McKinsey & Company

The Next Normal

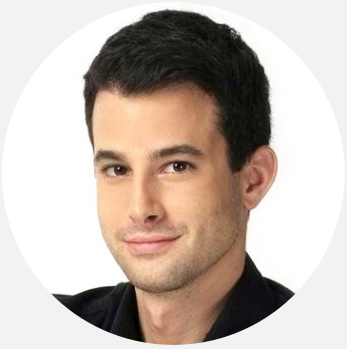
The future of biotech: AI-driven drug discovery

November 2022

Many diseases today don't have a cure. One reason is that drug discovery is difficult: finding and developing an effective medicine is a yearslong and very expensive process. But maybe it doesn't have to be. Experts say AI—if properly integrated into scientists' research—could revolutionize drug discovery, making it possible for more patients to get the treatments they need.

In this edition

Alex Devereson



Christoph Sandler

Lydia The

How AI could revolutionize drug discovery

Artificial intelligence could help scientists develop better medicines faster—and thus improve millions of people's lives. But for that to happen, companies will need to change the way they work.

Human bodies are incredibly complex. It takes many years to discover even just one new medicine to successfully treat a disease. Could artificial intelligence help speed up that process? McKinsey experts believe so. (The following transcript has been edited for clarity.)

Faster *and* better

Lydia The: What excites me about AI and drug discovery is the convergence between technology, drug development, and biology, which is going to lead to better drugs being developed faster—using all of the capabilities that Silicon Valley and the tech ecosystem have developed—to help us have even greater impact on patients.

Christoph Sandler: Today, to discover and develop a drug takes more than ten years.

Alex Devereson: We might be able to have drugs in one-tenth of the time, from being discovered to being able to treat patients. Today, many diseases simply have no treatments whatsoever. I think, and I hope, we're going to see a world where we can generate therapies that can treat those patients very effectively. Fundamentally, we will have life-changing, game-changing drugs—on a scale and at a pace that we've never seen before—getting to the right patient at the right time.

The promise of personalized medicine

Christoph Sandler: In the not-so-distant future, we might collect health data across different inputs: from wearables, from our electronic medical records, or from clinical or academic research. And we will have the opportunity, on a voluntary basis, to upload these data into a central, secure, trusted data storage system.

Lydia The: You could imagine using the data to figure out not just what drug might work for you but exactly what drug would work for you at what time, in what sequence, in what dose—really personalized to you.

Will AI replace scientists?

Lydia The: What we've found is that technology doesn't supplant the people. Rather, it will enable scientists to do things faster and better—and potentially develop insights that humans would not be able to develop at all.

Alex Devereson: Scientists will be able to discover things with machine learning that they could never have thought of by themselves, generate entirely new ideas, and move at a pace at which one person can do what it would have taken 100 to do before.

Lydia The: While in previous generations, scientists would spend a lot of their time—maybe even the majority of their time—on manual efforts such as pipetting from one tray to another tray or manually curating and cleaning data, I think AI will help us do all of those things in a more automated and quick way, and develop hypotheses that can then lead a scientist to think through, "What would the next experiment be? What are the implications of the data?"

What companies should do today

Alex Devereson: I think the challenge a lot of companies have is that they want to explore new ideas but, for good reason, they are reluctant to commit until they've seen some results and some tangible impact. So they explore a lot of pilots.

Lydia The: We have a term for that. We call it "pilot purgatory": companies focus on one pilot, and they see the returns in a single pilot, but they don't establish that approach and way of operating across their organization.

Define a 'North Star'

Lydia The: One of the most important things for companies to escape pilot purgatory is a real mindset shift, from the top end all the way through the rest of the organization.

Christoph Sandler: It is important for the organization to define what a North Star for them should be, so that the data and analytics transformation of the R&D function can be targeted toward that North Star.

Identify—and solve—the biggest pain points

Alex Devereson: Truly understand what your biggest scientific and operational pain points are. For the lab scientists, for the patients who are in your clinical trials, for the patients who might get the drug: What is the biggest unsolved problem today?

Embed analytics into decision making

Christoph Sandler: Bring the organization on board and help them understand the potential of data and analytics.

Alex Devereson: You need to really, truly redesign a process that embeds analytics, where it's not something on the side; it's truly a part of the decision making.

Christoph Sandler: And you need to establish trust in the data and in these models. Equally important as the technical part is the human part.

Deliver value quickly

Alex Devereson: Don't set up a project that delivers only after five years. Have a view on how you can deliver value in three months—and on what it takes in terms of analytics, data, and technology—with a relentless laser focus on value for the patient and for the scientific process.

