Research

The Era of Accelerated Discovery

1.

Discovery Today: *Slow, Costly & Inefficient*

All it took was a ball of discarded Scotch tape in the waste basket.

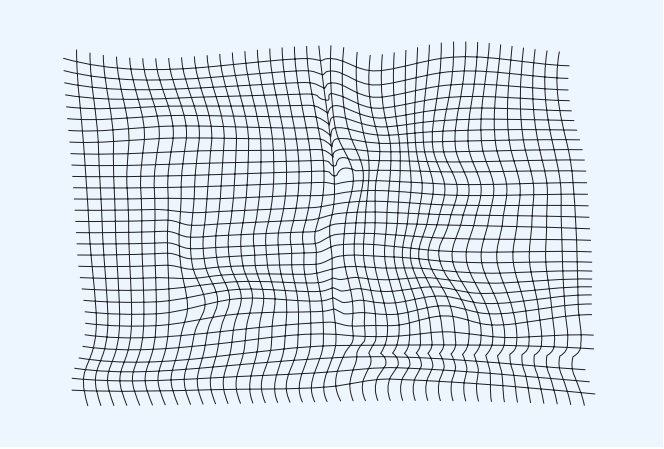
There was residue on the tape's sticky side—the result of researchers at the University of Manchester cleaning up graphite, the material in pencil lead. Curious, physics professor Andre Geim examined the residue under the microscope, then folded the tape in half and pulled it apart. The graphite flakes were thinner than any he'd ever seen. Geim had discovered graphene, an atom-thick layer of carbon—the thinnest, yet the strongest material known. The finding earned him and his colleague, physicist Kostya Novoselov, the 2010 Nobel Prize in Physics. The discovery was serendipitous—but just like many other accidental material discoveries, it could significantly impact our world, leading to flexible circuits and unbreakable touchscreens, more powerful gadgets, improved bulletproof vests and more efficient medical, chemical and industrial processes.

Graphene was discovered totally by chance in 2004. It is an extremely efficient heat and electricity conductor, just one atom thick—one million times thinner than the diameter of a human hair—yet about 100 times stronger than the strongest steel of such thickness.

New materials, and how we create them, have always been crucial to progress. They have taken humans to the Moon and to the bottom of the ocean, revolutionized science, healthcare, manufacturing, defense, biotechnology, communications—and nearly every other industry. But the process of materials discovery is incredibly hard and complex even today, in the age of material science. This is mainly because the chemical space of potential materials is so vast—there are more possible combinations for a new molecule than there are atoms in the universe.

Serendipity is how many important materials were discovered throughout history, including Teflon, Velcro, Vaseline and vulcanized rubber. Those that were not happy accidents were often discovered by trial and error, during a slow and iterative process of mixing different compounds and testing the result. That is how the ancient Greeks, for instance, learned to make concrete while mixing lime, water and crushed potshards. In the 19th century the approach became more scientific, with chemists starting to synthesize substances and adjusting experimental conditions to produce new versions of a

The demands on materials keep growing. We need them to store energy more efficiently and to feed the world's surging population.



material until they got the properties they needed.

Discovering new materials has helped our world progress—sometimes in small incremental hops, other times in giant leaps. Consider plastics. In 1869, American inventor John Wesley Hyatt, eager to win $10,000 promised to anyone who could come up with a substitute for ivory, discovered how to make synthetic polymers. The invention changed the way we live.

The plastic revolution that Hyatt had started made a significant leap forward in 1907 when Belgian chemist Leo Baekeland invented the first fully synthetic polymer (which he named after himself, 'bakelite') and coined the term 'plastics'. The subsequent development of polymers occupied researchers worldwide, and by the 1950s, materials science departments began to appear at universities.

A new progress jump followed when computers joined in, greatly accelerating the materials discovery process. Knowledge of a material's properties no longer had to be passed on from the master to the apprentice, but began to be encoded in early predictive models, enabling researchers to simulate the performance of a material. Machines replaced an artisan's—and later augmented a chemist's—understanding of matter and its characteristics, and the rate of materials discovery surged dramatically.

Nevertheless, discovering and designing materials using the time-consuming trial and error method, even when assisted by computation, has not been trivial. To come up with a new polymer, the first step is to set its desired characteristics. Then begins the tedious search for the correct molecule of the plastic's constituent parts, the monomers. There are over a hundred different options for molecules that could be potentially suitable and an infinite number of ways to synthesize the polymer resin—with variations in monomer type, composition, molecular weight and other properties. We won't know which molecule will give us the plastic we need until we test the material in the lab. The vast design space combined with multiple, lengthy experiments means that creating a new material using this traditional approach is expensive and typically takes years.

