FIRESIDE CHATS

Artificial Intelligence in Drug Discovery

DrugBank is a widely used data source for companies working in the pharmaceutical, healthcare, and technology space. The data integrates seamlessly into in-house products, to enhance outcomes for data-driven decision-making.

twoXAR Pharmaceuticals, a drug development and discovery company, uses DrugBank to assist in its artificial intelligence (AI) solutions. We chatted with Aaron Daugherty of twoXAR to discuss AI in the drug discovery industry, and how DrugBank plays a role.

twoXAR

Aaron Daugherty
VP, Discovery

ABOUT TWOXAR PHARMACEUTICALS

twoXAR Pharmaceuticals is a drug discovery and development company focused on first-in-class small molecules. The company currently has a development portfolio of over 18 diseases. twoXAR saves years in drug development while generating a 30x hit rate at in vivo efficacy milestones over traditional methods.

WEBSITE
twoxar.com

INDUSTRY
Drug Discovery

USE CASE
Artificial Intelligence

Please introduce yourself

My name is Aaron Daugherty and I'm the Vice President of Discovery at twoXAR. Since starting at twoXAR as one of their first employees, I have helped build twoXAR's drug discovery platform and lead the Discovery Science team's efforts to discover potential treatments across a wide range of diseases. I have a PhD in Genetics from Stanford University, and prior to that I was a Fulbright Scholar, receiving my BSc in Biology from the University of Richmond.

Can you tell us about twoXAR?

At a high level, twoXAR is a pharmaceutical company. We're finding new medicines across a wide range of diseases. Of course, what differentiates us is that we're an AI enabled company; we use AI software to improve success rates and find efficiencies of what we're working on. As an example, we work in the early stages of de novo drug discovery, and normally that's a process that will take on a scale of four-to-six years to come up with a hit, with a success rate of around 1%. For us, a normal project will start from scratch and be moving into in vivo or in vitro models of the disease in about a month. So by using our software, we can essentially cut off about three years of experiments. We also have a much higher success rate. We'll test somewhere around 10 unique, first-in-class mechanisms of action and we see about 30% of those being significantly efficacious at those preclinical milestones.

So what does a VP, Discovery do?

I should give a little bit of a background on how I got here, first. I originally came in to help develop the software we have at twoXAR. Essentially, I was hired because the CEO needed to step away from software development to focus on being a CEO full-time. I stepped in and was helping to build out our platform. Then over the years we've found some folks far smarter than myself to continue to develop our software even further. So today working with and managing that team of extremely talented developers is about half my time. The other half of my time is spent working with our equally talented drug discovery researchers who are using the tools we're building to discover new drugs across the 18+ diseases we've got going right now. Of course, there's also time to wedge in communicating the unique approach we use at twoXAR. So, at a really high level, I have one foot in computation and one foot in biology, and that's what makes my job pretty awesome.

Can you tell us a bit more about artificial intelligence in drug discovery? How'd you get interested in the field?

Prior to joining twoXAR, I was on the basic research side of things and was using AI to deal with the complexity of biology. It was a pretty natural transition to start applying some of those skills at a company that was trying to discover drugs. I really loved the basic research, but it was much more fulfilling to have that impact of actually going out and identifying new treatments for patients; it was hard to beat that. Then this role also allowed me to combine biology research with some of my other interests on the engineering side, like getting to build new analytical tools—it was well-timed.

Six years ago, when we started in drug discovery, AI was not all that common. You'd say AI and you maybe get a couple eyebrows raised and people wouldn't know what it was. Obviously, that's changed a lot in the intervening six years. Now there are a lot of people doing some aspect of AI in drug discovery.

There's so many different approaches to using data for drug discovery. Can you tell us about how twoXAR uses data and the approach you take?

One of the things that really separates us from others in the drug discovery space is that we use a really heterogeneous mix of data. Our strength is our ability to pull together so many different data sources—we are now using data from over 60 different data providers. Obviously, DrugBank is one of those and actually one of the first ones we included.

Over the years we've grown to include molecule data, from DrugBank and others, and a number of other foundational data sources. Things like systems biology data: protein-protein interaction networks and pathway information. As well as medical ontologies. Then we also integrate disease-specific, patient-derived data. Data types like gene expression, genetics data, micro RNA data; really all the omics types that folks may have heard of, as well as data that are phenotypic; claims data and electronic medical records.

If we have the opportunity to get our hands on additional data, we look to integrate it because one of our founding principles is to look at numerous independent data sources to help us get a more holistic model of the diseases we are studying.

You mention you've used DrugBank data since the beginning. In what ways do you use it?

Actually, we started using Drugbank data before twoXAR was twoXAR. Like DrugBank, twoXAR was started as an academic project, and it grew from there. Even from our initial beginnings, we've used DrugBank to source some of our molecule data. As I mentioned before, we like to source as much diverse data as we can to power our solutions, but sometimes it's really difficult to find good data. There are some datasets that have millions and millions of compounds but are really sparse data of intermittent quality. That's why I've always appreciated DrugBank—there's really good depth of information on any one molecule, but there are also something like 13,000 molecules. This is a powerful combination of depth and breadth for us, and that's the type of data we focus our energy on sourcing.