Alex Devereson is a partner in McKinsey's London office; **Christoph Sandler** is a senior expert in the Tokyo office; and **Lydia The** is a partner in the Bay Area office.

Watch this and other *The Next Normal* videos on McKinsey.com.

Vineeta Agarwala

November 2022 | Interview

Vineeta Agarwala on the promise—and limits—of AI in drug discovery

By helping us learn from every single patient and every piece of data, AI could enable medical breakthroughs, says Andreessen Horowitz general partner Vineeta Agarwala.

'I hope you're learning from my journey.' It's a sentiment that many patients have expressed to Vineeta Agarwala, who agrees with it wholeheartedly—and believes that the healthcare and scientific communities are, indeed, starting to learn from every patient's journey. Much of the learning is made possible through AI, which has tremendous potential to aid scientists and improve patient outcomes by rapidly analyzing reams of data—even from failed experiments and inconclusive lab tests.

Agarwala is a general partner at the venture capital firm Andreessen Horowitz (also known as a16z), where she leads investments in companies that straddle the worlds of biology, health, and technology. The a16z Bio + Health portfolio now comprises some 50 companies. Agarwala, who is also a physician, sees genuine opportunity for AI to break new ground in drug discovery, as she recently told McKinsey's Lydia The during an interview in San Francisco. An edited transcript of their conversation follows.

Lydia The: You're a physician, a scientist, and a venture capitalist. That gives you a unique vantage point from which to observe AI's growing role in drug discovery and development. What gives you confidence that AI can revolutionize the field? And in which specific parts of the drug discovery and development cycle do you think AI could be most powerful?

Vineeta Agarwala: I think the intuition for the application of AI in drug discovery comes from nature. Nature has created a lot of great drugs, a lot of incredible protein machinery, a lot of very complex sequence-encoded instruction manuals. I mean, we have protein sequences in our bodies crawling DNA that figured out how to do transcription and translation and all kinds of mechanics.

To me, that says the design space is huge. The design space of proteins, nucleotide sequences, and small molecules that could have therapeutic effect is vast. At the very highest level, my reason for believing in the role and application of AI to drug discovery is just that we've barely tapped the level of complexity that nature has proven can exist.

AI and machine learning [ML] could plug into the drug development life cycle at every step. Learning systems can help humans conduct an end-to-end task better and faster at any stage: in target discovery, in the design of a modality, in the design of preclinical experimentation, in the clinical-development phase, in deciding which patients are most informative for your clinical trial, and so on. I think we will see this technology help us inch closer toward the frontier of what's possible.

One thing that excites me is the breadth of ML platform companies that are now being stood up across the industry—and they're all taking different approaches and learning from different data sets. An ML company in the small-molecule space is not the same as an ML company in the nucleic acid space or in the biologic space. There's a growing list of companies that are picking a very specific problem and getting very, very good at it.

What patients care about

Lydia The: What will the proliferation of ML companies mean for patients? What's most exciting to you about AI's potential impact on people who need drug therapies?

Vineeta Agarwala: The reality is that patients don't care how their drugs were developed, but they care intensely about the speed and the access properties of those medicines. What I get most excited about is the potential for AI to fundamentally change the timescales, as well as the cost structures, for development in the therapeutic space. Development costs do ultimately transmit to patients directly, and that's something that patients care a lot about.

We're unlikely to see a patient express pride and say, "Hey, I'm on an AI-guided drug!" But we are very likely to see patients—and patient communities that have been underserved by a lack of drug development—say, "Wow, new technologies have really created a surge of shots on goal for me that weren't possible before."

One thing that a lot of patients tell me—and tell their providers—is, "I really hope you're learning from my journey. I've done so many lab tests; I hope you're learning from all of them." That's something that I think the field is starting to get better at. We're starting to build systems around harnessing clinical data at scale, and real-world data, and using that as evidence. We're starting to get better at making sure that every time a patient walks in the door, their data journey is not exhaust. It's harnessed and captured, and we're going to learn from it, and we'll get better at diagnosing or treating the next patient or selecting the next patient for a trial.

The satisfaction of learning-inspired drug development is that almost nothing you do gets lost. Nothing you do is useless. The non-hits might be just as informative and important for your underlying model as the hits. That's a really satisfying regime to live in. It helps you want to generate data. It helps frame a lot of the problems you're solving as so-called type-two scientific questions, where either answer is still informative and important. So that's a theme that really inspires me about this space.

