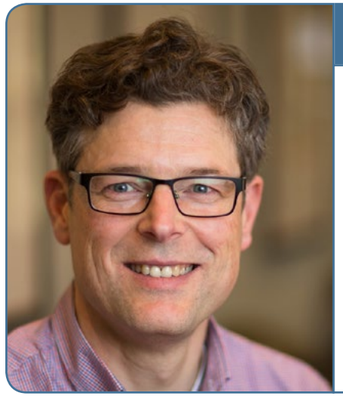
AI AND SUPERHUMAN DRUG DISCOVERY



WILLEM VAN HOORN, Chief Decision Scientist, **Exscientia**

Willem van Hoorn gained a MSc in Chemical Engineering and a PhD in computational chemistry at the University of Twente, the Netherlands followed by a postdoc at Yale with Bill Jorgensen. He subsequently spent a decade at Pfizer Sandwich focusing on computational techniques for HTS triage and combinatorial library design. This was followed by a position as IT consultant at Accelrys (now Biovia) assisting a range of clients from small biotech to big pharma. Since 2013 he is pursuing his long term interest of applying computer algorithms to drug discovery at Exscientia.

Why is artificial intelligence so important in drug discovery?

It's important in society as a whole, it's not only drug discovery. If you look at facial recognition, it has profound effects, and goes beyond playing chess. Drug discovery is this one other field where AI can have a tremendous impact on productivity and bringing it up instead of down. With any endeavor, you would be foolish not to look at it if it can help you. You might not need it in certain spheres such as the care industry where human contact is crucial. However, where you deal with data you would be a fool not to exploit AI. It's as simple as: this is the data that I am presented with. Can I learn from that? What am I doing with it?

Can you tell me more about your specific applications of AI?

We are very focused on drug discovery. We seldom do any target discovery; that's not really our differentiator to this, when you're trying to find a way to lead to candidate optimization trajectory. Sometimes, to get a lead, there is a lot of party material out there or collaborators share their HDS results. They might say they want a compound with certain properties, and then we come up with structures that fulfill said properties. What we do is still disruptive to some extent, but on the other hand, you make some compounds, you put them in assays, you learn from that, and then you know what next to make.

Could you tell me a little bit more about the benefits of the cutting-edge central chemist platform that you developed?

We developed the platform ourselves and we have been around since 2012. We have always worked with customers. The needs and the demands of customers have driven the development of the platform; we've been very much needs-driven. Given that most of us are ex pharma people, we know what problem we are trying to solve. We built that platform, it is basically an ongoing project. I don't think that will ever end. The goal is to make humans more efficient, it's not something that replaces people. It helps you make better decisions, but humans are still necessary. We call that superhuman drug discovery - if you work with a machine, you are better than anyone else who doesn't work with a machine. It's the same with chess; when Garry Kasparov was beaten. Since 1997, machines win, AI wins; the best player couldn't compete against the best AI. AI has improved a lot, but chess is still there. In chess it became quite clear that machines beat people, but people using machines beat machines. There is still something that we as humans bring to the table and that has to do with the strategy versus tactics. I'm a very bad chess player, and I tend to lose based on tactics. I just want to win so I move everything forward, then I get wiped off the board. I don't plan ahead. However a computer programme will be aware of that and will use that to its advantage. There's no way you can recover from that. It's similar to what we do, a lot of it is about not making a particular mistake because it won't work. Equally, active learning makes us ask ourselves what we need to know. Drug discovery is a problem of known unknowns, and you know exactly what you need to know. That compound is a drug because it's active, it's selective, it's absorbed and it's excreted, it doesn't kill the patient. You can make it in sufficient quantity, etc. If we talk about small molecules, we know exactly what we need to know. You could do brute force with all possible compounds versus all possible assays you could submit, but that is a theoretical concept. What is the most informative experiment to do? What should I do now in order to get the information that I need in the next round?

Would you say that that's one of the main challenges?

It's not learning from Big Data. It's like learning from small data. When you're working on drug discovery projects especially in the beginning, you've only made a handful of compounds. The question then becomes, what do I do next, given the handful of data that I have?

Regarding the future of AI in drug discovery, you said you're driven by the problems or the required solutions of your customers - what do you anticipate the future problems being?

It all revolves around productivity. Discovering a drug is too expensive; there are real costs involved. In the end, you need to increase the success rate. That has to do with target discovery. That's not really what we do. If you look at small molecule discovery, there are only two decisions that count. What is the target? Which chemical series are we going to pick? The first decision is out of the biologist's hand. It's only in phase 2 that you know if you're making the right decision or not. The challenge is, how can I keep the series alive and have it work with as many series in parallel as possible for as long as possible without wasting resources? A lot of it is resource allocation, which is usually fixed. I worked at Pfizer for about two years, and there are three chemists and so many biologists about at the end of the two years who either have a candidate or if not the project stops. If you can bring that down to one year, so you have twice as many shots. Statistics will favour you, if you do these things in a cheap and fast manner. Even if you don't improve your target selection, just by the sheer number you execute you will have a higher success rate.

Are there any key innovations or technologies in this field that would help you with those problems?

What we struggle with is usually ADMET data; the absorption distribution, the toxicity. The big pharma companies have been measuring these for a long time, and they have a big vault of data that's not in the public domain. It's very hard for us to predict these things. We have to generate the data on the series we work on while we work on it. Instead of a big general model, we work with small, localized models. There is a learning cost there that we would like to avoid. Additionally, deep learning works like magic sometimes and then doesn't work at all. I think there was a lot of research to be done in order to determine when these methods would be the most appropriate solution. In a certain context, something simpler may work. I think there are a lot of question marks surrounding these methods. When should we apply them? It's not an 'if', it's a 'when' and 'where' question. The whole field is developing in an interesting way, and things that were initially not applicable to our work are now a considerable part of it.

What are the top three takeaways from your presentation?

AI works, it's already useful. Secondly, it won't replace your job, but it will change your job. It moves fast. Things we did originally, we're not doing anymore. We're overtaken by ourselves.

What would you say is your main purpose for attending events such as this one?

It's twofold. To some extent, there is a business development point, because our potential customers are sitting in the audience. Furthermore, we can see what our competitors are presenting. Finally, I've worked for Pfizer, and a lot of people here have worked for Pfizer – there is the social aspect, i.e. to meet people you've already met and meet new people.

Are there any talks that you'd really like to attend?

This morning, I went to the bispecific antibodies talk; I didn't know they existed, because we are very focused on small molecule discovery and synthetic, organic molecules. If you look at the field of drug discovery itself, one of the best-selling products these days include antibodies, and that's something we don't do at all. So we have to consider whether this could potentially make our work irrelevant overnight.