing approaches will not generalize well from the standard benchmarks to practical drug discovery problems.

Unfortunately, there are no standard benchmarks for pharmacodynamics. We are currently constructing a benchmark based upon docking scores, which will allow molecular property optimization to be evaluated for pharmacodynamics. Docking scores have the same general form as binding affinity, with the advantage that the quality of the novel, optimized molecules can be evaluated inexpensively and reproducibly. Of course, experimental data will be used for non-benchmarking efforts, and the ultimate test is the design and experimental characterization of a novel drug molecule.

Can you briefly outline some key technologies that are used or planned to be used to build your platform? For example, do you use deep learning models? Where do you get data to train models?

We use a variety of standard deep learning technologies, including variational autoencoders, recurrent neural networks composed of long short-term memories (LSTMs) and gated recurrent units (GRUs), and attentional mechanisms. However, our exceptional performance is driven by our novel approaches, including molecular representations comprising multiple SMILES strings (simplified molecular-input line-entry system) of a single molecule that are processed in parallel, and atom-based information exchange between these parallel processing hierarchies.

Our algorithms implicitly construct a nonlinear, ligand-based pharmacophore directly from experimental data. Unlike traditional ligand-based techniques, our powerful deep learning architecture captures complicated interactions between fragments and can leverage related experimental data (e.g., binding to similar proteins) to infer a property of interest. In contrast to structure-based drug design, which cannot directly account for induced fit, solvation, entropy, and multiple binding domains, our approach implicitly models all contributions to experimental measurements. For example, by including data from in vivo animal models of disease, we can directly optimize with respect to the full effects of the molecule in the body. Small amounts of such valuable but expensive data can be buttressed with large amounts of inexpensive data from in vitro experiments, or even docking scores.

We have compiled an extensive collection of public datasets with which we train our machine learning models. In a commercial engagement, we combine this with data from the customer, which characterizes the exact properties they wish to optimize.

What is the major competitive differentiator that you believe will help your company “shine” in the market of artificial intelligence solutions for drug discovery? Or maybe there are a number of factors that create a unique synergy in your company, making it special?

It is possible to achieve some success in any domain using off-the-shelf machine learning algorithms designed for text or images. However, drug discovery differs significantly from English-to-French translation or autonomous driving. The best performance in a new domain like organic molecules and drug discovery can only be attained through fundamental machine learning research and the development of novel domain-specific algorithms. We have an exceptional team of accomplished

machine learning and chemoinformatics researchers, who have been working on this problem for two years.

According to a report by BPT Analytics, small molecules are the primary drug modality that is being developed by most AI-driven startups in the pharmaceutical space. Why do you think this is the case, and what are the specific reasons you decided to build your company around designing small molecules and not, let's say antibodies?

Small molecules present a unique opportunity for AI. Humans have either synthesized or extracted from natural sources no more than 10⁹ molecules. If you synthesized one copy of each of these molecules, they would fit in a single bacterial cell. In contrast, it is commonly estimated that there are 10⁶⁰ drug-like molecules. One copy of each of these molecules would fill a space much larger than the earth. Humans have studied almost none of the available drug-like molecules.

As you have reported in your blog, most leads are either known compounds or identified via high-throughput screening of existing molecules. From this starting point, lead optimization is by its nature a local search over molecular structures. The overwhelming majority of the 10⁶⁰ drug-like molecules differ significantly from any molecule in a comparatively small million-molecule high-throughput screening library, and could never be found via this approach.

Machine learning offers the possibility to model and search over a much larger space of molecules, identifying lead candidates that would never be contemplated by high-throughput screening. Moreover, it can simultaneously optimize for strong target binding, minimal off-target effects, good pharmacokinetics, and low toxicity, while maintaining synthesizability. Its suggestions are truly drug-like, rather than hit-like or lead-like. Machine learning can ingest vast amounts of experimental data, leveraging subtle correlations to other related experimental measurements, and accounting for exceptions and idiosyncrasies that diverge from simple models and theories.

What is your business model and how you are planning to get early customers onboard? Have you already raised money for your operations and building the platform?

Enki is offered as a molecule discovery service. We deliver molecules to our customers, not code. We charge a technology access fee, milestone payments, and potentially royalties on net sales, so we are aligned with customer success. We are working with on-boarding initial customers and are concurrently raising a seed round of funding.

What is it like to grow a technological startup in Vancouver? How local government supports technology entrepreneurship in the region, and do they have specific programs for the life sciences and pharmaceutical artificial intelligence?

While Canada is a well-known AI powerhouse, with AI pioneers and luminaries such as Geoffrey Hinton, Yoshua Bengio, and Richard Sutton based in Toronto, Montreal and Edmonton, respectively. Vancouver is less well-known. However, in the life sciences, Vancouver is home to biopharma leaders such as Zymeworks (NYSE: ZYME), Xenon (NASDAQ: XENE), Arbutus Biopharma (NASDAQ: ABUS), STEMCELL Technologies, and AbCellera. Thus far, we have not seen many other AI for drug discovery start-ups here, but fully expect that this is changing: there is far too much talent and expertise here to sit on the sidelines.

Moreover, Vancouver deep technology start-ups tend to “swing for the fences,” with highly ambitious, moonshot companies such as General Fusion, Carbon Engineering, and D-Wave all being local champions of the high-risk-high-reward, transformative mindset that seems to be common here. It might be something in the water, but many Vancouver companies are literally trying to change the world.

As for government support, there are a myriad of programs from the Canadian Digital Technology Supercluster, and National Research Council (NRC) AI Challenges, to the Scientific Research & Experimental Development (SR&ED) program that all Canadian tech start-ups enjoy. Most relevant to AI for drug discovery in Vancouver, is the Industry Innovation (I2) Program from Genome BC that provides critical non-dilutive, repayable funding of up to \$1M.

Handol, thank you very much for your time!

Dear readers, if you liked the interview, please, share it and leave your comments below. I'd love to hear your thoughts!

If you have more questions about Variational AI and its drug design offering, please, follow the company on LinkedIn.