Vineeta Agarwala biography

Education
Earned both a PhD in biophysics and an MD from the Broad Institute of MIT and Harvard; holds a BS in biophysics from Stanford University

Career highlights
Andreessen Horowitz
(2020–present)
General partner

GV (Google Ventures)
(2015–19)
Venture partner

Flatiron Health
(2016–18)
Director of product management

McKinsey
(2007–08)
Business analyst

Fast facts
Did graduate work in computational biology and human genetics

Serves on a number of company boards, including BigHat Biosciences, GC Therapeutics, Memora Health, Orbital Therapeutics, Pearl Health, Thyme Care, and Waymark

Is board certified in internal medicine and continues to see patients at Stanford as an adjunct clinical professor in the Division of Primary Care and Population Health

Competed on the Stanford debate team

'A constant loop between the lab and the learning'

Lydia The: As you say, there is potential for AI across the entire discovery and development workflow. With that enormous scope and opportunity, how do you prioritize? How should a company—and the industry—think about which areas are ripe to start with first, versus areas where we still need more to get there?

Vineeta Agarwala: For a company, it is productive to ask the question, "Where do I have training data? Where have I run an experiment so many times that I might already have the data I need?" I would flag lead optimization as a part of the drug development cycle where large pharma companies do have a lot of data.

But taking a step back, where should the *industry* invest resources in machine learning? That's a much vaster question, which should be untethered from where there already is training data, because our capacity to generate new biological data and to profile patient samples is at an all-time high. I don't love the framework that says, "I have a lot of data, so let me throw it into a learning system and make a list of all my best insights and then hope I become a smarter actor going forward." Sometimes large companies gravitate to that framework because they're sitting on a lot of data.

More interesting, though, is to ask, "How can the learning system change the design of my next experiment such that I only gather the data that most incrementally improves my understanding of the problem at hand?" That becomes a big unlock because now you're designing the most informative experiment you can, using an underlying learning model. It's not a one-time learning effort. It's a constant loop between the lab and the learning.

Could AI make drug development cheaper?

Lydia The: The elephant in the room is the capital intensity required for a lot of these companies to develop machines and robots so they can automate their data collection. What's the advice you're giving to companies about investing up front in order to build the platform that allows for learning-enabled drug development?

Vineeta Agarwala: I often hear, "Well, to build an AI platform, you're going to have to burn millions on accumulating the training data needed to learn." I think it's a misconception that an AI platform needs to have a capital-intensive initial build phase. There are some settings where that will be true, but there are others where the AI platform can actually improve your spend efficiency and help you conduct just the right set of experiments that are most informative in helping you iteratively design a great therapeutic. The ultimate promise of AI-guided drug development is that you become more capital efficient because you're learning from every piece of data generated in your lab.

Part of the reason that AI-enabled business models could be so exciting is that the cash-out needed to conduct all the preclinical work required to even get a program into the clinic, and then ultimately through trials, was historically very steep. The promise, in principle, is that AI-first technologies could fundamentally shift that cost curve such that your wall clock time and your cash requirement to nominate a program to advance into clinical trials could be blunted. We're spending, like, $1.5 billion per drug that we get approved at the moment, so there's a lot of room for improvement.

Efficiency and efficacy

Lydia The: We've built a case around AI, and now I feel like we need to address the skeptics. What should people look for to be convinced? Is it, "We have demonstrated that we were able to increase PTS [probability to success]" or "We identified a target we would have never been able to identify just using manual experiments"? Are there goalposts that you're looking for?

Vineeta Agarwala: The proof points for AI-guided drug development, in my mind, fall broadly into two buckets. One is around efficiency of drug development. The other is around properties like efficacy and safety.

In the category of efficiency, there's now starting to be some literature on this topic: For a company that said it was leveraging AI in its drug development process, how long did it take to get to an IND [investigational new drug] application? There's so much variation in the baseline that it's hard to do these analyses, but there's early data suggesting that those timelines have shrunk—to three to five years, versus historically six to ten years. That's an important goalpost in the efficiency domain. Also, can you burn less capital in getting to an IND? We're very keen to back companies that can say, "We can stretch it. We can take more shots on goal with the same capital investment."

For what it's worth, I think it's going to be really hard to figure out what program AI touched and what program AI never touched. I don't think it's black and white. But if you were to try to separate programs into more AI-guided versus not, these are the metrics that matter: PTS, time to IND, and capital expenditure per IND.

The other important bucket is efficacy at tolerable safety profiles. That's the mantra of our industry, and if AI can help us get that more frequently, more quickly, and at lower cost, that would be huge: Have you been able to unlock development goals that were previously not

possible? Could you drug a target because you found a pocket that no one else could see? Could you design an antibody to be delivered in a particular formulation that was really important for patients that couldn't have been designed without AI-guided insights? These are all examples of fundamentally new—and very exciting—unlocks in the industry.