Continuous advances in high-performance computing (HPC) have helped to reduce vast libraries of potentially useful molecules to a much smaller set of probable leads for experimental chemists to follow up on. For decades, researchers have been relying on ever-more powerful computational software to help find better catalysts to design more durable plastics, more efficient components for batteries and solar cells and lighter and stronger aluminum alloys or carbon composites for larger, faster and more cost-efficient aircraft. Computers have enabled us to perform simulations of complex molecules to predict properties of materials that do not yet exist. Yet even with the help of ever higher-performance computers, the materials discovery and design process stayed relatively lengthy and inefficient throughout the second half of the 20th century, and remains very costly today, both in human expertise and computational power.

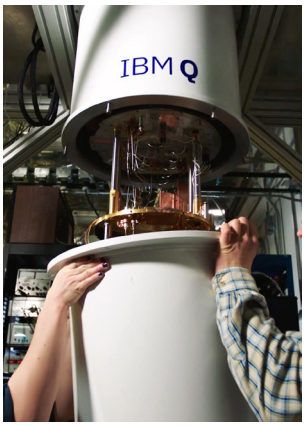
At the same time, the demands on materials keep growing. We urgently need new materials to develop drugs to fight novel and dangerous diseases like COVID-19. We need them to store energy more efficiently and to feed the world's surging population. We need them to 3D-print bone implants and better prosthetics, and to create recyclable plastics that fall apart into their original constituents to enable a waste-free circular economy. We need them to make more environmentally-friendly photoresists—crucial enablers of computer chip manufacturing. These tiny components have for decades allowed our integrated circuits to keep shrinking while doubling the number of transistors, in accordance with Moore's Law, making our gadgets slimmer and more powerful. We need new materials for so much more—from innovative medical tools to lightweight and sustainable cars and faster and cheaper spacecraft, to new superconducting alloys for quantum information technologies and better communication and cryptography systems.

The goal is clear. To be able to discover and design novel materials faster and more efficiently than ever, we need a fundamentally new way to do science. And today, we have all the ingredients to make it happen—with an impact that will touch every sector of our society, our economy—our lives.

At IBM Research, scientists such as Teo Laino from the Zurich lab are pushing the limits of material discovery. Here, Laino is working on RoboRXN for Chemistry, an cloud-based, AI-driven robot that automates material design.

2.

Supercharging Discovery with AI and Quantum



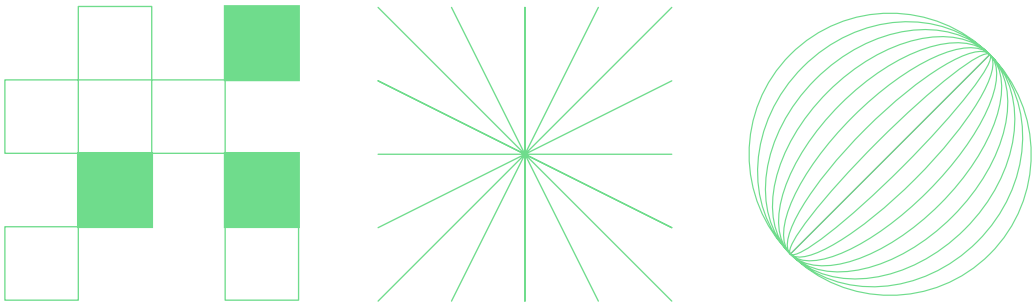
Quantum computers will add an invaluable ingredient to the future of materials discovery

Enter quantum computing and artificial intelligence. Once labelled promising but distant, these technologies are now very real.

Quantum computers are still at the dawn of their innovation journey. But they exist, are developing rapidly and show unique potential for extremely accurately simulating molecules and rapidly predicting the outcome of chemical reactions. AI, for its part, is already making a mark on science and the process of discovery. Researchers are increasingly relying on machine learning to screen novel high-performance materials, create models to assess the crucial relationship between the behavior of matter and its chemical structure, and to predict properties of unknown substances.

