

# Alex Zhavoronkov: Linking Chemistry and Biology Using Artificial Intelligence

March 28, 2021 by Andrii Buvailo

The team at Insilico Medicine has recently reported that they achieved a major milestone for their comprehensive artificial intelligence (AI)-based drug discovery system: a novel molecule for a novel target discovered with their AI demonstrated efficacy in a broad therapeutic area and reached preclinical candidate stage in Idiopathic Pulmonary Fibrosis (IPF).

Insilico Medicine is a Hong Kong-based company that pioneered the application of deep learning, specifically -- generative adversarial networks (GANs), for drug discovery and generative chemistry.

**Dr. Alex Zhavoronkov**, Co-founder and CEO of Insilico Medicine, agreed to answer several questions about his path into biotech, the company, the discovery, and how they managed to build an end-to-end AI system capable of uncovering novel targets and molecules, linking biology and chemistry into an integral research process.

*This interview text has been edited for clarity and size. Watch the original video for the complete interview.*

**Andrii: Alex, let's start with some personal story. You are from a mixed background combining computer science and biotech, and you held a number of executive positions in tech companies. Then you switched to biotech and started Insilico Medicine, can you tell a bit more about how you got where you are today?**

**Alex:** I did my first two bachelor's degrees in Canada at Queen's University -- one at Queen's School of Business, and another one in life and sciences and computer science. Then I worked for a number of semiconductor companies and I also worked at ATI Technologies, which makes graphics chips and which competes with NVIDIA. Now it is part of AMD and back then it was actually much more competitive with

NVIDIA than it is right now. Today, NVIDIA is a dominant force on GPUs, powering the deep neural networks that we are currently using. Also, very early in my career, I managed to make some money on the stocks in semiconductors, enough money to sustain myself for a few years. I thought, "OK, what am I gonna do next?". "Should I continue a rat race of making money, or better do something more impactful?"

One of my major interests for a long time was aging research. I kind of don't want to accept the grim vicious cycle where people grow, reproduce, and gradually decline and die. If you think about the marginal contribution of curing any single disease, like for example if you completely cure cancer -- if you eliminate just cancer -- it would add 2.6 years to life expectancy in the United States, that's the current consensus estimate, so it is actually a very marginal increase. It does not really change the picture and you need to look deeper and look at aging. Anyway, I decided to go into biotech and apply the computer science approach to problem-solving to try to deconvolute the very complex process into many fewer complex processes and try to see if we can affect some of those processes to make a bigger impact.

I did my grad work at Moscow State University and at Johns Hopkins, and I then worked for a number of biotech companies and then started my own lab in academia in cancer biology and cancer bioinformatics, I started applying aging research concepts looking at data longitudinally, comparing biology at different ages and trying to train different algorithms to predict age and track age using biological data to see if we can train machine learning systems for age prediction and then see if we can deconvolute that knowledge of human biology of aging into something that is more related to disease and if we can find some relevant targets or pathways. It was kind of a shotgun approach. Also in between, we looked at different biological processes that transpired during aging and published a number of papers on the topic. Then in 2013, I started focusing on deep learning, after visiting an annual graphics technology conference organized by NVIDIA. I knew quite a bit about neural nets at that time, but it was only the beginning of the deep learning revolution – notably after neural networks demonstrated amazing accuracy in image recognition competition Imagenet in 2012. At that time, I also realized if deep neural networks were so good at image recognition why don't we use them for age prediction and maybe then we can look at various ways to derive biologically relevant features.

We started developing a range of algorithms that could reduce the dimensionality of data and allow us to train deep neural networks on smaller data sets to identify the most relevant features and by 2014 we had quite a bit of experience on that and we started the company. So, Insilico Medicine was originally founded as a target discovery company from the perspective of aging research and age-associated diseases. Later we expanded into chemistry and then combined chemistry and biology to get where we are today, discovering novel targets and novel therapeutics.

**Andrii:** It sounds unusual, because aging research is a broader, more complex area than discovering therapeutics for a particular target, and you actually started with aging and later shifted to a more classical pharmaceutical research. Anyway, from what I know about you and Insilico Medicine, you are really obsessed with artificial intelligence. I know you even wrote a book about how to understand a robot and the relationship with a kind of imaginary artificial intelligence creature (“Dating AI, A Guide to Falling In Love with Artificial Intelligence”). And your obsession transformed into tangible results – your team seems to have discovered something really big this time, having applied artificial intelligence to uncover a novel target and a new molecule active on that target which is to me is amazing. But do you think this is a major shift in the way drug discovery is done or it is more like an opportunistic breakthrough, being lucky in a sense?