AI won't be a 'catchall' solution

Lydia The: What do you think people get wrong about applying AI to drug development?

Vineeta Agarwala: I think what people get wrong is to assume that AI is a catchall silver bullet for problems that are hard. Those problems will still be hard. There are vexing problems of target biology, of interconnected pathways that we don't understand yet, of preclinical models not being fully predictive of human biology.

AI will create insights that make those problems a little bit more tractable, more scalable, more engineerable. But many of the problems won't be solved by AI alone. They will continue to be solved by really creative molecular biologists who come up with new assays, or through great synthetic solutions that help us generate a lot of diversity in different spaces, or through thorough preclinical studies. So you can't assume that there's an AI solution for every hard problem. Most likely, it's going to be a *scientific* solution that might benefit from AI.

My bet is that learning platforms and technologies that help us learn from large amounts of data are going to be ubiquitous. I hope that, some years from now, people won't just try to pluck software modules and incorporate them into a specific problem set, but rather that there will be a much more integrated experimental, computational, and clinical development regime in which the goal is to learn from the ever-mounting biological data that we're collecting.

Vineeta Agarwala is a general partner at the venture capital firm Andreessen Horowitz. **Lydia The** is a partner in McKinsey's Bay Area office.

The authors wish to thank Alex Devereson, Christian Fougner, Olivier Leclerc, and Christoph Sandler for their contributions to this article.

For more from Vineeta Agarwala, see the videos accompanying this article on McKinsey.com.

November 2022 | Interview

‘We can invent new biology’: Molly Gibson on the power of AI

Molly Gibson

The cofounder of Generate Biomedicines, a pioneering new drug development platform, describes how AI and machine learning are transforming the way we discover new medicines.

Some diseases have confounded researchers for decades. But the application of artificial intelligence (AI) to medicine promises to drastically accelerate the discovery of new drugs to treat them. Tackling hard-to-crack challenges and reaching more patients excites Molly Gibson, cofounder and chief strategy and innovation officer at Generate Biomedicines. Based in the Boston area, Generate is a therapeutics company launched in 2018.

Gibson recently spoke with McKinsey’s Christian Fougner and Lydia The about how AI and machine learning can speed up drug discovery—and what it could mean for patients everywhere. An edited version of their conversation follows.

Lydia The: You were part of several biotech firms before you cofounded Generate Biomedicines. Tell us about Generate’s potential for revolutionizing medicine.

Molly Gibson: Generate is focused on transforming drug discovery in the protein therapeutics space. Instead of thinking about how to discover new molecules from natural processes and evolution, we use computers to learn the rules of proteins and how proteins function. We can then generate completely *de novo* molecules that do the things that we want so that we can create better, cheaper, safer drugs.

When we first started thinking about the idea of Generate, we drew upon examples of different fields that have been revolutionized by the idea that a computer can learn principles of systems, then take those principles and invent new and novel things that never have existed before. An example is in the field of computer vision. A computer can look at hundreds of millions of pictures of human faces and videos of people talking and ultimately learn the principles by which a human face comes together. It can then create images or videos of people that look realistic, like you or me, but don’t exist and statistically never will exist.

‘What if, in the future, we actually know the connection between a DNA sequence and a protein that not only elicits an immune response but treats a disease? We could completely change the lives of patients.’

We’re excited about applying that to the world of proteins and biologics and therapeutics. We’ve tried to predict biology for decades. But what if you could actually invent new biology? The algorithms can get better over time and are constantly learning. You could then start to *program* therapeutics versus *discover* them. Nature has given us this huge set of beautiful proteins to learn from, and it’s given us the rules that have been discovered through evolution. We can learn those rules, extract from them, and generate proteins that have never existed before.

‘More shots on goal’

Lydia The: We know that drug development has a 90 percent failure rate. Why do you believe that AI could improve on this?

Molly Gibson: That’s an incredibly daunting number, and one that I feel we, as an industry, should be obsessed about solving. It’s a complex problem with lots of steps—from selecting the target, to creating the molecule, to identifying the patient populations. We’re seeing AI have an impact across all spectrums of that complex equation. AI will give us an improved understanding of biology, so we’ll be able to identify better targets and then create molecules that better intervene in that biology and are safer for patients, as well as identify the patients who are going to be most affected by those medicines in disease populations.

We’re going to see an inflection point in our ability to learn and optimize for molecules and diseases that we know. We know how to affect the disease, but we just don’t have the tools to create the medicines that we want at the speed that we want. My hope is that we’re going to see an acceleration of molecules that are invented by computers, and that, for those diseases where we understand the biology, we’re going to find ways to treat them more quickly.

