Science & Technology Outlook 2021

Accelerating discovery to solve our biggest challenges

IBM **Research** | January 2021 | research.ibm.com

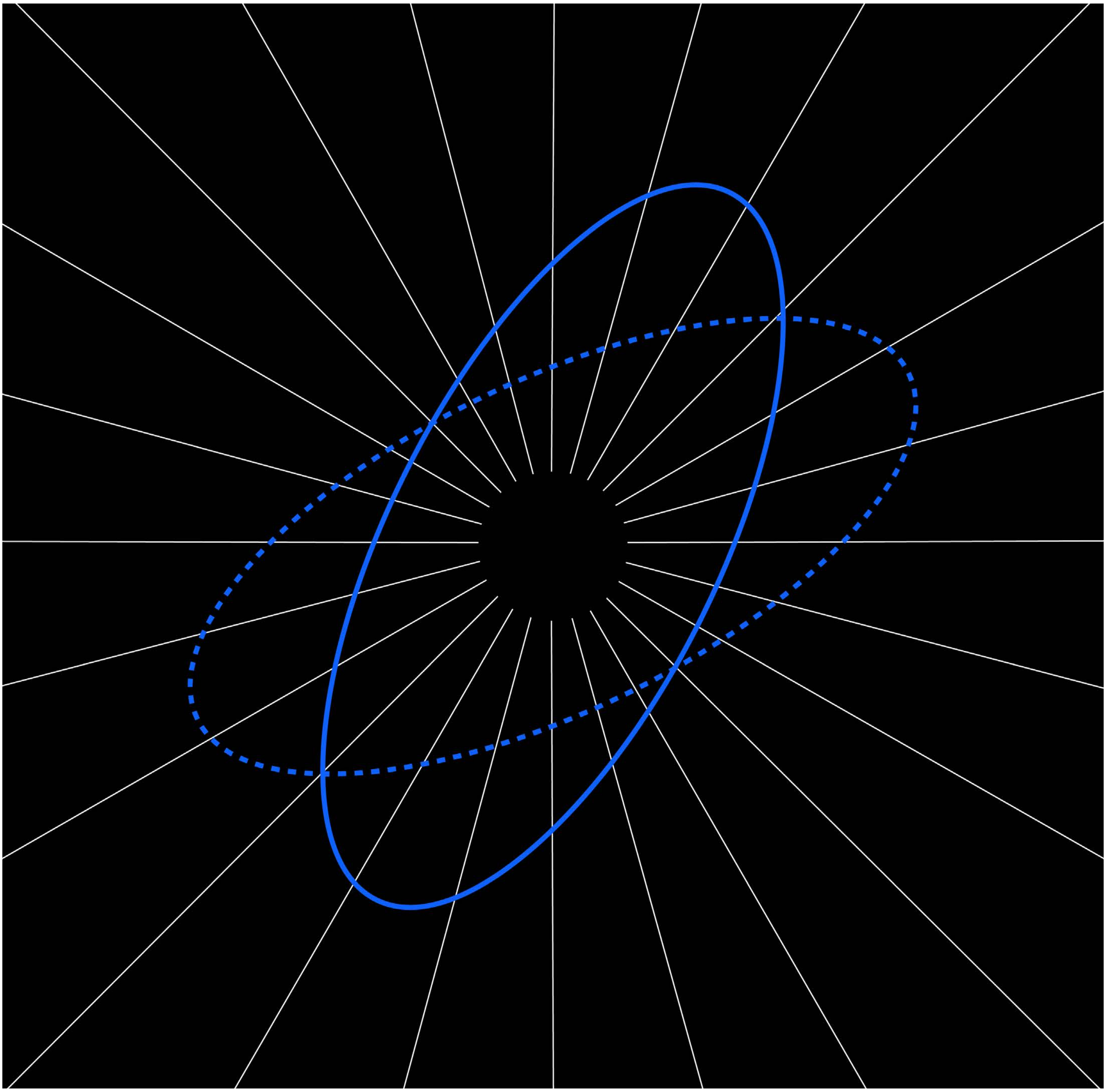


Table of Contents

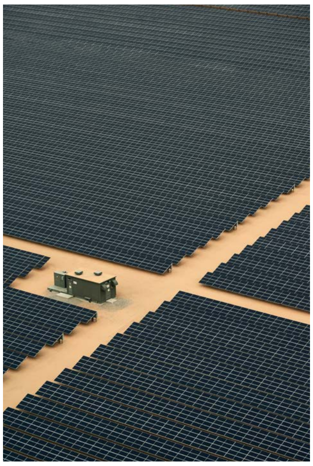
Summary	01
Section One: The Urgency of Science	03
Science Focus Must Transcend This Crisis	03
The Emerging Discovery-Driven Enterprise	08
Hybrid Cloud is Essential for Discovery	11
Section Two: Scaling the Scientific Method	14
Accelerated Discovery	14
Intelligent Infrastructure	17
Communities of Discovery	19
Section Three: A Vision and Roadmap	21
Future of Climate	21
Future of Health	23
Future of Work	25
Utopias and Dystopias	27
Policies and Prevention	29
References	31

#R210001
Written and designed
by IBM Research

Summary

COVID-19's impact on the world has emphasized the importance of science. We are amidst a once-in-a-century global crisis, where the COVID-19 pandemic has spawned one of the greatest races in the history of scientific discovery—one that demands unprecedented agility and speed. At the same time, science is experiencing a sea change of its own, with data and artificial intelligence being used in new ways to break through long-standing bottlenecks in scientific discovery. The timing could not be better. Numerous societal challenges are demanding a dramatically faster pace in science. Simultaneously, there is a growing trend with industry and governments to use the scientific method of discovery, and experimentation at scale as rigorous processes to build knowledge and inform decisions. The idea of the "discovery-driven enterprise"—an organization whose culture is defined by rigorous experimentation, even applied to its own internal processes—is a powerful one, broadly, for enabling more informed decisions and more impactful actions across all aspects of society. These combined forces are giving shape to "accelerated discovery"—in which parts of the scientific process are automated—which will drive new generations of information technology, produce important advances in science, and create new opportunities in business.

History reflects that crises like COVID-19 permanently shape the world in profound ways. The pandemic has highlighted the potential of science both to produce critical breakthroughs and serve as a rigorous methodology to build knowledge



and make decisions. The health, safety, and prosperity of society depend on it. In the "new normal," advanced computing, accessible in the hybrid cloud—a seamless melding of on-site and off-site computing resources—plays a defining role for discovery-driven enterprises. As will communities of discovery—networks of people spanning multiple institutions who share a common scientific purpose, such as creating a vaccine. Powered by pervasive artificial intelligence-infused workflows, they will further accelerate and scale the application of the scientific method to problems in science as well as across broader domains. This new era of accelerated discovery will enable critical advances in climate, health, and work. With the faster pace of scientific discovery comes new responsibility to mitigate risks of unintended consequences while ensuring important beneficial outcomes to society.

This report describes the urgency of science along with the opportunities for accelerating and scaling discovery using the scientific method to produce a more agile and rigorous approach to tackle complex problems in science, business, and society broadly, as follows:

1. **The Urgency of Science** addresses the role of the scientific method in producing solutions to the world's most urgent challenges. Accelerating and scaling discovery is more critical than ever, and transformations in science and business necessitate the development of information technology based on hybrid cloud for discovery.

2. **Scaling the Scientific Method** investigates how the confluence of *Accelerated Discovery*, *Intelligent Infrastructure*, and *Communities of Discovery* is accelerating and scaling discovery—applying the scientific method to achieve unprecedented speed, scale, and automation.

3. **A Vision and Roadmap** examines implications of these advances for the *Future of Climate*, *Future of Health*, and *Future of Work* and explores how to proactively govern the development of technology during its entire lifecycle and responsibly advance innovation and maximize beneficial impacts.

