BIOPHARMA THOUGHT LEADERS

HOW AI IS ACCELERATING AND TRANSFORMING DRUG DISCOVERY

Leonard Lee, head of growth and customer success for Accelerated Discovery at IBM, discusses the ways AI is transforming drug discovery and assisting scientists today.

Artificial intelligence (AI) in drug discovery is moving beyond the proof-of-concept stage. Advanced machine learning (ML) techniques are powering technologies that slash the time needed to value and enable exploration of uncharted space. Researchers are moving into a new era of AI-enabled drug discovery that will accelerate the development of life-changing medicines.

What is IBM's Accelerated Discovery initiative?

The focus of this group is to build new applications and tools to help researchers pursuing molecular discovery. It's really about how technologies can deliver impact on complex global challenges in the physical world. For example, how can we significantly compress the timeline and cost of developing new drugs? To do that, we're focused on fundamental building blocks—atoms, molecules, proteins—and the discovery of new materials that solve problems currently intractable in the health, materials, climate, and sustainability domains. Our ethos and focus are best summed up in the tagline 'matter that matters'.

We're harnessing the power of IBM Research—a history of cutting-edge computing innovation and our knowledge of AI, combined with a deep understanding of material science. Accelerated Discovery builds on that heritage by leveraging AI for knowledge ingestion at scale, enriched simulation to augment gaps in human understanding, automated hypothesis generation, and cloud-based experimentation—to accelerate the entire discovery pipeline.

Where do you see the biggest opportunities today?

Right now, there are four areas we're actively pursuing that we believe can have a lot of impact within drug discovery: scientific- and chemistry-knowledge integration; AI-enhanced scientific simulations; generative AI for molecular-lead generation; and AI for retrosynthesis planning.

We're trying to both speed up innovation and expand the window for discovery with these collective capabilities.

Could you talk more about the four priority areas, starting with generative AI?

On the generative front, we're using vast datasets representing the structure of molecules to make predictions of new molecular design-based constraints like the ones researchers apply to their discovery

processes. Think of it as reverse engineering. We're asking, given a specific set of requirements, how do I create a molecule that fits what I'm looking for?

We're using AI to generate novel hypotheses for exploration and to create new molecules that domain experts may not come up with on their own, expanding their discovery space.

What about using AI in the chemical synthesis stage?

The drug discovery process must consider what is synthetically feasible and what is the best way to synthesize a candidate molecule. Even prior to the language-model revolution, we have been building language models for chemical synthesis, trained with over 3 million chemical reactions, to learn the fundamentals of how organic molecules react. Based on that, we can start to predict how to synthesize something and create step-by-step recipes for how to make a molecule in a manual or automated lab. This formed the basis for our RXN for Chemistry tool.

We've actually automated the execution of these guiding recipes at our robotic AI lab in Zürich. The lab, which knows what reagents it can access, uses the language model-based approach to recommend different ways to synthesize a molecule, and can infer the detailed sequence of steps to then synthesize the molecule.

> **One example comes from researchers at the University of Oxford, who we worked with to use our generative AI and surrogate simulations to look for antimicrobial candidates**
>
> Leonard Lee
>
> Head of Growth and Customer Success for Accelerated Discovery
>
> IBM

With RXN for Chemistry itself, we're starting to dig deeper with partners, incorporating their crucial domain expertise to go beyond simple demonstrations and apply it to real-world challenges facing drug discovery teams.

Why are AI-enhanced scientific simulations important?

In molecular simulations, a lot of computation is required, and extremely long execution times can vastly limit the aperture of discovery. One of the things we've learned to use is what we call 'deep-learning surrogates'. Essentially, AI-powered shortcuts replace expensive physics-based evaluations with rapid data-driven ones. This allows a researcher to run many more simulations with the same computational budget and quickly test one idea after another to validate or reject a hypothesis.

In addition to making simulations faster, we use AI to make smarter decisions about what to run, and how. We have developed a methodology based on the principles of Bayesian optimization to balance knowledge generation and knowledge exploitation effectively across a range of techniques. By understanding what our AI models know, and don't know, this system makes decisions about which molecules to simulate, and what kind of simulations to run, ensuring that the overall budget is used efficiently[1].

You mentioned knowledge ingestion at scale. Can you explain how you help researchers with that?

Our goal is to extract useful information from existing scientific literature or datasets available within research-focused organizations. We've created a capability to take massive amounts of unstructured data, including tables, and make sense of them by organizing them and extracting relationships, so that someone can start to explore the vast repository of available knowledge. To use the tool, known as Deep Search, for drug discovery, we've ingested drug patents, journal articles, and data from millions of documents and sources, and we can start to use that corpus of information to help researchers.

If you're looking at a molecule, you can learn whether anyone has ever discussed it and what interesting side effects or related targets they have explored. All that knowledge brought into the picture together can arm researchers with invaluable insights.