**Alex:** I think this time it is a major shift. We have indeed had quite a few of those opportunistic incremental improvements and some opportunistic discoveries in both target discovery and small molecule chemistry -- you probably remember our work with a company called BioTime that we published in 2017. By the way, we did the work in 2016 and they had to spend quite a bit of time on experimental validation and patenting and then lots of bureaucracy in terms of publishing so we didn't publish in a top-tier journal. But it was still a traceable publication where for the first time we reported an application of a deep neural network to target discovery. We identified an embryonic fetal transition factor called COX7A1 so that was more or less an opportunistic paper – we actually got paid for a little and BioTime spun off the entire company called Ajax and took it public. A substantial part of intellectual property was COX7A1.

We also identified a few other targets during that time using large datasets of transcriptomic data coming from embryonic stem cells that were differentiated into mature cells. We used similar approaches to target discovery for many chronic diseases like fibrosis. We published a number of papers that were also quite opportunistic on generative chemistry – the first one was in 2016, called “The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology”, where we showed that we could generate molecules with desired properties.

The first major proof of concept was done in 2018 and the results were published in 2019 -- you remember our kind of “reasonably famous” paper in Nature Biotech where we actually synthesized and tested predicted compounds all the way to experiment on mice. It was done for a well-known old target, but the catch was that we managed to do the whole thing very quickly – within days, not years. That was a cool milestone for our artificial intelligence system, but that was still opportunistic in nature, and in order for us to convince big pharma that we were preparing a paradigm shift in the way drug discovery should be done, we realized we would need to take some drug candidates discovered by our system into clinic ourselves, and we better do it for a completely new target also discovered by our AI. So we had to combine biology and chemistry into one AI-driven workflow and even more than that – including some ability to predict clinical trial outcomes to add maximum value to this AI-based drug candidate generation process. Now big pharma is largely convinced that artificial intelligence is a big deal in drug discovery – we have lots of partnerships and top corporations are using our system to discover targets and drug candidates.

By the way, besides partnering with companies like ours, every big pharmaceutical company has invested dramatically into AI, some of them hired 600-700 people to focus on data science on AI – and this is on top of what they already had in terms of traditional bioinformatics and cheminformatics resources. So, the whole industry is moving into this AI-centric drug discovery paradigm and they got some experience already. But they have been quite slow so far, and I have not seen any similar result, like what we did, coming out of big pharma -- not a single reported case. Probably, they already spent billions of dollars on AI adoption, but where publications or announcements that would go beyond our capabilities as a small but agile “AI-first” company? Anyway, I think that now we have developed the complete end-to-end system and we have the ability to replicate our recent success over and over again in a systematic way.

While we will continue offering services for pharma corporations discovering targets and drug candidates, like we are doing it today, we will also focus on that kind of very quick “rat races” going after specific new targets in a variety of conditions because that's where we can generate the maximum amount of value and once we reach Phase 2 trial, those assets become dramatically more valuable regardless of how they are discovered – by AI or not -- so people do not need to look under the hood trying to understand and see if they can do it themselves in order to partner, they would just license the ready-to-commercialize asset. It is much better for us to have those assets ready to license, rather than, you know, try to convince people to partner with us. We have done this before, but it is not easy – it is just like fundraising. We have been convincing people to license our AI-driven software and for that, we had to constantly prove ourselves in order to find partnerships to earn money.

You know It is interesting how things work: some companies partner with big pharma and get tens of millions of dollars up-front without having any tangible targets coming out of their AI-systems. People are just doing those crazy deals on the expectation of something coming out of quite hype-prone technology. It wasn't the case for Insilico Medicine, it was quite hard for us to get large sums of deals and we had to constantly publish and show results to have even those. Who knows why.

Now with our latest discovery, I think it is irrefutable evidence that our system is capable of not just discovering something interesting, but actually delivering rapid innovation, at low cost, at a fraction of the time needed for traditional drug discovery. We showed AI in action, having identified preclinical candidates and having reached IND studies for a reasonably difficult disease with no real cures.