Section One: The Urgency of Science

Science Focus Must Transcend This Crisis

The urgency of science has never been stronger than it is today. COVID-19 caught the world off-guard and disrupted nearly every facet of life: work, the economy, and health. Outbreaks like SARS, MERS, and Ebola should have served as warning signs, but those warnings went largely unheeded. Now, as the latest coronavirus has become a global disruptor, we are struggling to respond. More than ever, the future of our health, our safety, and our prosperity depend upon science. *We need science to move faster.*

The rapid spread of COVID-19 revealed critical gaps in our knowledge and preparedness, which limited our ability to make vital decisions and take swift action to mitigate the crisis. Yet, it is science that gives us hope as we continue to face uncertainties resulting from the pandemic. It is science that has led us to treatments and vaccines. It is science that will help us develop strategies for preventing and mitigating future crises. Beyond COVID-19, numerous challenges—such as those defined in the United Nations' Sustainable Development Goals[1]—are demanding new focus. To address health around the world, climate change, social inequities, and more, we must place the scientific method at the center of our efforts. The scientific method will be a cornerstone to how emerging discovery-driven enterprises operate with greater agility and resiliency. It will be a foundation for how we govern technology and ensure beneficial outcomes for society. *We need to scale our use of the scientific method for greater impact.*

Figure 1:
Science is transforming to a new era of accelerated discovery.

1st paradigm	2nd paradigm	3rd paradigm	4th paradigm	
Empirical science	**Theoretical science**	**Computational science**	**Big data-driven science**	**Accelerated discovery**
Observations Experimentation	Scientific laws Physics Biology Chemistry	Simulations Molecular dynamics Mechanistic models	Big data Machine learning Patterns Anomalies Visualzation	Scientific knowledge at scale AI-generated hypotheses Autonomous testing
Pre-Renaissance	**~1600s**	**~1950**	**~2000**	**~2020**

Increasing speed, automation, and scale

AI-powered autonomous lab
in Zürich, Switzerland

We can revamp the innovation process to address urgent societal challenges by accelerating discovery and more rapidly translating scientific knowledge into practice. We can extend scientific thinking to a broader set of domains and move beyond the natural sciences. We can build on the scientific method as an essential foundation for emerging discovery-driven enterprises. Accelerated discovery can have a vital role in all kinds of decision making, such as enabling agile development of policies and regulations.[2] It can catalyze the development of resilient supply chains and create more responsive and real-world-aware risk models in manufacturing, finance, and healthcare. (See A Vision and Roadmap)

Historically, science has seen a number of major paradigms shifts, as depicted in Figure 1.[3] In the earliest days, science was mostly empirical—it was about observing nature and making measurements. Important changes came later with the emergence of theoretical science, which was about establishing theories and using observations to validate or refute hypotheses.[4] A classic example is Kepler's law of planetary motion, which hypothesized that planets follow an elliptical orbit around the sun and tested and validated this hypothesis using astronomical observations. Later computers brought another big transformation with development of modeling and simulation, including tools for simulating biological processes like ligand receptor binding and protein folding that could be applied for problems like drug discovery. These allowed more rapid cycles of hypothesizing and testing, leading to a faster pace in scientific discovery.

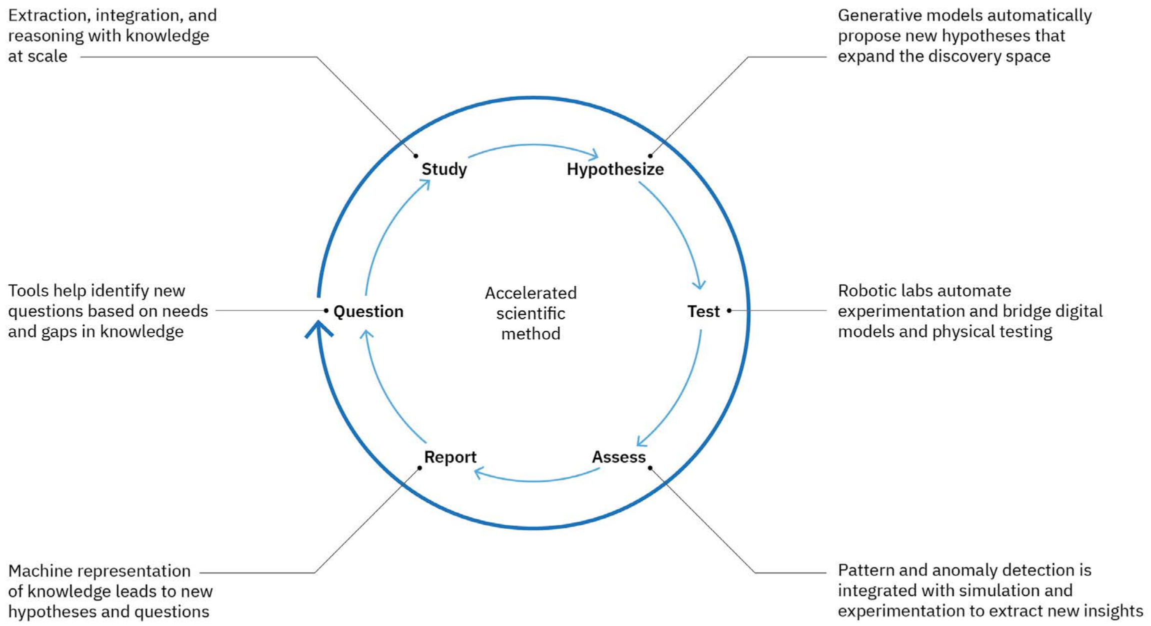
The last two decades have seen the emergence of the Fourth Paradigm of big-data-driven science, dominated by exa-scale systems and an exa-flood of data.[5] The Fourth Paradigm has definitively made science a big-data problem.[6] For example, today virtual chemical databases contain billions of identified and characterized compounds.[7] This same big data is now providing the basis for a new era of accelerated discovery, where artificial intelligence (AI) is enabling unprecedented levels of speed, automation, and scale.[8] This includes greatly improved integration and reasoning with scientific knowledge, automatic generation of hypotheses using novel techniques that expand the search for new discoveries, and automation of experimentation using robotic labs. (We illustrate these components in a case study below, in Accelerated Discovery.) Now building on this data, scientific discovery is entering a new era, where AI, and increasingly quantum computing, is applied by communities of discovery—in the hybrid cloud—to transform the scientific method and overcome long-standing bottlenecks.

The scientific method is composed of a set of steps.[9] Often, it starts with a question, which is followed by study, formation of a hypothesis, testing and assessing the results, and finally reporting. This scientific process is gated by numerous bottlenecks.[10] It is human-expert-driven and episodic. There are challenges right from the beginning in coming up with questions, which require increasingly deep and broad expertise. There are difficulties keeping up with the flood of scientific papers and growing knowledge. More than two million articles are published in 30,000 scientific journals each year. More than 74,000 new COVID-19 scientific related papers alone were added to the National Institute of Health library PubMed[A] in 2020. There are significant challenges in developing hypotheses. Chemical space is infinite. Estimates say there are 10^{63} potential drug-like molecules. Our knowledge is incredibly sparse compared to what is possible. Similarly, there are gaps in testing, including bridging digital models and physical testing, and ensuring reproducibility. It has been reported that 70% of scientists have at least once tried and failed to replicate the experiment of another scientist.[11]

[A] Explore the papers on PubMed at: pubmed.ncbi.nlm.nih.gov

The changes that are making science AI-powered, increasingly automated, and community-driven are closing the loop of scientific discovery in significant ways, as depicted in Figure 2.

Figure 2:
The loop of scientific discovery is closing in significant ways.



This includes advances at each step, for example, to extract, integrate, and reason with knowledge at scale to better respond to questions,[12] to the use of deep generative models to automatically propose new hypotheses, to automating testing and experimentation using robotic labs.[13] Important advances in machine representation of knowledge also enable new results that lead to new questions and hypotheses.[14] For the first time, the loop in the scientific method is closing—*each breakthrough is a step towards realizing the dream of discovery as a self-propelled, continuous, and never-ending process.*

Deep generative modeling is one important example of an emerging AI technology for scientific discovery. The last decade has brought about a revolution in AI based on deep learning and neural networks, which has created significant advances in capabilities for discrimination tasks. More recently new developments in AI technology based on pre-trained language models—which can write sentences—and other deep generative models are being used to automatically generate images, speech, and natural language.[15] In domains like chemistry, deep generative models can generate new candidate chemicals, molecules,[16, 17, 18] and materials,[19] and expand both the discovery space and the creativity of scientists. This is critical in applications like materials science, where the design and discovery of new molecules is faced with a chemical space that is uncountably vast.[20, 21]

These accelerated discovery methods, together with advances in computer infrastructure and increasing focus from communities of discovery, will speed up scientific discovery to address urgent challenges, as depicted in Figure 3(a). There are two other dimensions that are important for scaling the scientific method. First, the scientific method can and does apply to problems beyond natural science, as depicted in Figure 3(b).