One recent example of how this could look in the future is what we did for the development of a vaccine for SARS-CoV-2. We were able to program mRNA molecules to express the specific antigen that our immune system responded to, in order to create protection against future infection. We understood enough about the connection between an mRNA molecule and the protein structure of that antigen. But what if, in the future, we actually know the connection between a DNA sequence and a protein that not only elicits an immune response but treats a disease? We could completely change the lives of patients. Our ability to program biology is going to change the pace, speed, and cost at which patients get new medicines.

Molly Gibson biography

Education
Earned PhD in computational and systems biology from Washington University in St. Louis in the Center for Genome Sciences & Systems Biology; received BS in computer science from Truman State University

Career highlights
Flagship Pioneering
(2017–present)
Senior principal

Generate Biomedicines
(2018–present)
Cofounder and chief innovation officer

Tessera Therapeutics
(2018–present)
Founding team member

Kaleido Biosciences
(2015–18)
Scientific strategy and operations executive
Senior scientist, computational biology

Fast facts
Has published in multiple journals, including *Science* and *Nature*

Has filed for multiple pending patents

Appeared on *Business Insider*'s 2021 list of 12 young serial entrepreneurs building the next generation of biotech start-ups, and *Endpoints News*'s 2020 20 Under 40 list in biopharma

While working on her PhD, collaborated with St. Louis Children's Hospital to study the effects of early life interventions on development of the preterm infant gut microbiome and health outcomes

Previously served as director and board member of the Young Scientist Program at Washington University in St. Louis, leading more than 100 active volunteer scientists with the mission of bringing inquiry-based science to disadvantaged K–12 students

The combination of all of that is going to start to chip away at the failure rate [for new drugs] that we see today. We're going to go from a world where we see one or two therapeutics a year come out to a world where you could imagine a tenfold jump—where we could see ten to 20 therapeutics, if not more, that are ready to go into clinical development every year.

Lydia The: The biopharmaceutical industry is focused on a handful of diseases or on small patient populations. People always say it's hard to move beyond that in ways that are cost-effective or that apply to infectious diseases that mostly affect the developing world. If we think that AI's going to make drug development faster and cheaper, could that also mean it will address a broader reach of patients?

Molly Gibson: Yes. I'm really excited about the opportunity for AI to change the economic equation for drug development. We should see a complete change in priorities. We'll be able to pursue different types of patient populations: those in developing countries or places where clinical trials are higher risk or cost more. If you can bring the drug development costs and time down significantly, you have more shots on goal, and you can do new and novel things that just haven't been possible with models in the past.

Not just faster, cheaper, and better

Lydia The: What do you think the lab of the future will look like? What types of work will labs be focused on?

Molly Gibson: This is an exciting topic. When we think about what type of data we need to generate for AI, it's not going to look like the types of experiments that you've done before. We're going to constantly be thinking about how to miniaturize experiments and how you make them high throughput in a completely different way so that you constantly have data feeding the algorithms, and you're learning from every protein you generate.

One of the things that we've invested heavily in at Generate is fabricating our own microfluidics devices to be able to assay proteins with high throughput rates that traditional techniques just can't do. A lot of times, people think about layering AI on top of how we already do things. What gets me more excited is the idea that AI gets to transform *what* we do. It's not just about doing the things that we already do faster, cheaper, and better. It's about being able to do things that we weren't able to do before.

Lydia The: This seems like it's going to be very expensive and require a lot of data. How will it affect larger, well-funded biotech companies or institutions, versus smaller upstarts or researchers at a university that may not have the same resources?

Molly Gibson: There's going to be a play across all these different sizes of companies, from the large companies that can afford to generate large amounts of data to start-ups. As much as I believe that there's going to be a huge advantage to having data, I still think human ingenuity will have an important role in innovation. How you approach the problem and how you execute upon that problem will be important.

We'll also likely see more and more public data sets being generated for start-ups, and people with new ideas and benchmarks to test new algorithms against. And with all those new ideas, it's execution—your ability to turn those ideas into reality—that will make all the difference.

Of course, the data that we have access to is going to dictate the types of problems that we solve and the areas in which AI and machine learning will make the biggest impact the earliest. A good example of this is DeepMind.[1] They have been able to make enormous strides and solve the protein-folding problem—which had been a holy grail in the field for decades—because there is just such a well-curated data set in protein structure. That's been a key part of their ability to innovate.

Mastering complexity

Lydia The: What do you think are the main barriers to realizing the vision of AI in changing the cost structure and ability to develop drugs?