Since the explosion of the AI technique dubbed “deep learning” that relies on making computational analogs of our brain’s nerve cells—neurons—around 2012, researchers and companies have realized that their data has more value than previously thought. Today, some consider data the world’s greatest natural resource. Our industries produce 2.5 exabytes of data daily, similar to the amount stored in 250,000 Libraries of Congress or about five million laptops.

In 2018, some 450,000 new papers were published in the field of material science alone. That is a staggering number—and a mountain of data for scientists to tackle. Thanks to AI, we are deriving more information than ever from these data resources; AI systems are beginning to help us to sift through and sort vast materials literature. For instance, researchers have shown that it’s possible to accurately ingest more than 100,000 PDF pages in a day on a single server node—and then train AI models to extract content from them to create a knowledge base for human experts. This *Deep Search* capability is one of the pillars of the new Accelerated Discovery workflow, enabling us to build on the known, tapping into the information tucked away in academic papers, articles, books, patents, and much more.



Once we build on the known, we can fill data and knowledge gaps with classical and quantum-based simulations, together with AI-assisted automated experimentation. While the ability of HPC to simulate physical and chemical processes is well known, the use of quantum computers for these tasks is still in its infancy.

But the potential for enhanced molecular simulation quantum computers offer is very real—and extremely needed.

Molecular simulation requires mimicking the interactions between electrons in each atom with the nuclei of all other atoms. The bigger and more complex a molecule is, the more difficult this process becomes. Today's supercomputers can simulate fairly simple molecules, but when researchers try to develop new complex compounds—whether it's for life-saving drugs or better batteries—classical computers cannot maintain the accuracy they have at smaller molecular scales, or may not be able to carry out a simulation at all. The solution has typically been to approximate the behavior of unknown molecules in simulation, synthesize them, and finally test them experimentally to verify whether they turn out as predicted.

Even when researchers use AI to assist with such searches for 'designer molecules', the first result produced on a classical computer is usually not precise. The back and forth process between tests in the lab and the computer calculations can take months and even years. This is the traditional trial and error iterative method of the centuries past, upgraded to a more systematic approach by use of HPC and machine learning, or AI. The most complex molecule researchers have so far been able to simulate on a classical computer with the help of AI is around the size of pentacene, which has 22 electrons and 22 orbitals. Anything larger, and calculations become prohibitively slow.

In other words, powerful as they might be, AI and HPC alone will not get us to where we need to be to innovate on material development in the future.

Quantum computing promises a solution. Today, this emerging technology is just entering the phase of commercialization as a cloud service. Current performance indicates that these machines will provide us with a quantum advantage—meaning outperforming any classical computer, even if all classical computers were combined together—in certain use cases within this decade. Quantum computers promise to have the most power when it comes to predicting outcomes based on many possibilities, such as simulating a molecule to identify the properties of a specific new material or calculating the investment risk of a financial portfolio.

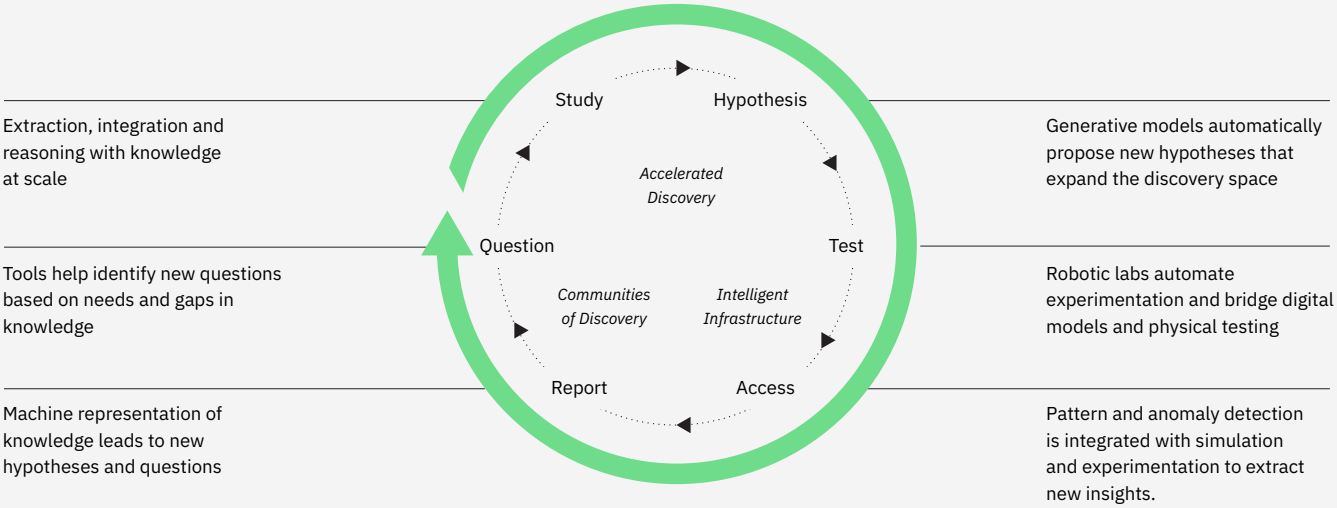
How do they do that? While a classical computer has to sift through all likely outcomes one by one, a quantum machine will go through many possibilities simultaneously, dealing with a much larger range of possible computational states. It relies on "quantum bits" or "qubits" to make computations —the quantum counterparts of the binary digits or "bits" that transistors in a standard computing circuit work with. We can think of qubits as artificial atoms that behave just like the real atoms researchers are trying to understand, but in a programmable way. Atoms are held together with chemical bonds to form a molecule, and qubits are a great way to simulate the full and natural behavior of these bonds—to simulate a new molecule.