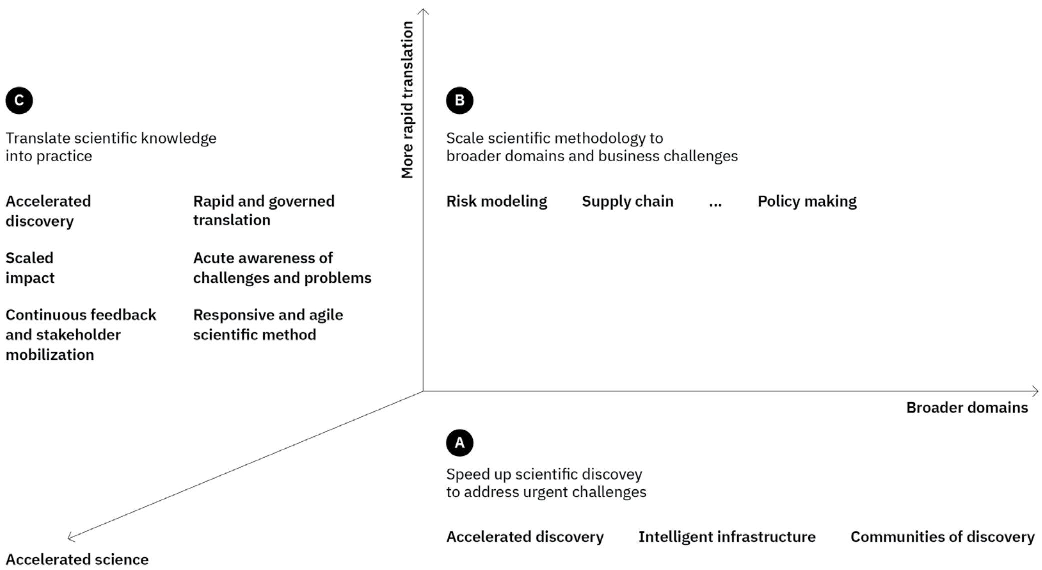


Figure 3:
Opportunity to (A) speed up, (B) scale up, and (C) scale out scientific discovery with greater agility and for larger impact.

A good example is the organization J-PAL,[B] which employs the scientific method through field experiments to test hypotheses and make policy decisions in areas including microfinance, public health, and agriculture. For example, to the question of "what interventions increase educational outcomes at the lowest cost?", the answer is determined through randomized controlled trials that find "targeted help," in specific contexts is the best intervention.[2] Similarly, to questions about pricing medicines, hypotheses drive randomized controlled trials that test them and ultimately guide policies and actions. Esther Duflo, J-PAL's co-founder, won a Nobel Prize in Economics in 2019. Accelerating science will benefit such applications outside of the natural sciences.

[B] Visit the poverty action lab: povertyactionlab.org

Similarly, in a business context, there are enterprises today that are making increasing use of the scientific method—where discovery goes beyond traditional analytics, and is an active process of hypothesizing, testing and experimentation.[22] Enterprise discovery builds on data and AI and provides an accelerated wheel of exploration that allows enterprises to build knowledge, answer questions, and enhance operations, decisions, and offerings. This ability is vitally important to emerging discovery-driven enterprises in the context of COVID-19, as enterprises have continuous streams of important questions related to external factors and impacts on their business. Enterprises need discovery to answer these questions.

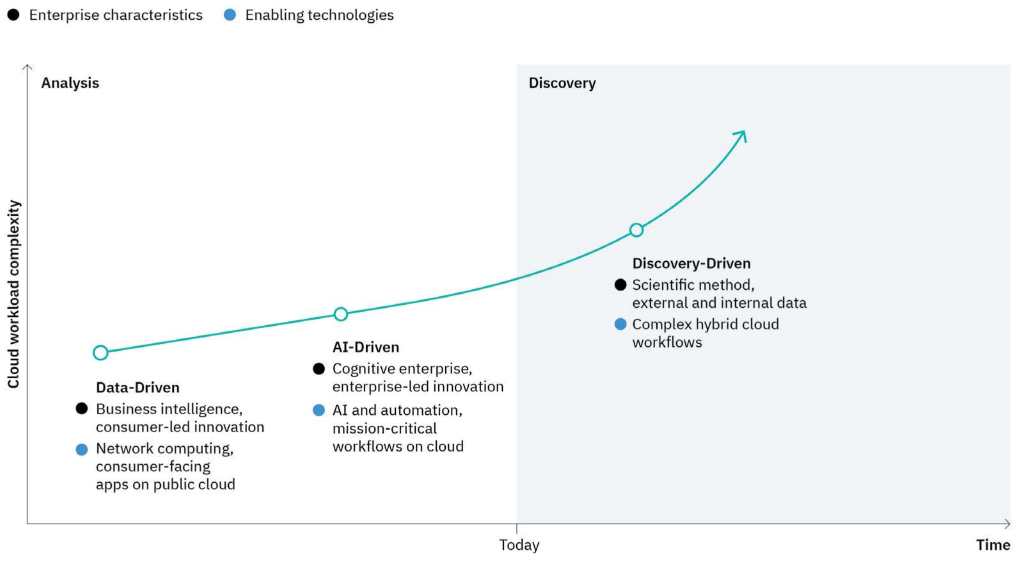
In addition to accelerating science and applying it to broader domains, the third dimension of scaling, as shown in Figure 3(c), is the more rapid translation of scientific knowledge into applications. It is important for the wheels of scientific discovery to go faster. But science is needed for more than science's sake. Today, the translation of knowledge or diffusion of innovation towards impact does not take a direct route, as some scientists publish fundamental advances and move on, leaving others to mine published results for applicable findings. A more rapid and governed approach to translation is needed to achieve impacts at greater scale. Deeper engagement from communities of discovery is needed across the entire process, to help not only with discovery and its translation and impact, but also to obtain acute awareness of challenges and problems. For example, in developing medicines, research labs and pharmaceutical companies should provide feedback on how to improve software environments for their needs.

The consequences of these transformations of discovery are dramatic: The acceleration of the scientific method, the scaling of scientific expertise, and the potential to expand discovery to a broader set of problems—*achieving critical breakthroughs that impact society and industry.*

The Emerging Discovery-Driven Enterprise

The COVID-19 crisis has changed businesses in profound ways. Many of these changes will be permanent. COVID-19 has affected entire supply chains from resources to production and manufacturing to distribution. Workers are affected at a global scale. Businesses are operating inconsistently. Consumer behavior is changing daily. COVID-19 has created new uncertainties that require enterprises to be more agile and responsive than ever and to constantly make sense of external data. Prior to COVID-19, many businesses had begun the journey of increasing use of analytics and AI to improve business processes. This created a growing focus on exploitation of core enterprise data assets—such as user or transaction data—and transformation of enterprise workflows—by removing, digitizing, or automating tasks related to, say, production or billing. More than ever, businesses must now tap into a wealth of external information—such as related to global health and climate—to guide decisions and adapt their operations and strategy.[23] In an interconnected world, a virus incubated in bats from Asia can bring down a company in America.

Figure 4:
Discovery-driven enterprises use the scientific method to guide decisions and adapt strategy and operations.



Obtaining a deep understanding of external forces and the changing landscape is critical for business continuity and resiliency. Businesses need discovery tools and platforms that allow them to assimilate information from sources beyond the core business—information on politics, the environment, social movements, and other industries. The rapid collection of data will inform decisions, and the application of scientific rigor will help create knowledge and manage risk, allowing businesses to adapt to the changing environment. Post-COVID-19, the emerging enterprise will be a discovery-driven enterprise.

Discovery will drive the business of science along three segments in industry, as enterprises utilize discovery within various parts of the value chain:

1. Science as a business – which includes industries such as life sciences, chemicals, and materials, where the practice of science is core to the business;
2. Businesses that rely on science – which includes sectors such as energy and utilities, transportation, healthcare, and technology hardware that rely on results and outputs of science (e.g., geology, weather, medicine, and physics); and
3. Information-driven enterprises – which use the scientific method and experimentation at scale—building on data and AI—to create new critical knowledge about markets or management practices that improves business decisions, development of products, and operations.