So yes, this time this is a major shift, not an incremental improvement.

**Andrii: As I understand, Insilico Medicine started as a sort of contract research company, as many AI-driven companies in this space, and then now you are kind ready to take on a more challenging thing to create your own pipeline of targets and ultimately drug candidates and basically grow into clinical-stage biotech. This seems like building a pipeline on steroids, using AI.**

**Alex:** Yes, this is correct. In fact, we hired a bunch of really high caliber people to do that, which by the way was very difficult in this market. Because, let's say, in China, where we also do operate quite extensively, it is almost impossible to hire amazing drug discovery people -- most of them are showered with cash and they start their own companies, rather than work for someone. It is a lot of competition for talent, and also a lot of new companies that pop up here, the market being very hot. But we managed to get some of the best in the world drug hunters, so we also have humans with a lot of experience with demonstrated capability of taking hypothesis level molecules all the way into the clinic. So now we have resources and talent to take our programs forward ourselves -- not as a bunch of "AI guys" but as a diverse team of experts in both AI and drug discovery. And you remember -- we are AI-first, unlike many other "regular biotechs", so it is a little different story. We did not start with drug discovery then trying to adopt AI somehow. But we started building the AI-first platform from the very beginning, it is in the DNA of Insilico Medicine.

---

**Andrii: Right, I have been reviewing a lot of AI companies in the drug discovery space over the last several years and it does seem easier to build an “AI-first” R&D company from scratch than it is to transform an existing big pharma corporation or biotech corporation into a data-driven organization.**

**Alex:** Yes, I can not agree more. One way to solve this challenge is by acquiring AI-first companies. And it is strange to me -- we have been on the radar of many biotechs and many pharma companies for a while and a year ago or maybe one and a half years ago some of them could have bought us for relatively little cost. Pharma corporations have a lot of data in-house and they can merge those assets with an advanced AI. But I think that pharma companies should focus on what they're good at -- clinical trials and ensuring that the market is satisfied with the right products so also manufacturing and sales is the pharma's priority. And the actual innovation -- it should be outsourced as much as possible. I guess big pharma CEOs are not that great at building internal R&D pipelines, because research is not what is generating a lot of shareholder value. They need to be ensuring that the Phase 2-3 pipelines are full of really amazing assets and I hope that we will be able to contribute to that. So now we are basically surveying the market of what is going to be “hot” five-six years from now and trying to predict the trends ensuring that we can now generate those molecules and identify those targets much faster to satisfy the big pharma demand five years from now.

Big pharma has an issue with its early discovery programs in that they can't afford to be failing a lot. In contrast, when we were developing our generative pipelines and when we were validating our models we were failing a lot -- and it was not necessarily a bad thing, because AI learns on good and bad results. We were experimenting to see what worked, what didn't, to see which approach to implement. We were asking our outsourcing partners to synthesize even molecules that they believed to be “weird” or outside of regular medicinal chemistry rules -- but still, we needed that activity data to calibrate our models.

Most of the pharma does not have the ability to go testing molecules that are perceived “wrong” based on the opinion of their medicinal chemists. So their “AI guys” rarely have a chance to include such data into the training process, and so they have biased data towards what their medicinal chemists like. When we actually started selling our Chemistry42 software we realized that we are solving a huge problem -- allowing the AI guys in big pharma to go train their models, plug them into our Chemistry42, compare them with everything we have done and generate really cool molecules that would make medicinal chemists happy. It is very important that generative AI systems can synthesize both “normal” types of molecules, that a medicinal chemist would love, and “weird” but potentially innovative molecules, “outside

the box” options. Once humans can see that AI can do great stuff, they are more likely to accept new stuff that AI suggests. And they are more likely to approve testing of such molecules. So you can prioritize novelty or you can prioritize “standard” medicinal chemistry practices and you can satisfy all camps: medicinal chemists, computational chemists, and AI guys. Currently, those areas are quite disconnected at big pharma organizations.