Quantum computers will add an invaluable ingredient to the future of materials discovery: accurate simulations for larger scale molecular systems than those currently accessible by HPC. Together, AI, HPC and quantum would enable the new Accelerated Discovery workflow (Figure 1), assisting researchers with the creative process, design, and execution of physical experiments.

What is this workflow of the future?

It is composed of *Deep Search* (step A), *Intelligent Simulation* (step B), *Generative Models* (step C) and *AI-Driven Experimentation* (step D).

The loop of scientific discovery is closing in significant ways
Steps towards scientific advancement as self-propelled, continuous, and never-ending process



Imagine we want to create a new plastic with specific properties: extremely flexible but also lightweight and able to fall apart into its original monomers for recycling purposes. Having outlined these desired properties—having created a hypothesis—the next step is to use AI to sift through the past knowledge on polymer manufacturing to see all the previous research, patents, and fabrication attempts. The AI would identify knowledge gaps and create predictive, so-called *generative models* of a possible molecule for this new polymer. More and more companies are starting to apply generative models already, to a degree, and they are getting ever better. Relying on a desired physical or chemical property of a non-existent material, machines generate the chemical composition and structure of candidate molecules—based solely on data and large numbers of examples of past chemical reactions.

Then HPC or, in future, quantum computing, would conduct a precise simulation of this new candidate molecule and the reactions it should have with its neighbors. And finally, we would use AI again for AI-driven experimentation, where machines would validate the predictions of designer molecules in a lab.

Importantly, this new workflow—the new Scientific Method—doesn't need to proceed from left to right (from A to B to C to D), like the traditional, centuries-old scientific method would. For example, in situations where there is a lack of relevant previous publications, simulations (step B) and/or experiments (step D) would be key to the design of generative models (steps C). Many other flows are possible depending on the situation, and eventually, the method would turn into a loop—one that starts with a question that turns into a hypothesis, goes through A, B, C and D stages and eventually culminates with new knowledge that invokes a new question—and the loop starts again.

Following this workflow should make scientists and engineers involved in the process of discovery much more productive. They would play a key role in each step—curating the data AI extracts from literature to increase its relevance and influencing the selection of the computational method, as well as its duration and cost. They would also contribute their own creativity to the designs proposed by the AI models and select the overall scale of the AI-driven experimentation process.

Crucially, anyone in the world would be able to run AI and quantum processes through hybrid cloud. This cutting edge technology combines on-premises infrastructure—a private cloud—with a public cloud and allows data and apps to flow freely between the two environments. Hybrid cloud gives businesses greater control over their private data, which is key for discovery-driven enterprises of the future.

Using this technology, you won't have to have a quantum computer inside your company—it will be somewhere in a quantum data center. Same for AI. To support experimentation at scale, hybrid cloud platforms for accelerated discovery will enable the user to switch at a drop of a hat between classical, AI, and quantum computational resources, providing the IT infrastructure that will drive the transition of enterprises towards a discovery-driven approach.

3. Opportunities for Accelerated Discovery

Our goal is to cut the time to design a new material from a decade to a year, and to cut the cost from $10 million to 1 million. A challenge? Yes. Doable? Absolutely.