The science efforts across these segments are built on a $2.4 trillion annual research and development (R&D) expenditure worldwide, driven by R&D across governments, industry, and non-governmental organizations.[24] As these three segments respond to the urgency of science and increase investment, new opportunities will be created. The use of the scientific method presents opportunities to reduce unnecessary spending and increase profitability across all parts of the business value chain. An example is the increasingly large number of companies who are regularly experimenting in sales and marketing. A common example is A/B testing to compare different versions of websites, the way Netflix experiments with movie recommendations. A more extensive example by some retailers is the experimentation of the entire supply chain to allow more rapid testing of different products and quick responses to consumer interests and trends.[25] Inditex, which owns Zara, goes beyond predicting fashion trends each season. Instead, they continually introduce new items on small scale and measure performance before releasing them broadly.

IBM Cloud data center in Dallas, TX
for AI and Hybrid workloads

As enterprises become more discovery driven, transformations are required in the culture, skills, business process, tools, and platforms.[26] For experimentation to be effective, it needs to be performed at scale and in a frictionless manner throughout the organization. A discovery culture is evidence-based, which requires adaptivity and openness. Fluency in data science, including statistical methods, is important to accelerate such changes. Beyond traditional AI tools, enterprises need hybrid cloud platforms to support experimentation at scale. These transformations power enterprise discovery efforts, drive advances in domains such as climate, work, and health, and enable activities in accelerated discovery broadly.

Hybrid Cloud is Essential for Discovery

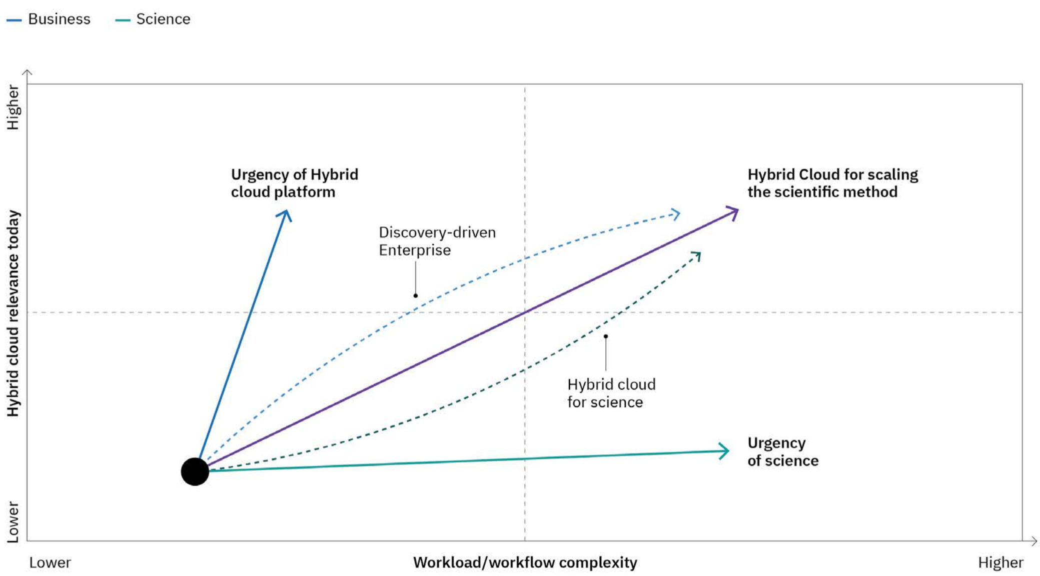
Transformations in science and business make clear that the *hybrid cloud for discovery* is necessary. Science has long pushed the frontier of advanced computing—generating some of the most complex data and computationally intensive workloads in industry.[5, 8] On the other hand, business is leading the adoption of cloud—driven by software application modernization, digital transformation, and other enterprise priorities. With the emergence of accelerated discovery, the paths for science and business are now converging, as shown in Figure 5. Science is moving beyond dedicated advanced compute systems to make greater use of hybrid cloud.[27] Business is building on data and AI to become discovery-driven in hybrid cloud.

Most businesses today have diverse information technology (IT) environments that involve multiple public- and private-clouds and on-premise resources. This mixture of existing resources drives the need for hybrid cloud for four reasons, which we call *history*, *choice*, *physics*, and *law*.[C] History means leveraging investments already made in infrastructure. Choice means freedom to host and move as needed and not be locked into one infrastructure. Physics refers to the need for IT to be close to data or services in order to reduce latency, for example when managing assembly robots. Law refers to compliance with legal frameworks that dictate where data and compute reside; hospitals might not be allowed to move patient data off-site.

Concurrently, recent initiatives in science, such as *European Open Science Cloud*,[28] *NIH STRIDES*, and *NSF CloudBank*, are making cloud the destination for science, by connecting researchers to online data, software, and processing. Helix

[C] Arvind Krishna gives the IBM Think Digital Opening Keynote: ibm.com/about/arvind/speeches/05-05-2020

Figure 5:
Needs of business and science converge with hybrid cloud for scaling the scientific method.



Nebula Science Cloud from the European Commission[29] and the U.S. Department of Energy's Research Hybrid Cloud at Oak Ridge National Laboratory are exploring hybrid cloud platforms for science. These efforts are uncovering new imperatives for hybrid cloud for science, namely *heterogeneity, reproducibility, gravity*, and *open* as follows:

- *Heterogeneity* to support seamless workflows across highly diverse resources including scientific instruments, sensors, physical devices, and entire labs and research organizations.
- *Reproducibility* to enable replication of scientific experiments and results regardless of differences in IT infrastructures or location of data and resources.[30]
- *Gravity* refers to the strong pull from extremely large data sets—some at a petabyte scale—as well as proximity required due to physical manifestation of experiments and instruments.
- *Open* refers to the prominence of open science practices of communities that may dictate open- or *FAIR*- (findability, accessibility, interoperability, and reusability) data access.[31, 32]

Discovery using the scientific method will transform business enterprises, evolving rather than replacing data- and AI-driven transformations enabled today by hybrid cloud. Discovery entails new continuous cycles of experimentation that feed enterprise analytics to further enhance operations, decision making, and offerings. Discovery-driven enterprises need hybrid cloud to enable this end-to-end *exploration*, *experimentation*, *orchestration*, and *exploitation* building on data and AI.

Accelerated discovery tasks are uniquely defined by both a high intensity of workload and a high complexity of workflow. A compelling example is the science conducted by CERN and the World-Wide Large Hadron Collider (LHC) Computing Grid.[D] This is a deeply coordinated big-science effort that conducts two million experiments per day involving more than one hundred sites across dozens of counties in carefully orchestrated workflows that involve large amounts of compute and data. Another example is the COVID-19 High Performance Computing consortium[E] that rapidly mobilized 6.8 million processing cores to drive scientific research efforts responding to COVID-19. These examples show the large amounts of computation, data, and data movement that make up

[D] Learn more about the grid at: home.cern/science/computing/grid

[E] Visit the consortium at: covid19-hpc-consortium.org intensive workloads embedded in complex workflows. Accelerated discovery, which often entails intensive workloads and complex workflows, is a leading candidate to drive the next frontier of hybrid cloud. The accelerated scientific method will be hybrid cloud's proving ground, showing its potential for both science and business.

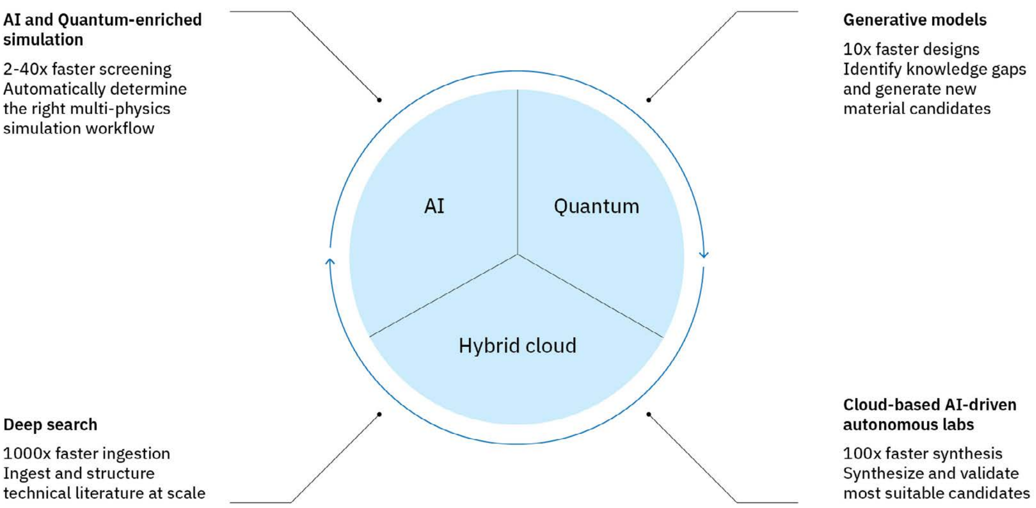
IBM Quantum lab at the Thomas J. Watson Research Center

**Section Two:
Scaling the Scientific
Method**

Accelerated Discovery

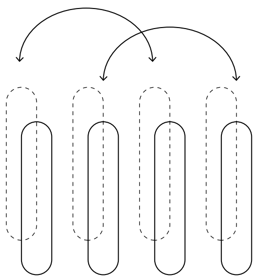
The pace of progress in science has been historically limited by bottlenecks, as described previously. These bottlenecks are increasingly being overcome with the application of AI, quantum computing, and hybrid cloud technologies. New technologies are enabling accelerated methods of discovery that include deep search, AI- and quantum-enrich simulation, generative models, and cloud-based AI-driven autonomous labs, as shown in Figure 6.