**Andrii: As I understand from your announcement, your AI system is linking chemistry and biology, kind of creating an “end-to-end” drug design system, which is not available elsewhere. From what I have seen with most AI-driven startups in drug discovery, they are usually focused on some particular stage of the process, like, you know, docking or screening, target discovery, but not a complete process from idea to preclinical stuff.**

**Alex:** There are many startups out there and often they have AI in their decks to fundraise easier, some of them do have strong systems and innovation. There are not a lot of end-to-end systems in the market, but I think there will be convergence over time and you will see more and more of those examples. For now, to my knowledge, our system is the most experimentally validated because we have published most of what we did, every stage of building our AI system. Now we have brought it all together and now we have got it working, and validated with the discovery of a new target and a new preclinical candidate heading to clinical trials soon. Though I deeply respect many other scientists and companies, like for example Brendan Frey from Deep Genomics -- they also managed to do a preclinical candidate a while ago, for a monogenic disease. They identified it using genomics data and then generated oligonucleotide candidates, so they deserve praise and recognition. But many others are just using AI as a marketing driver. And an easy way to understand that is to check if they published anything regularly, like us. For us, it was important to publish every milestone down the road, and now we can show a complete story published in papers how we built the end-to-end AI system and finally got to discovering novel assets for fibrosis.

One more thing, to build great technologies with AI components, you need to partner a lot. We are very collaborative and actually, one partnership I would really like to highlight is Arctoris -- they do laboratory automation and robotics. We had a few case studies where we sent the molecules that we generated for a range of targets just to test them out. The guys at Arctoris produced the results very quickly and they

got the biology to work. We like such partnerships and are open to those kinds of things.

**Andrii: It seems like you're not just discovering cool stuff but also building the community!**

**Alex:** I am convinced we are at the beginning of something really big in this case -- a big change and the AI will be in the center of it, and we will need great community and partnering to move things forward.

**Andrii: To summarize our interview I would like to ask your opinion about the future of the biotech industry, especially considering the impact of pandemics and all this economic turbulence. What will be the role of Insilico in all that?**

**Alex:** Well, I think that the future of our industry also depends on the state of the general economy because the economy is not behaving rationally right now. I am just looking at the global economy and I can see a lot of inefficiencies, a lot of people are staying at home, people are not driving, people are not doing a lot of useful work, farming is going down, everything is going down, production is going down -- but the stock market is going up! And the countries are printing money like crazy, there are many restrictions on doing business so the economy, in theory, is in a very depressed state, some areas are on complete life support and I hope that the state of the general economy does not really affect us. I can see that there is a major increase in money supply and the "good" and "bad" companies get the funding and more often "bad" than "good" ones. I really hope that this mad ride in the stock market is not going to lead to some crash that will "inhibit" innovative businesses like ours. Because if there is, you know, blood on the streets and people are forced to search for food and shelter, they do not care about drug discovery and many of those projects might be buried. If the economy continues to grow or there is, at least, a mild landing -- I think companies like ours are going to do very well because we have got an arsenal of tools to not only support the others but also discover some gold for our own pipeline -- those gold mines are gonna open up in time so we are going to pass the Phase 2 and start licensing, those products will propagate on to the market and we are going to help a lot of people and as more funding and more attention are on our industry and on our company -- well we would be able to do more with those

---

resources.

Also, regarding the future, I will be getting into robotics in a very dramatic way a very different way compared to everybody else. It is a completely different data type, completely different approach, completely different people, and very scalable very cheap architecture. So I will expect a lot of discoveries to come from that. I think a lot of people are going to pivot into robotics now because that is the only way to prove that you are not using a human anymore and you are allowing some processes to be completely run on autopilot so we want to get there. I think that once we get a little bit more resources we will be able to go after aging in a bigger way. We also have to go after some kind of quicker wins and safer bets where you have very complex but known targets, maybe in cancer, in immunology. But we of course do not give up on aging, because that's my life and passion, and we have many projects in that area, so with more resources, we would be able to go after more chronic diseases and maybe repurpose some of the molecules. I think that from this standpoint if you completely abstract yourself from the general state of the economy the outlook is very bright for the industry because many other companies are building a novel paradigm with AI, they got resources they got attention, they got people and it is still early days of "AI-revolution". It reminds me of the early days of the Internet -- a lot of "dot-coms" will happen, but there will be some "Amazons" and "Googles" that will be taking the leadership positions and I hope we will become one of those companies.

- Insilico Medicine