A Accelerating Drug and Therapy Discovery
B Boosting Manufacturing
C Strengthening Agriculture

Accelerating Drug and Therapy Discovery

Healthcare needs a boost. We need better and faster ways to discover new drugs to combat pandemics like the one triggered by the devastating infectious disease COVID-19. We need to continue to develop truly personalized medicine, create better prosthetics, more efficient bioprinting of organs, and vaccines to protect us from deadly viruses.

Just like with any new materials, in early stages, drug discovery and design rely on molecular comparison and choosing the best candidates. An adapted Accelerated Discovery workflow can greatly improve today's process of a pharmaceutical company running millions of simulations and comparisons on high-performance computers. A *Deep Search* of publications and patents could not only help find previous design attempts for a given treatment target, but also help spot beneficial side effects that could be used for drug repurposing. Quantum simulations of selected parts of large drug molecules or specific molecular interactions could selectively increase simulation accuracy of the largely classical simulations. Generative models could identify new drug designs with multiple optimization goals, such as maximum efficacy combined with minimal toxicity.

When it comes to the synthesis of new drugs, biotechnology has great potential—for instance, for growing proteins in genetically engineered crops to create vaccines. Combining classical, AI, and—later—quantum computing to simulate the effectiveness of new biotechnological techniques could be invaluable in the future.

Drug discovery and design still rely on molecular comparison and choosing the best candidates. The new Accelerated Discovery workflow supercharged by AI, HPC, and soon quantum, will give that process a great boost.

Beyond drug discovery, AI has already been used in tandem with classical computers to speed up imaging, diagnosis and data analysis. Neural networks are particularly effective in recognizing patterns, and AI algorithms are already analyzing vast datasets of patient information, detecting specific patterns a human expert would struggle to find. Quantum computers could, in the future, help AI algorithms to find new patterns thanks to their ability to explore extremely high dimensional feature spaces, impacting fields such as imaging and pathology.

And with both AI and quantum computers, it could be possible to tackle, in novel ways, the long-established protein folding challenge—helping us understand why proteins malfunction when they are wrongly folded, leading to the development of new therapies and drugs.

That’s just scratching the surface—the powers unleashed by the current revolution in computing will undoubtedly impact many different areas of our lives.

Boosting Manufacturing

Manufacturing is an extremely complex and multi-process industry that has undergone a massive change in the past years. We are witnessing an acceleration of intelligent automation and robotics and an even more fundamental shift to additive manufacturing (3D printing) and the broad use of composite materials.

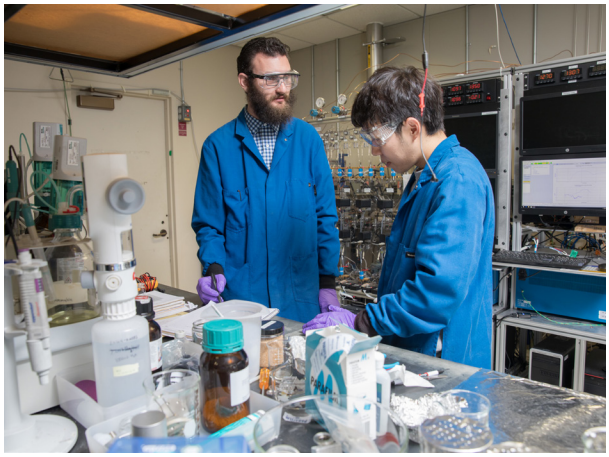
Manufacturing running shoes is a good example. A key requirement is to make soles that provide stable support on uneven terrain, are lightweight, promote foot flexing, and are made of a durable material. No single material can meet all these criteria simultaneously. While a foam cushioning component reduces vertical shock, such cushioning normally contributes to lateral instability. To make the sole, we need to combine the cushioning layer with reinforcements—such as rigid or semi-rigid vertically extending braces. The assembly is typically manual, very time-consuming and tricky—the core reasons why manufacturing of such shoes mostly takes place offshore, in low-cost labor countries.

By using HPC, AI and quantum computing to supercharge the materials discovery and design process, we could engineer novel composite materials that would combine the desired properties of the individual components. This way, we could dramatically cut the labor intensity of manufacturing and be able to much more often make the products in the U.S. It is already starting to happen, and the Accelerated Discovery workflow applied to composite materials can accelerate this shift. It could greatly boost a broad range of material assembly applications, beyond shoes—from architecture to transportation, medical devices to defense and scientific equipment. The combined power of classical, AI, and quantum computing for manufacturing goes far beyond its application to material design.