Figure 6:
Accelerated discovery for materials design.



Recent developments in materials science have demonstrated the potential of accelerated discovery.[13, 33, 34] For example, deep search using AI is speeding up ingestion of scientific papers and extraction of knowledge by 1000x compared to human experts[35]—ingesting tens of pages per second, compared to multiple minutes per page required by human readers. The AI first parses PDFs into different kinds of content, then builds knowledge graphs, networks in which different concepts are linked together according to their relationships. A molecule might connect to its properties, related molecules, and databases of experiments on the substance. AI is also being used for predictive simulation to automatically choose and optimize what simulations to run and in what order, as well as the methods to use—including quantum simulation, which speeds up overall simulation efforts by 2-40x. Similar techniques can also select what real-world experiments to run and in what order, reducing the overall experimentation effort by 50%, as we found in our experience with a chemicals company.

Generative models and deep learning architectures like transformers are also radically changing material design.[16, 17, 18, 36, 37] Generative models for hypothesis design are speeding up early phases of chemical discovery by 10x.[38] Transformers have enabled a first-of-a-kind cloud-accessible AI-driven



robotic lab that has demonstrated speeds of experimentation up to 100x faster than traditional methods.[13] An integrated set of accelerated discovery tools is being applied to the design of high-performance sustainable photoresist materials—used in the manufacture of computer chips—with a goal of speeding up the overall scientific discovery process by 10x and lowering cost by a similar factor. Unifying these technologies will create a complete, closed-loop capability for accelerating discovery. With the first end-to-end implementation in the context of materials science, the discovery workflows can be captured and reimplemented as flexible, continuous, AI-driven processes to solve other kinds of problems.

A multi-disciplinary team of computer scientists, systems researchers, data scientists, and subject matter experts are building a computational workflow using the tools depicted in Figure 6 to drive the discovery of new photoresist materials. The deep search tools are being used to extract photoresist materials and properties from 6,000 papers and patents. The performance properties needed for design were sparse, so the AI-enriched simulation platform was used to augment the extracted data with predictive simulations of each material.[F] If, say, the melting point of a certain material had not yet been measured and published, simulation would fill in the blank. This augmented dataset is being used to train a generative model that created thousands of materials candidates with targeted properties in a matter of hours. While simulations model the properties of a given material, generative models learn from the data and suggest new materials with a given set of properties. The candidates from the generative model are then evaluated by human experts to select top choices for the next step. In late 2020, the AI-driven automated lab synthesized the first candidate photoresist material. This was the first time this entire discovery process was exercised, and the new material was created 2-3x faster than prior manual approaches. Measurements of the newly created material can then feed into other stages of the process, creating a feedback loop that further accelerates discovery.

The integration required to build this system was challenging. Each computational tool was a complex application with unique requirements. Variability in compute intensity, data volume, persistence, and security introduced additional challenges. Some components emphasize traditional classical computing, while others require accelerators for AI training and inference. Other components are tied to a specific physical location, like the automated synthesis robot. A hybrid cloud platform that enables scientists to quickly

[F] Try IBM RXN for Chemistry here: rxn.res.ibm.com/

build this kind of complex discovery workflows and switch between classical, AI, and quantum computational resources will further supercharge the discovery process. Accelerated discovery requires integration of multiple complex workflows with different experts, implementers, and stakeholders. Cloud software often uses containers—bundles of programs and all the auxiliary software they depend on—allowing for consistent performance across different computers. Modern, container-based architectures on OpenShift—software for developing and running programs in the cloud—will help realize benefits of this integration.

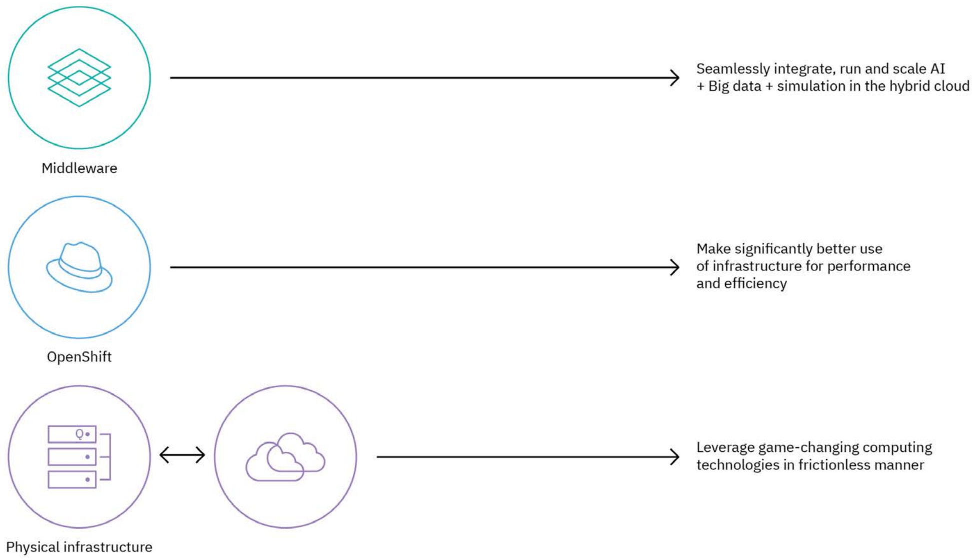
Computational discovery is an intensive and complex workload. Even in its simplest form of virtual screening and simulation, scientific discovery accounts for a significant fraction of the computation done using supercomputing resources (~one million processing cores per year for materials science in the U.S. alone[G]). An expanded accelerated discovery workload is an opportunity for hybrid cloud to bridge across a mix of data and compute resources and drive the integration of quantum computing. There is further opportunity to transform the process of discovery through AI, intelligent simulation, quantum computing, and generative models—and in so doing accelerate and scale impact from scientific discovery. Making this possible will require innovation at the workflow level but also in the underlying infrastructure supported by hybrid cloud.

[G] Explore computing data at: xdmod.ccr.buffalo.edu

Intelligent Infrastructure

Emerging technologies like AI and quantum computing demonstrate enormous potential to accelerate scientific discovery.[8, 39, 40] This opportunity requires innovations across the full hybrid cloud stack, as shown in Figure 7, from reimagining middleware—software that sits between the operating system and user applications—to enhancing the way processing gets distributed across computers.[41, 42] It also requires redefining the way infrastructure is virtualized—such that one computer can be treated by simultaneous users as several independent computers, to optimize both agility and security—and enabling seamless exploitation of quantum computing. These advances will make it simpler to define discovery workflows, flexibly manage and deploy them, and enable accelerated scientific discovery at scale. As described above for accelerated materials design, defining a discovery workflow today requires significant manual effort. It involves configuring multiple computer processing tasks, writing ad hoc scripts to do tasks like moving data, and managing execution across disparate resources. New types of middleware and unified runtimes—systems that exploit all available processors—are needed to tie these steps together and tightly coordinate tasks, ensure elasticity and resiliency, and evolve a so-called "serverless" computing foundation in which underlying server infrastructure is invisible to the user.

Figure 7:
Hybrid cloud innovations for accelerated discovery.