The full value of Deep Search is becoming clear now that we're working with lots of companies to help them understand and organize their own data. These companies are feeding in private documents as well, so that their employees can assess all public information together with their own proprietary data when researching a molecule.

How can researchers access these technologies today?

These are all available to the public to experiment with. RXN for Chemistry and Deep Search are accessed via the web. Our Generative AI and simulations toolkits have been released as open-source technologies that people can download and start to use, while we work on the best way for them to be consumed at scale and more broadly. We'll get a clear picture of the best form factor as we work with customers to understand how they use these capabilities.

How do you lower the barriers to use?

That's why our team exists. It's about bridging the gap between the cutting edge of AI/ML and the capabilities of people working in life sciences. In AI/ML, there's a lot of brilliance, there are a lot of new techniques that come out, but it's not always as accessible or consumable as it needs to be for widespread adoption.

We're here to figure out how to translate these technologies into actual applications and tools that people can easily use. We've developed graphical user interfaces and application programming interface (API) calls that abstract away a lot of the AI techniques. Our guiding principle is that AI expertise should not be required from users. Pre-training the models, configuring the models, and building training pipelines to be able to tune the models easily is all built in. It eliminates the need for the kinds of special skills required to use some of the other libraries that are on the market. That barrier is too restrictive. Very few people are experts in both AI/ML and drug discovery. We want to put these tools into the hands of all researchers. At the same time, there's enough domain knowledge that we can guide them and understand how to insert a new technology into existing workflows.
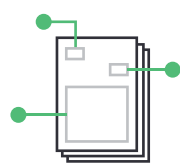
Deep Search

Knowledge extraction using AI for text and image understanding

Teams of scientists used Deep Search to extract knowledge from the COVID-19 Open Research Dataset

Generative Toolkit for Scientific Discovery

Accelerated hypothesis-generation for molecular research

Researchers from IBM and the University of Oxford used generative modeling for de novo design of antimicrobial therapeutics in only 48 days

Simulation Toolkit for Scientific Discovery

AI-enhanced molecular simulation for virtual experiments

IBM Research demonstrated advanced simulation techniques offer a 30–40x speedup in the discovery of anti-malarial drugs

RXN for Chemistry

An AI lab partner for chemical synthesis

IBM Researchers used RXN to shortlist candidates and narrow chemical recipes to use for the creation of new photoresists

Leverage AI tools to accelerate the discovery process. Transform traditional trial-and-error experimentation or rules-based modelling with tools that see beyond existing knowledge and require less manual processing of scientific information and data, for 10x faster discovery.

We are asking how to make these technologies something that thousands of people can use to improve their workflows today. That's why it's exciting. We're applying this to the field at scale.

What impact will these technologies have?

We're already seeing the impact. In practice, tools like Deep Search have made ingestion, structuring, and reasoning with scientific literature 1,000 times faster; simulations with AI surrogates can be between two and 40 times faster; AI-generative models can be 10 times faster at coming up with novel hypotheses; and AI-driven autonomous labs have run experiments 100 times faster.

Those improvements cover each step in the drug discovery process. Given how long the process takes today, how much it costs, and how many clinical candidates fail, there's huge potential for improvement. There's now a clear path to achieving those improvements.

Do you have any real-world examples of what is possible?

Yes, we're already engaging organizations and scientists to deploy the technologies and seeing the results. One example comes from researchers at the University of Oxford, who we worked with to use our generative AI and surrogate simulations to look for antimicrobial candidates[2].

Using generative AI, they looked for molecules with specific antimicrobial functions and minimal toxicity. They then ran the candidates through simulations to understand which molecules were feasible and could be synthesized. In 48 days, the researchers identified, synthesized, and experimentally tested 20 candidates, two of which were highly potent and displayed low toxicity in mice.

The 10% success rate, compared to numbers that can fall below 1%, and the 48-day turnaround time represent vast improvements over standard discovery processes for de novo therapeutic-molecule design. This is what we want to bring to researchers. Scaling access to these technologies will provide a step change in drug discovery timelines.

What's next for Accelerated Discovery?

We're working at a rapidly advancing scientific frontier. A big part of our focus will be to keep turning the latest AI innovations into usable products. As we do, our offerings will get better and better at helping researchers accelerate drug discovery.

As more researchers adopt the technologies, we'll enter a fundamentally different era of drug discovery. Until now, most work has built incrementally on other projects, deviating a little to expand the space but remaining tied to prior knowledge; venturing outside of the known space meant running around in the dark, hoping to catch something.

AI changes that. Now, the exploration space is essentially unlimited, and we have a way to navigate it. There are so many unmapped areas, and advanced technology is essential to help us take a bigger step to explore the possibilities. Whereas most existing drugs are, on some level, sequels to earlier projects, we can now create wholly new molecules, untethered from existing knowledge. I'm excited to see how our life-science partners will use the capabilities.

1. Fare, C. et al. *npj Comput. Mater.* **8**, 257 (2022).
2. Das, P. et al. *Nat. Biomed. Eng.* **5**, 613–623 (2021).