Another rapidly emerging trend is the use of Digital Twins—the exact simulations of whatever is being built. These digital models allow manufacturing companies to better optimize their performance. Digital Twins are the reason we have very low failure rates in well-run factories. AI and quantum, though, should greatly improve the existing capabilities of Digital Twins, boosting factory operations further. Using these two technologies could help us, for instance, to better understand the dynamics of fluids in a pipe or to determine the rate of the pipe's corrosion.

Then there are electronics. Companies are already starting to incorporate AI across the entire manufacturing process of computer chip fabrication. Quantum computing should be key for optimizing the inputs to the system, analyzing different options and making the best decisions in seconds.

Quantum and AI will be crucial to making the supply chain much more adaptive and event-driven. They will help it to respond to the real-time demands of the manufacturing process—like a living organism, modeling natural processes.



At the IBM Research Almaden lab in San Jose, California, researchers experiment with material design to produce next-generation batteries, going beyond Li-ion. Some battery research already involves molecular simulations on our quantum computers.

The global population is surging, and so is the demand for food. We urgently need to design new, more efficient fertilizers to boost crop yields—and feed the world.

Strengthening Agriculture

Another challenge the Accelerated Discovery workflow could help address is food security—by solving the decades-long nitrogen fixation problem.

The world's growing population could reach 10 billion people by 2050, up from seven billion in 2010. To boost crop yields, farmers have long relied on fertilizers—nitrogen-based compounds that make the soil richer and improve crop growth, such as manure, compost, and, later, chemical fertilizers. Nitrogen is the most abundant gas in the atmosphere, making up four-fifths of the air we breathe, and is a key ingredient of proteins, DNA and other molecules essential to life. But most living systems, including humans and plants, can only use nitrogen in a 'fixed' form, when it is turned into ammonia, having been combined with organic compounds. Ammonia then gets converted into nitrates plants use to make proteins for healthy growth.

Certain bacteria on the roots of plants fix nitrogen naturally—that's nature's clever way to make its own fertilizers to feed plants that feed us. And bacteria still outwit us in this process as they are much more energy efficient than whatever we can come up with in the lab. Researchers have been trying to engineer a catalyst to rival bacteria since the 1960s, in a bid to address the limited supply of naturally fixed nitrogen and tackle the looming global food crisis.

We are not there yet. But the discovery of new materials, in this case catalysts, could help. Currently, nitrogen for fertilization is produced using the so-called Haber-Bosch process that relies on a very energy-intensive iron-based catalyst. Some 10 MWh (or 20 Giga Joules) of energy is needed to produce one ton of ammonia, equivalent to the energy contained in one ton of fossil fuel. The process accounts for two percent of global carbon emissions.

For decades, researchers have been trying to develop a more sustainable, cheaper and less energy-intensive approach. One solution could be creating new materials to facilitate the reaction between nitrogen and hydrogen. This could be achieved by using fuel cells—devices that convert the chemical energy of a fuel into electricity. It's like a reverse battery—instead of storing energy, it uses energy from renewable sources to combine nitrogen from the atmosphere and hydrogen from water to produce ammonia.

AI and quantum computing could help us find new catalytic molecules to lower the amount of energy needed to sustain this process. First, in the *Deep Search* step of the Accelerated Discovery flow, AI would sift through the existing knowledge about catalysts. Then, in the *Intelligent Simulation* step, a quantum computer could precisely simulate different molecules and their behavior, further augmenting our knowledge. Then researchers would use the resulting data to construct *Generative Models* and determine possible configurations of the new molecules. And finally, the candidate materials would be tested in *AI-driven chemical labs* and screened to check for effectiveness. This experimental data would also be crucial to improve the predictive capabilities of the models, with the objective of finding the correct catalyst for nitrogen fixation.

Healthcare, manufacturing, agriculture... The new, upgraded Accelerated Discovery workflow, the closed-loop scientific method that benefits from AI, HPC and quantum computing will without a doubt revolutionize material design.

But material design isn't the limit.

The new workflow should and will greatly impact all aspects of our lives. And with all the global crises of today and tomorrow, the need for it has never been greater.

The world hasn’t run out of problems to solve—and won’t any time soon. We have to stop procrastinating. It’s time to apply the new Accelerated Discovery workflow with AI, HPC and, in future, quantum computing—and to apply it widely. For material design and beyond. Because this world is all we have.

research.ibm.com