Accelerated discovery workloads also bring unique requirements for resources. Some steps, such as AI inferencing for hypothesis design, may require hardware acceleration; others, such as tightly coupled simulations, may need very low network latency. Today, Kubernetes—software for managing software containers—does not support the sophistication needed in fine-grained computer resource allocation. This gap can be closed by improving Kubernetes' scheduling of program execution, improving Kubernetes' and OpenShift's abilities to discover devices connected to a network, and creating an API—a way for programs to interface with each other—that enables coordination between software that discovers computing resources and software that assigns tasks to them.

Deploying discovery workloads in cloud environments is also challenging. After mapping a job to a node, the infrastructure gets virtualized so that the workload can be deployed, for instance inside a virtual machine (VM). VMs are slow to spin up and spin down—but are used in the public cloud due to security concerns. In contrast, containers are much lighter weight and faster to start and stop–but share functionality in the kernel (the core of an operating system) which makes them vulnerable to hacking. There is opportunity to define the next generation of virtualization technologies, a forward evolution towards a "microVM," to enable the agility of containers with the security of VMs to support discovery workloads.

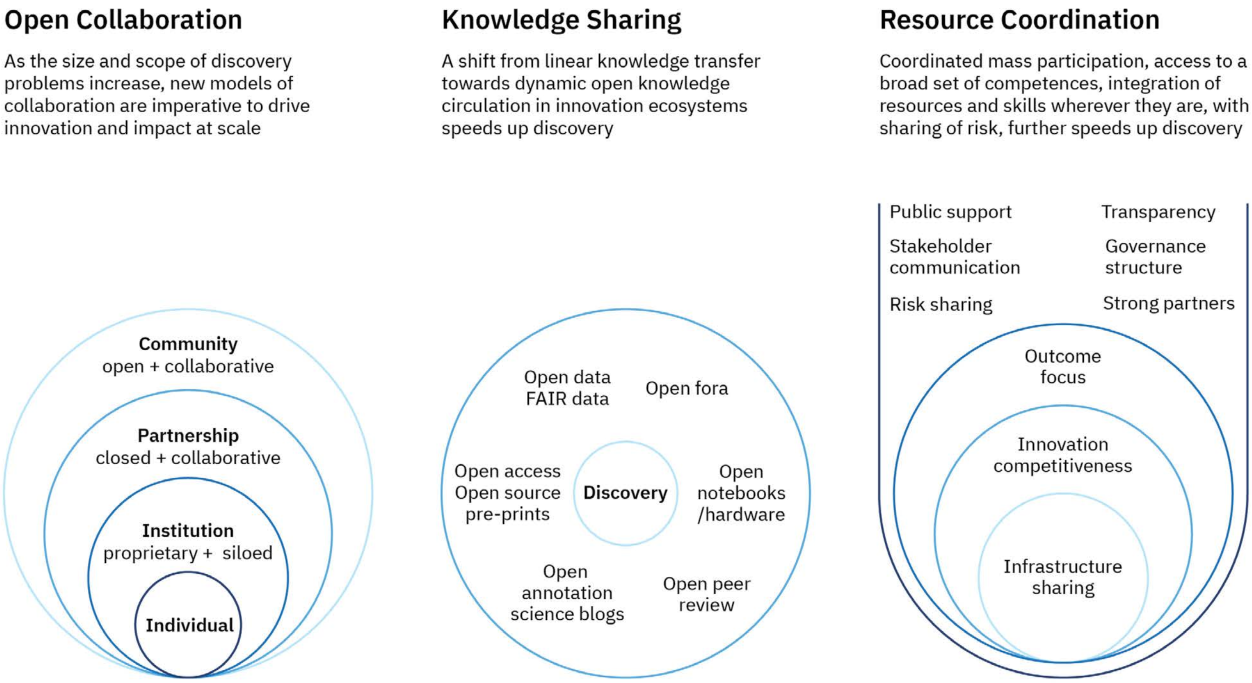
Quantum[H] computing will be transformative for scientific discovery.[39, 40] In order to make quantum computing practical in the hybrid cloud for discovery, it must mature in several ways. The quality with which we can run quantum circuits needs to be enhanced, which will be essential for increasing quantum volume, a measure of both number of qubits and how long they last. The variety of quantum circuits that can be run on a quantum computer needs to be enhanced to enable users run new kinds of experiments. A new quantum runtime—code that underlies programs for quantum computers—is needed in order to speed up how quickly quantum circuits can execute a quantum system. Finally, quantum development needs to be simplified with a lower barrier to entry by making quantum programming truly frictionless.

[H] Visit IBM Quantum at: ibm.com/quantum-computing/

Communities of Discovery

Communities of Discovery are becoming the new paradigm for the practice and advancement of scientific discovery.[43] As illustrated in Figure 8, they build on open science practices and are characterized by dynamic knowledge circulation and well-coordinated collaboration.[44] Discovery communities are purpose driven, and the impetus for their formation includes infrastructure sharing, innovation competitiveness, and a collective mission focus.[45] Enterprises cannot remain competitive without leveraging the abundant knowledge, creativity, and resources in these communities. The scientific discovery practiced in these communities are precursors to the next generation of high-value discovery workflows and workloads in industry.

Figure 8:
Scaling the scientific method requires communities of discovery.



The COVID-19 crisis has energized the spirit of collaboration and brought science to the attention of a global society confronted by a common threat. Collaborations have emerged across academia, government labs, industry, and non-profit organizations, bringing together diverse expertise and resources. The COVID-19 *High Performance Computing Consortium*,[I] a public-private partnership, is made up of dozens of consortium members from government, industry, and academia. The consortium supports COVID-19 research by providing researchers access to the world's most powerful cloud and high-performance computing resources.

[I] Visit the consortium at: covid19-hpc-consortium.org

The modern scientific discovery process demands reproducibility of results, collaboration, and effective communication of knowledge for further expansion. This, together with the digitization and acceleration of discovery, creates the need for portability, elastic capacity, AI-based tools, and security across multiple clouds. These needs can be met by a hybrid cloud for discovery with OpenShift orchestrated multi-clouds and platforms and tools for discovery. OpenShift creates a standard interface for services from multiple clouds. That software layer allows for easy access to platforms and tools for tasks like computers security and open-source software development. A hybrid cloud for discovery helps productivity, collaboration, integration, and scientific reproducibility, while providing a way to obtain feedback to improve the platform and further grow adoption. Leveraging communities of discovery to accelerate solutions to large-scale problems is essential for scaling impact and creating a virtual cycle that accelerates value creation. These communities will capture the next dominant workflows and workloads for accelerated discovery and will drive a robust supply chain for innovation and value creation and achieve a scale of impact that is critical for society.

**Section Three:
A Vision and Roadmap**

Future of Climate

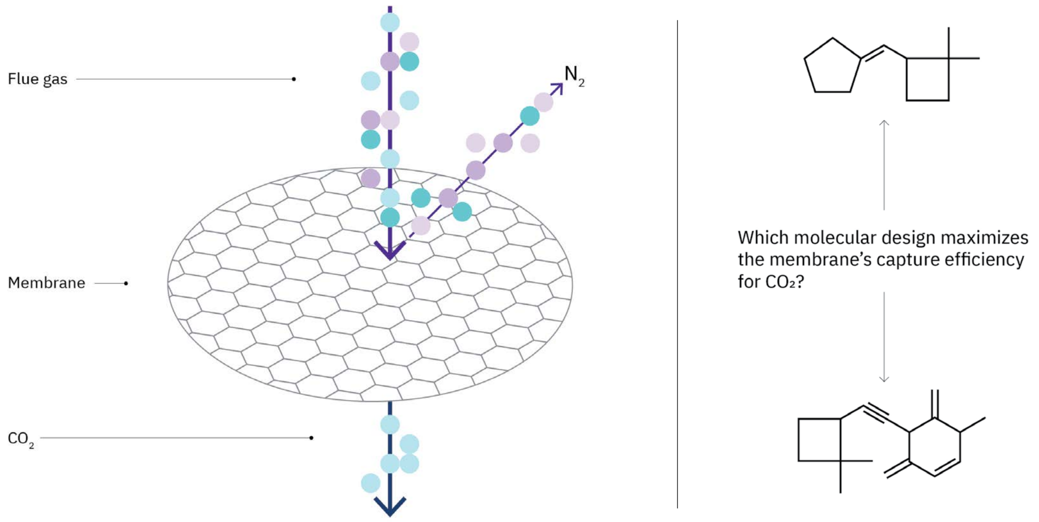
The crisis of COVID-19 parallels that of climate change. Both need an urgent, science-based, coordinated response to mitigate devastating impacts. This urgency is amplified by consumer and investor demands, policy and regulation changes, and business needs for resiliency. Advances in AI and hybrid cloud accelerate the ability of companies, policymakers, and communities to address climate change.[46]

Sustainable hybrid cloud: Datacenter energy consumption is expected to grow to 8%–20% of the world's electricity use by 2030.[J] Companies are increasingly forced to take responsibility for their carbon footprint, and IT is a natural place to start. IDC declared the datacenter carbon footprint as the new battleground for cloud providers. Increasingly, cloud users call out carbon-footprint reduction as a key motivator for cloud provider selection. This has resulted in innovations like underwater datacenters and AI-optimized datacenter operations. Building on an open and secure hybrid cloud, there are further opportunities to develop sustainable architectures that allow customers to measure, quantify, and optimize their workloads' carbon footprint on premise and off premise.