Molly Gibson: One of the challenges will be to create data sets without bias. This is a huge area of focus across biology and in patient data. I saw recently that most of the medical data that have been used to train AI algorithms came from three states: California, Massachusetts, and New York. So we're only seeing a very small portion of the population. Different types of bias creep into all different types of data sets, so it'll be very important to make sure that the data sets that everybody is using to develop new models are fair and representative.

[1] Over the past six years, Google's DeepMind has developed AI software called AlphaFold, a deep-learning system that can predict a protein's 3-D structure from its amino acid sequence. In July 2022, DeepMind released predicted structures for nearly all known catalogued proteins, increasing its AlphaFold database to more than 200 million structures; *Research Blog*, "AlphaFold reveals the structure of the protein universe," blog entry by Demis Hassabis, July 28, 2022.

But the biggest barrier, by far, is that biology is hard. Human biology is incredibly complex. Integrating our understanding of biology with engineering, translating that into an ability to create novel molecules—bringing it all together will be very challenging. And moving into more complex diseases, expanding our reach to the most in-need patient populations, understanding disease at an earlier state so we can treat or reverse disease before it even starts—mastering all of that complexity and being able to do these impactful things is going to be really hard.

But it's such an exciting opportunity. The future is not going to look like the past. The way that we discover drugs will not look like it did ten years ago. What we have done in drug discovery historically is going to look incredibly analog to the way that we discover drugs in the future.

Molly Gibson is the cofounder and chief innovation officer at Generate Biomedicines. This interview was conducted by **Christian Fougner,** a consultant in McKinsey's New York office, and **Lydia The,** a partner in the Bay Area office.

The authors wish to thank Alex Devereson, Olivier Leclerc, and Christoph Sandler for their contributions to this article.

Comments and opinions expressed by interviewees are their own and do not represent or reflect the opinions, policies, or positions of McKinsey & Company or have its endorsement.

For more from Molly Gibson, see the videos accompanying this article on McKinsey.com.

Daphne Koller

November 2022 | Interview

‘It will be a paradigm shift’: Daphne Koller on machine learning in drug discovery

We’re entering a “new era of science”—we finally have enough data and technology to truly enable better drugs for patients, says insitro CEO Daphne Koller.

Daphne Koller has a knack for using technology to improve the human condition. She’s won some of computing’s highest awards and been at the center of a few of Silicon Valley’s efforts to improve lives—one example being Coursera, the global online learning platform she cofounded. A recipient of the MacArthur Foundation’s “genius grant” and one of *Time* magazine’s Most Influential People, Koller is a leading authority on machine learning. Her current mission, as the founder and CEO of drug discovery and development company insitro, is to harness the power of machine learning to create better medicines for patients in need.

Koller recently spoke with McKinsey’s Lydia The at insitro’s headquarters in South San Francisco. Excerpts of their conversation follow.

Lydia The: Let’s start with a big-picture question: How do you think AI and ML [machine learning] can change drug discovery? How would you describe the opportunity?

Daphne Koller: Drug discovery in the past 50 years is a tale of glass half full and glass half empty. On the half-full side, we have transformative medicines that have made a very big difference to patients. On the half-empty side is the so-called Eroom’s Law, the reverse of Moore’s Law, where the cost of drug discovery has grown exponentially year on year without an increase in new drug output.

Why is that? It’s because there are multiple places in the drug discovery process where we need to make significant decisions. If we’re lucky, one decision will get us to a good outcome, but the rest will get us to a dead end. Every successful drug has to bear on its back the cost of all the failures. That means for many diseases there’s just no drug—because either it’s never been a priority and no one’s working on it, or people are working on it but haven’t figured out a path to create an effective medicine. Disease-modifying medicines are few and far between, and cures are almost nonexistent.

With AI, we're able to use large amounts of data to build "compasses" that allow us to know, when we get to these forks in the road, which path will most likely lead to success. We aspire to create a much more engineered process, with a higher success rate. The aim is to go faster from identifying the genetics of a disease, or of a group of patients, to developing a disease-modifying intervention—so that maybe, when we get to 2035, there will be a lot more treatments to help patients live a long and healthy life.

Lydia The: How exactly is insitro tackling that? What's the first angle you're taking?

Daphne Koller: Our focus is "de-convoluting" the biology of human disease. Often, clinicians tackle disease without really understanding what the disease even is. Disease is often defined by coarse-grained symptomatic manifestations, some of which use classifications that date back 50 years or more. These are typically filtered through a subjective lens of both the patient and the clinician, so we end up with a mishmash that really doesn't speak to the underlying biological causes of the disease.

