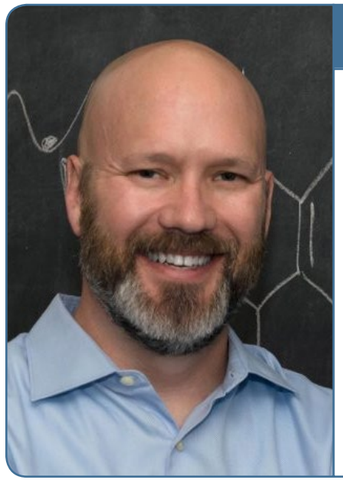
THE FUTURE OF MACHINE LEARNING AND AI WITHIN DRUG DISCOVERY



BRANDON ALLGOOD, CTO and cofounder, **Numerate Inc.**

Brandon Allgood is the CTO and cofounder at Numerate, Inc, an AI driven drug discovery company. Brandon leads research and development of Numerate's AI platform. He also leads the data science group, is the primary technical lead on both internal and collaborative drug discovery programs and is responsible for the technology vision at Numerate. Brandon has previously served as Director of Computational Science at Numerate and as a Research Scientist at Pharmix. He received a B.S. in Physics from the University of Washington, Seattle, and a Ph.D. in Computational Physics from the University of California, Santa Cruz. Brandon has authored scientific publications in astrophysics, solid-state physics, and computational chemistry and biology and has 15 years of experience in AI, mathematical modeling, and large scale cloud and distributed computing. He advises a number of venture capital firms and start-up companies, is a founder of the Alliance for AI in Healthcare, a member of the Forbes Technology Council and a UCSC Foundation Trustee.

What do you think are the most important technologies impacting discovery chemistry at the moment?

Obviously, machine learning, specifically. On the attrition side, it would be deep learning. On the hit identification and activity prediction, the data that you have is often too small to support deep learning. That's a little more difficult there. Cryo-EM is becoming a nice tool to start to bring more structure-based work to targets that are generally not structure-based. I'm personally excited about quantum. WuXi Nextcode is the only company out there that's shown how would it could be used. I think that's still 10 years off.

What makes artificial intelligence so important in discoveries of chemistry?

There was a study done a couple years ago, about the IRR of internal R&D in the pharmaceutical space. The internal rate of return is essentially approaching zero. We're already below the cost of capital. Soon, pharma should just burn the money instead of investing in R&D. What are the drivers of this? The fact that we have machinery set up for a very different reality than the one that exists today; we've machinery set up for the blockbuster drugs. Now we are faced with the fact that all the low hanging fruit is gone. No longer can you inhibit an enzyme, lower cholesterol and make billions of dollars - the diseases we're tackling today are generally more complex. They're likely polypharmacological. Additionally, the diseases have much smaller patient populations. We need to take a different approach to drug discovery. I can back that up with the statistics, and the work that I've been doing for the last 20 years; I think that machine learning is the solution, if not one of the solutions, to making drug discovery profitable again, being able to unlock complex diseases, and to finally have an effect on the failure rate. If you look at PhRMA's funnel slide, the numbers have not changed in over 10 years. That tells me we're not learning from our mistakes. With machine learning, we can. If you make every decision in discovery chemistry, informed by every molecule you've ever made before through machine learning, you will start to learn things from your previous mistakes, and you should start seeing lower and lower attrition.

What do you see as the key benefits in improving translation through AI modeling?

With machine learning one can get to a drug faster/cheaper, having made fewer compounds and having spent less money and less time, which is great. But that's just taking an existing process and making it more efficient. Yet, what's interesting about machine learning is now one can now ask what are the things we can do today that we couldn't do without machine learning? Those are the nonlinearities and they exist around translation. That's translating a low-throughput, high-content biological assay that would never be amenable to a high-throughput screen. An assay that is expressing a phenotype that's using primary human cells and very close to this disease. If you show this to a doctor or a biologist, they'll say that phenotype is the disease phenotype. If you can drive your program based on that, you should expect a much lower rate of failure in the clinic due to efficacy. On top of that, pharma is still making approximately 3000 compounds a program, with only one in four making it into the clinic. That's 12,000 compounds synthesized and tested to get one compound in the clinic. Our ability to find compounds with good ADMET and toxicity properties has been dismal. It's largely been left to the medicinal chemist to remember a certain molecule they made, or a certain problem they ran into in the past. Yet if you had machine learning models that are good, and you evaluated every compound against every ADMET property, every off target that could possibly exist, then for every decision you made, you should start to see attrition. I'm not saying every program will end up in the clinic. However, if your failure rate is one in four, just a small change could be huge.

What are the main challenges that you see in this field?

This is an industry that has been data-rich and algorithm-poor and has been forever. People weren't thinking about machine learning when they collected the data. We started with the idea that we'd never get access to data inside of pharma. We built machine-learning algorithms that could deal with publicly available data with the noise, the bias, the non-reproducibility. More recently, now that the hype wave has taken over around AI, pharma's starting to share their data with us, which means I was excited to get access to all this clean data from pharma. I realized quickly that I'm very happy that we had to break our pick on the publicly available data, because the tools we use there have been very helpful in wading through the data we're getting now from pharma, which is slightly better quality, but not what my imagination had come up with. The data was inconsistently collected, assay conditions were not recorded properly, they use magic numbers. It is awful. Things are not stored in a consistent format. There's a lot of issues there. There are recent articles that have come out asking: why hasn't AI had a bigger impact on industry generally. Most industries are in the same boat, the data is the thing that's holding back AI, not the other way around. The second thing is that there are too few people that understand machine learning and the drug discovery and development process. You need some of these experts. The problem is we don't have them yet, and we can't train them fast enough. There's a data problem and a personnel problem.

What are the next steps for your company's research in this area?

We were forced to focus on assets and not hype because we started working on this before the hype. We've evolved to the point where we are pursuing our own internal programs. That's the vast majority of our work - internal programs. We do have multi-target collaborations with Takeda and Lundbeck, but it's really around starting more internal programs and pushing them further into the clinic. In terms of machine learning research, we probably have the biggest suite of models around ADMETox than, at least, any of the small companies. However, now with our partners, we are starting to move into taking clinical data and building models of primary toxicology, patterns of neuro and cardio toxicology, and applying those in a pre-clinical setting, optimize compounds, based on primary outcomes.

What are the top three takeaways from your presentation?

The machine learning and AI effect is real, it's not just hype. We need to focus on the data and cleaning that data up. We're making big strides in machine learning and applying it here, but there's still a long way to go. This is far from being a magic bullet.

What are you looking to get out of events like this?

There's an altruistic part of this; I want to stem the hype around machine learning and AI and show people what it can and can't do. It's not perfect. From a company perspective, it's engaging with decision makers within larger companies to start talking about how we could apply these machine learning models and learning more about what their pain points are, because as a small company I can imagine what pain points exist, and how to think about these different programs. I'm not sitting inside these big companies, so understanding their pain points is one of my main objectives.