[J] How to stop data centers from gobbling up the world's electricity: nature.com/articles/d41586-018-06610-y

Climate-smart supply chains: Supply chain disruptions during COVID-19 increased focus on supply chain resiliency. Pressure from businesses, investors, and governments is increasingly the urgency to make supply chains climate resilient and carbon responsible. Advances in data, AI, and compute enable climate models at regional scales and help optimize supply chains to increase climate resiliency and reduce the carbon footprint of supply-chain and cloud operations.

Figure 9:
Accelerated discovery can speed up the design of molecules for carbon capture.



Materials for carbon capture: The removal of carbon dioxide (CO2), the dominant greenhouse gas affecting climate change, from power plant emissions and eventually, the atmosphere, requires new cost-effective carbon capture materials and scalable processes. Electricity production generates 40% of CO2 emissions but less than 1% of that emission is captured today. Accelerated discovery can identify and optimize new materials for carbon capture, such as to design membranes for CO2 capture from flue gas, as shown in Figure 9, by leveraging AI-enabled scientific workflows that scan thousands of publications and databases, organize the properties of millions of chemical structures, and predict, synthesize, and validate the optimal molecular designs for efficient carbon capture and reuse at scale.

Future of Health

COVID-19 has changed the world in many ways—impacting employment, travel, global supply chains, greenhouse gas emissions, remote work, and the economy. But make no mistake that at its core this is a public health crisis of the scale the world has not seen in over a century. The devastating pandemic has highlighted critical opportunities related to public health, as summarized in Figure 10: treatments, vaccines, and cures might be more rapidly identified by mining real world evidence data; emerging infectious diseases might be continuously monitored for using genomic surveillance; and healthcare might be delivered in a more personalized and dynamic manner. We must advance capabilities on all these fronts to better anticipate and respond to future public health crises.

Figure 10:
Accelerated discovery is critical for improving our preparedness and response to health crises by (A) allowing drug repurposing at scale, (B) ensuring safer workplaces, and (C) enabling more personalized, dynamic, and efficient healthcare delivery.

	Platform		Technology		
A	Real world evidence	+	Casual inference	=	Drug repurposing at scale
B	Genomic surveillance	+	Anomaly analytics	=	Safer workplaces
C	Healthcare system readiness	+	Personalized health prognostics & Dynamic matching	=	Reduced disease burden

Accelerated discovery of treatments, vaccines, and cures: The COVID-19 crisis has created a new urgency to have the right drug at the right time and better understand what that requires. Drug discovery is a lengthy process; it can take up to $2.6 billion and more than 10 years for a new drug to reach the market. A third of this overall cost and time is spent in the drug discovery phase, during which researchers synthesize thousands of molecules to develop a single pre-clinical lead candidate.[47] To generate treatments for emerging viruses like COVID-19 more quickly, there is opportunity to identify potential therapies from safe and proven drugs.[48, 49] As shown in Figure 10(a), large datasets of real-world evidence can be mined using causal inferencing technologies to identify candidates for drug repurposing at scale.[50, 51] This would jumpstart subsequent research to enable more rapid clinical trials and regulatory review.

Continuous disease surveillance: COVID-19 emphasized the importance of early warnings and preparedness. A continuous monitoring system can detect, model, and track emerging infectious diseases, such as by performing anomaly detection on genomic surveillance data, as shown in Figure 10(b).[52] Such a system would include three key elements: entry point detection (signals from wildlife, breeding farms, and dairies), outbreak surveillance (signals from hospitals, social media, smart thermometers, and more), and microbe sleuthing (signals from patient microbiomes, urban surfaces, and wastewater[53]). Early studies around SARS-CoV-2 have already shown that these signals can provide effective means of pathogen detection.[54] Each disease surveillance element requires advanced methods for data acquisition, integration, and analysis. This monitoring would take advantage of the scalability, portability, and flexibility of hybrid cloud.

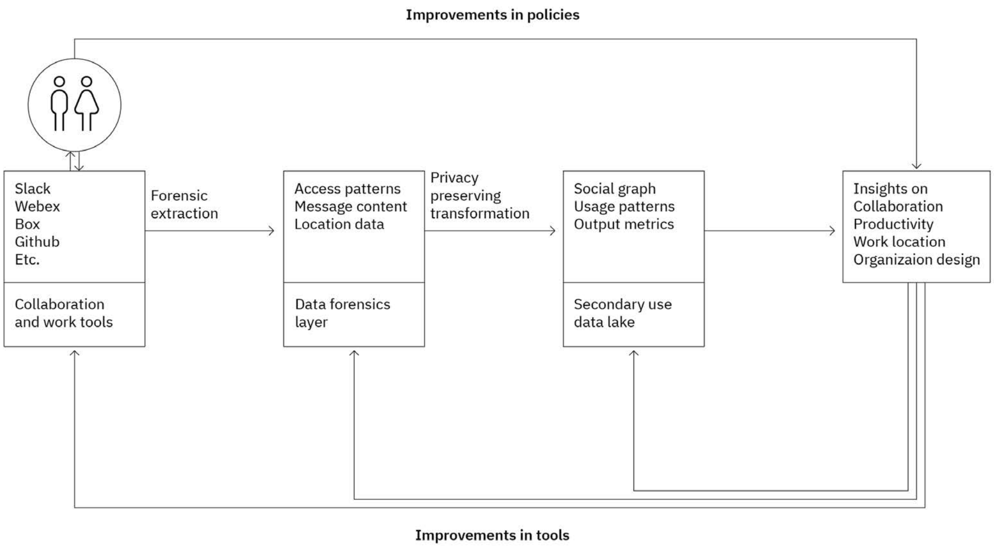
Adaptive delivery of healthcare: The COVID-19 pandemic revealed weaknesses in global healthcare systems. Capacity constraints and equipment shortfalls led governments around the world to impose social restrictions to limit virus spread and flatten the curve of infection. The health impacts of COVID-19 impacted many more people than those infected with the virus. People with pre-existing conditions risk increased morbidity or mortality due to delays in all but the most essential care. This triggered greater adoption of telemedicine and presents a significant opportunity to tailor delivery across all levels of care. To deliver the most effective care in a timely manner, there is a need to better understand the health of individuals, population trends, and the readiness of care facilities. Platforms using AI and big data are emerging that combine patients' clinical tests to draw health insights, or monitor transactions between patients, providers, and payers to identify system inefficiencies.

Future of Work

The COVID-19 pandemic is radically reshaping our ways of working.[55] The practices, policies, and technologies developed will define work for decades to come. The pandemic hit when most work was performed at legacy workplaces, where worker safety was a primary concern only in hazardous industries. The labor market was strong but economy-wide productivity growth in the U.S. was the lowest since the 1970s. New technologies entering the economy were not finding broadly productive uses and historically marginalized groups continued to face inequalities in the labor market. A massive shift toward remote work has disrupted our workforce.[56] Jobs that cannot be done remotely are increasingly reduced or eliminated, creating additional, recession-driven job loss. These disruptions are causing a redesign of work for higher productivity. This redesign includes workplace choices, skills development, and job design.

Understanding productivity: Understanding what drives productivity in complex, technical work has been a longstanding problem in economics and management science. The ability to measure changes in work practice using privacy-preserving technologies can help. This process starts by extracting a "tools signature" for a team from existing collaboration tools, such as Slack, Webex, and GitHub, as shown in Figure 11, and studying how it correlates to management reported productivity.[57] Over time, the discovery process can help develop an understanding of the components of productivity—the types of personnel, teams, tools, and collaborations—for highly complex technical work and adapt workplaces, practices, and policies.[58] An understanding might take the form of a computational model that predicts productivity based on a team's makeup, location, and tool use.