At insitro, we collect high-content data to help us understand underlying biological processes that correspond to disease. Some of those data sets come from patients. For example, we collect imaging data, such as MRI and histopathology; various molecular measurements; and other data that allow us to identify, via machine learning, subtle patterns to disentangle distinct patient subsets.

At the same time, we generate in our lab large amounts of human-derived cells, called induced pluripotent stem cells. These are human cells that were reverted to stem cell status, from which we then create neurons or hepatocytes that carry our genetics. We can further introduce into those cells genetic variations that we know are likely to cause disease. Then we can measure those cells and interrogate—with microscopy or RNA sequencing—what disease looks like at the cellular level. This system gives us a rapid approach for testing therapeutic interventions that could potentially work in humans.

'Let machines loose'

Lydia The: What role do AI and ML play in this process? In other words, what is AI doing that a scientist or researcher can't do?

Daphne Koller: I'll give you a couple of examples. In our recent work studying a fatty liver disease, we were able to identify—using ML—patterns within the liver tissue that correspond with known genetic drivers of disease. Human pathologists couldn't see those patterns because they don't even know what to look for. We found that if we let ML loose on the samples—if we let machines have an unfettered, unanchored look at the data—they're able to identify disease-causing and disease-modifying associations that a human just can't see.

Another example is our work on tuberous sclerosis complex, which is a rare but not ultra-rare disease: there are 50,000 patients diagnosed with it in the United States, a million worldwide, and we believe it's underdiagnosed. We created an in vitro cellular disease model by introducing the genetic variant that causes the disease into our cellular systems via CRISPR, and then we were able to phenotype those cells using different methods—including some live cell imaging via our proprietary ML-enabled microscope. We were able to demonstrate reversions that had never been identified before. We're now assessing those as potential novel drug targets.

Daphne Koller biography

Education
Hebrew University of Jerusalem, BSc and MSc
Stanford University, PhD in computer science
University of California, Berkeley, Postdoctoral researcher, Computer Science Division

Career highlights
insitro
Founder and CEO (2018–present)

Coursera
Co-chair (2016–19)
Founder and president (2014–16)
Founder and co-CEO (2012–14)

Calico Labs
Chief computing officer (2016–18)

Stanford University
Professor (1995–2014)

Fast facts
Cofounded Engageli, an online learning platform, in 2020

Was named one of *Fast Company*'s Most Creative People in Business (2014), *Time* magazine's 100 Most Influential People (2013), and *Newsweek*'s 10 Most Influential People (2010)

Is an elected member of the American Academy of Arts and Sciences, the National Academy of Engineering, and the International Society for Computational Biology

Received the MacArthur Foundation Fellowship in 2004

Has won many awards, including the ACM Prize in Computing in 2008 and the Presidential Early Career Award for Scientists and Engineers in 1999

Authored more than 300 peer-reviewed publications in journals and venues, such as *Science, Cell, Nature Genetics,* NeurIPS, and the International Conference on Machine Learning, with an h-index of over 145

Ensuring data integrity

Lydia The: Good data is critical to AI and ML. A question in my mind about AI and ML in drug discovery is, "Why now?" What makes you believe that we can now get enough data that's fit for purpose for AI applications? And how do you ensure you have good data coming out of your models?

Daphne Koller: One of the things that led me to come back to this field after a bit of a digression into online education at Coursera is that I felt like now is a time when we can really make a difference in applying machine learning to biomedical data. When I was at Stanford, a large data set was 200 samples. You felt lucky if you had 500. Now we're in a world where there's an unbelievable ability to both access and generate data that is fit for purpose for machine learning.

On human data, one of the earliest efforts is the UK Biobank, which has been able to create deep phenotypic data from 500,000 individuals—measuring everything from whole body imaging, brain imaging, blood biomarkers, urine biomarkers, predisposing factors, and so on, as well as longitudinal outcomes. It has limitations, of course, not least of which is that its composition is very Eurocentric, but—both in the UK and elsewhere—others are building on this effort and creating additional cohorts, making it more diverse. These data sets are only going to get bigger and more useful due to the growing availability of electronic health records.

Separately, in the last decade or so, there have been major advancements in life science tools like CRISPR—which is a therapeutic intervention but is at least as powerful as a research tool—and in measurement technology, with things like super-resolution microscopy, single-cell RNA sequencing, and single-cell proteomics. All of these enable the creation of incredibly large data sets that allow us to interrogate, in very fine detail, the underlying biology of disease.

Lydia The: How do you think about the balance between using publicly available data—which exists but isn't always clean or easy to use and doesn't necessarily confer competitive advantage—and generating your own data?