So how do you go about deciding what to source and how do you decide its quality?

Right now we're looking for things that are independent. We have 60+ different data sources, and so we search for data that's statistically independent, and thus adds a new perspective to our data. We want data that can tell us something new about molecules or diseases or systems biology. Once we have our bases covered in that regard is when the real challenge arises: measuring the data against our stringent statistical analysis to make sure we're seeing the distributions we expect, or there aren't many outliers; standard data science approaches.

Another big thing for us is that we really value interpretability. We are putting together a ton of visualizations. Our drug-discovery researchers are the ones who ultimately verify our predictions. So for any given prediction we're spot checking ten, a dozen, or hundreds of

data points to make sure they make sense. In other words, do all the pieces add up to tell a factual story and form a novel disease hypothesis?

Interpretability is an interesting problem when it comes to AI because there are a lot of techniques that use a 'black-box approach.' Is that something you focus on at twoXAR when you're designing or picking your approaches?

That's one thing I hope I've added to at twoXAR over the years, and something I can speak to. I'm very familiar with the need to get accurate results at both the research bench and at the computer, not just the computer. Everything we've built is fully interpretable. When we share with a partner or our internal team that a molecule is likely to treat a disease, we can trace back exactly what algorithms were used, exactly the data points it came from, and we do a huge amount of quality control on all of our data. This helps us improve our success across teams because now our non-clinical team understands the disease hypothesis and that's all because of the care we have put into our interpretability.

That's a really great approach, and that it helps with further experimentation is a really interesting insight. In regards to the broader field of AI in drug discovery, what are some of the trends you're seeing in the industry?

The biggest trend is that AI is fairly prevalent throughout drug discovery now. It's at least being talked about a lot. When twoXAR started using AI it was really just getting a toehold in the broader industry, and now it's a much bigger trend. A year or two ago a ton of companies rushed to use AI in drug discovery. There was huge hype, and a lot of companies thought it was going to immediately solve all of their problems, but that's honestly not a reasonable expectation for any single tool at the end of the day. That's what AI is—a really powerful tool. It might not immediately solve all of the problems we face in drug discovery, but we are starting to see the impacts in the industry and at twoXAR we've been seeing those results in detail. For example, we are now consistently generating very promising preclinical results for disease areas from lupus to glioblastoma.

> "That's what AI is—a really powerful tool. It might not immediately solve all of the problems we face in drug discovery, but we are starting to see the impacts in the industry and at twoXAR we've been seeing those results in detail.

Omics data is an advance we are seeing a lot in the medical field. Services like 23andMe are providing affordable genetic testing and make it more approachable for the average person. Do you think this sort of thing will impact your field?

In my opinion, the more open data we get out there, keeping in mind the importance of security of said data, the more vital information we have related to health and wellbeing of society, the better our world will be. People can learn from that data and improve on it. So I think companies like 23andMe have done a great job with that. And in general, there's just so much open data becoming publicly available, that now it's not so much a matter of not having enough data to make a real impact, it's more about finding the time and resources to evaluate the data sources. It's a good problem to have, mind you. The more data we have, the more folks like us that can go out and build algorithms to take advantage of the data pool and hopefully, positively impact human health.

Awesome. What's the most game-changing experience you've had while working at twoXAR?

It's a bit of a cop-out but I'm probably going to say the whole thing. I was going down that academic route, just like everyone who goes to do a Ph.D. Then I started to lean towards data science and thought I was going to have to set aside my biology experience just because I didn't see as many opportunities at that time to harness both of them. Right at that time twoXAR stepped in and filled that gap. In fact, I found this amazing role that allows me to do both. It allows me to do the computation that I enjoy nerding out on, but I also get to apply my background in biology that I spent a lot of time learning. I think it has a really big impact. Maybe to improve my cop-out, it really was getting to that first stepping stone with twoXAR and realizing I can use both of these passions I have, in a career I love, with a company that is actually making a big impact in discovering treatments for patients in need.

Thank you to Aaron and the twoXAR team for their insightful conversation about AI in Drug Discovery. If you are looking for more information about twoXAR please reach out to them directly.

DrugBank offers a number of commercial products to help customers like twoXAR get the structured and comprehensive drug they need for artificial intelligence. If you are interested in learning how our products can help your company, please email us at info@drugbank.ca.

DRUGBANK

DrugBank is the world's largest knowledge base of structured pharmaceutical data. DrugBank enables data-driven decision making in the healthcare, pharmaceutical, and medical spaces. By using DrugBank, companies and researchers can more efficiently discover or repurpose new drugs, and improve treatments for patients. The core of DrugBank's knowledge base consists of proprietary authored content describing the clinical level information about drugs such as side effects, drug interactions, as well as molecular level data such as chemical structures, metabolism, and what proteins a drug interacts with.

Visit www.drugbankplus.com to learn more

twoXAR

twoXAR Pharmaceuticals is a drug discovery and development company focused on first-in-class small molecules. The company currently has a development portfolio of over 18 diseases. twoXAR saves years in drug development while generating a 30x hit rate at in vivo efficacy milestones over traditional methods.

Visit www.twoxar.com to learn more

DRUGBANK

FIRESIDE CHATS

drugbankplus.com/**downloads**