Daphne Koller: Many people believe that by simply collecting a bunch of data haphazardly from different places and creating a sufficiently big pile, you'll have something that is fit for purpose for machine learning. That's very rarely the case, especially because some of that data comes from small experiments that were all done in a different way, in a different assay, under different conditions, with different definitions of what success looks like. That strategy is very dangerous, especially when mistakes are made at the early stage and you discover the mistake five years later in a very expensive clinical trial.

There are high-quality data sets available to everyone—not many, but they do exist. As I mentioned, the UK Biobank is an example. So we've onboarded data sets that we think add value to machine learning. There are novel methods that we can apply to these data sets that give us unique differentiated insights.

Now, is that a permanent competitive advantage? Probably not, because there are a lot of smart people out there. But it certainly gives us a head start.

Our bigger advantage is that we generate, in-house, complementary forms of data that align with what's available publicly but allow us to do experiments. Our data set allows us to intervene and assess causality of a disease variant or of an intervention, so it's a huge amplifier to what is available in public data sets. Using both, the whole is considerably greater than the sum of the parts.

A spirit of collaboration

Lydia The: Something I think about a lot is talent and culture. At insitro, you're mainly looking for two different types of talent: people with a biology background and people with a computer science background. Everyone talks about how hard it is to avoid silos and how challenging it can be to get everyone working together and understanding one another. How do you make sure that you don't create second-class citizens and that your employees truly see one another as equals?

Daphne Koller: That's one of my favorite topics. One of the things I'm most proud of is the way in which insitro has brought together people with diverse backgrounds: machine learning scientists, software engineers, automation engineers, stem cell scientists, discovery biologists, drug hunters, and more.

We've laid out behavioral norms from the very beginning to make sure that there's collaboration and team spirit within our company. We expect employees to engage with one another openly, which means you are willing to ask "naive" questions and accept "naive" ideas from people outside your discipline; constructively, which means you seek to make the outcome better rather than trying to be the smartest person in the room; and with respect for what everyone brings to the table.

This spirit of collaboration permeates insitro and is something that every new "insitrocyte" comments on when I have my 60-day meeting with them. And we've consistently found that it gives rise to not only better solutions but also better problems. Questions that we would never have thought to tackle just emerge when people with different backgrounds come together and say, "What is it that we're really trying to do here?"

Lydia The: On that point and looking forward, what can we reasonably hope for in this space? There's a wide range of opinions about the potential of AI and ML in drug discovery. What's your take?

Daphne Koller: I find that there's an interesting bimodal distribution of opinions regarding the role that machine learning could play in drug discovery. Just four to five years ago, skeptics thought AI was going to be completely useless. More recently, there's been a greater recognition of the value, but there are still people who think it will be a point solution, like combinatorial chemistry—something that will help in a narrow niche but won't have broader impact.

On the other side, there are people who are starry-eyed true believers who say, "This is going to be artificial *general* intelligence!" and who believe they're going to "get every drug approved within six months!" Biology is really hard, and we need to be very careful when intervening in something as precious as human life. If people tell you that they're going to have "100 drugs in the clinic in three years" and have predictions along those lines, that too is wrong, in a different way.

Where I think we'll be in 15 years is that machine learning will be an absolutely critical, pivotal shift—a paradigm shift in how we discover and develop medicines. It's going to be more akin to computers as a tool than it is to combinatorial chemistry as a tool, in the sense that it will touch every single facet of how we discover and develop medicines, and accelerate and improve every single one of them.

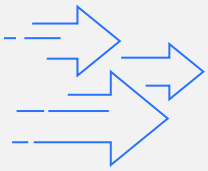
Lydia The is a partner in McKinsey's Bay Area office.

The author wishes to thank Alex Devereson, Christian Fougner, Olivier Leclerc, and Christoph Sandler for their contributions to this article.

Comments and opinions expressed by interviewees are their own and do not represent or reflect the opinions, policies, or positions of McKinsey & Company or have its endorsement.

For more from Daphne Koller, see the videos accompanying this article on McKinsey.com.

Related reading



For more on the future of biotech, see these articles on McKinsey.com.

AI in biopharma research: A time to focus and scale

October 2022

Self-learning: The dawn of a new biomedical R&D paradigm

July 2022

What are the biotech investment themes that will shape the industry?

June 2022

Transforming biopharma R&D at scale

May 2022

Better data for better therapies: The case for building health data platforms

April 2022

Generating real-world evidence at scale using advanced analytics

March 2022

Ten battlegrounds for digital and analytics in life sciences

August 2020

Creating value from next-generation real-world evidence

July 2020

Designed by McKinsey Global Publishing