Figure 11:
Steps to measuring output-productivity in hybrid work.



Transforming the workplace: In order to transition work-from-home to hybrid work, workplaces need to be safe from infectious disease. Even pre-COVID-19, infectious disease in the U.S. employee population cost employers between $10 billion and $30 billion per year in paid absences. Employers also have a general regulatory obligation to provide workplaces "safe from hazards that are likely to cause death or serious physical harm." Using technologies that span hyper-local epidemiology, genomics, and edge computing, there is need to create risk assessments for worksites. Combined with surface and airflow sensing, employee testing, personal protective equipment monitoring, and optimizing for social distancing will lower risks in moving towards hybrid work.

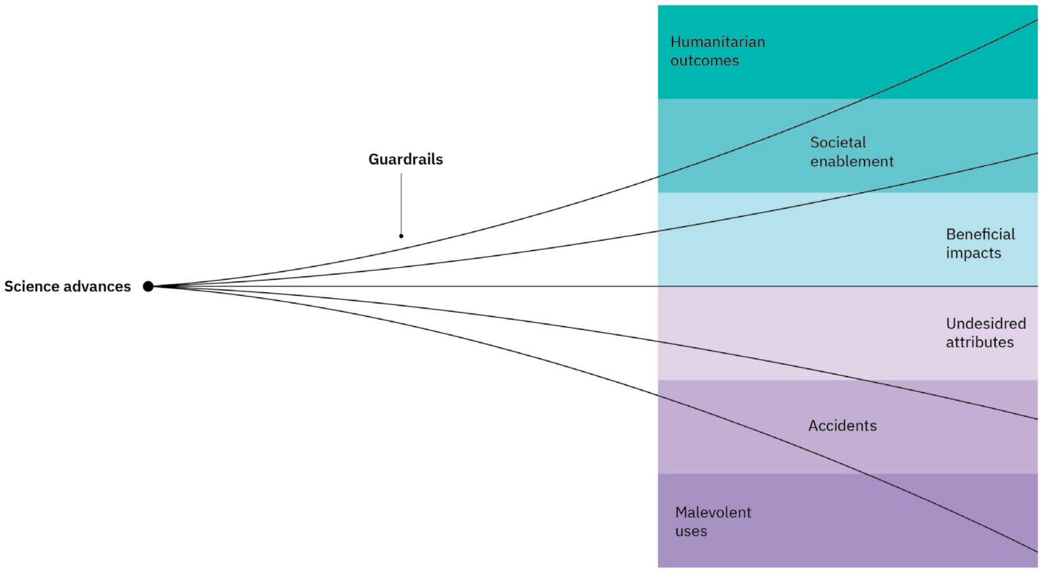
Transforming the workforce: The COVID-19 pandemic resulted in the largest labor market disruption ever recorded in the U.S. Firms eliminated millions of jobs, and while some of them will come back as economies reopen, others will not. Workers from marginalized groups have been disproportionally impacted by this disruption. Firms under financial stress are looking to restructure workforces for hybrid work, new skills, and greater productivity. Workers are in the process of retraining, working to develop skills that will be well-matched to the labor market. Using datasets like aggregate job postings, it is possible to create granular predictions of future demand for skills up to two years in the future.[59] This helps businesses understand whether the skills being developed are likely to meet future demands. At a national scale, as governments react to COVID-19-driven unemployment, this capability can guide job-creation programs.

Utopias and Dystopias

Scientific discoveries have created extraordinary positive impacts for society: curing diseases, increasing longevity, and improving quality of life. Unfortunately, advances in science can also have dystopian consequences through malevolent uses or other unintended effects.[60] Nuclear energy allows the most efficient production of electricity, with several hundred nuclear power plants alone producing approximately 10% of worldwide energy. Nuclear waste disposal, however, remains a tremendous challenge. Nuclear accidents have caused unprecedented devastation and the development of nuclear weapons causes geopolitical instability.

As discovery processes accelerate, we must be cognizant of potential outcomes—good and bad. As shown in Figure 12, this means establishing safeguards and solutions to ensure responsible advances to bring about beneficial impacts. Disruptive events like the COVID-19 pandemic bring large uncertainties. Predicting possible scenarios is essential for guiding effective response and ensuring preparedness for the near and distant future.

Figure 12:
Thoughtful guardrails are needed to maximize beneficial outcomes from accelerated discovery.



With recent scientific discoveries and technological developments, numerous challenges are emerging in areas such as CRISPR gene editing and neurotechnology. CRISPR gene editing has produced outstanding results in the treatment of congenital diseases responsible, in one example, for blindness in children. In 2018, the world was shocked by the revelation that the same CRISPR technology was used to edit embryos of two currently living babies. Germline editing can produce unpredictable consequences, which can propagate to descending generations. The editing of a single gene can influence multiple traits and the lack of understanding of all potential implications can lead to dangerous mutations.[61]

Neurotechnology is advancing rapidly. Breakthroughs have allowed people to recover movement after a debilitating brain injury and helped patients regain the ability to communicate after a degenerative brain disease. Neurotechnology is also at the verge of allowing the reading and writing of memories in living organisms. Human identity is constructed on experiences and memories, and thus, the misuse of this technology has the potential to alter the essence of what we are as individuals.

As we develop computing technology, we want systems that guarantee privacy and security, that are transparent and trustworthy, that ensure fairness and equity, and that are environmentally responsible.

Policies and Prevention

Technology advances that allow us to accelerate and scale scientific discovery have the potential to drive positive change for society. But change often elicits fear and resistance, which triggers blunt, sub-optimal, or poorly timed regulation. This context is further complicated by a widening technology-regulation gap, or "pacing problem," illustrated in Figure 13. Compounding this is the fact that two-thirds of U.S. Federal regulations are never changed after their inception, risking poorly applied guardrails informed by an outdated view of technology. Additionally, different countries impose different regulations, increasing compliance risks. This regulatory landscape poses a significant risk to accelerated discovery and its technology components and can hinder business growth.

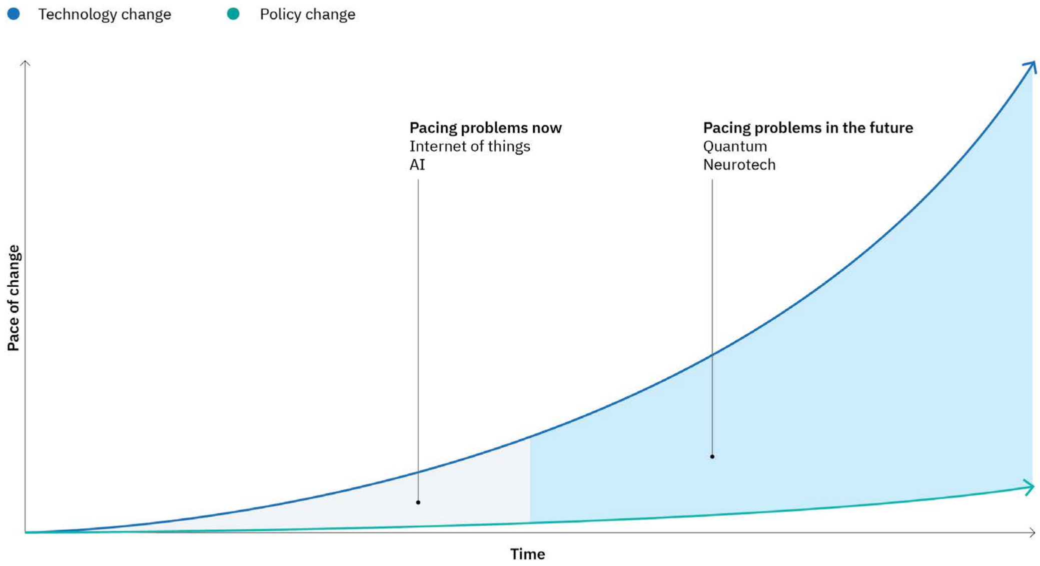


Figure 13:
"Pacing Problem" – innovation outpaces the ability of laws and regulations to keep up with technological change.

To address these concerns, we need a new values-based governance framework to provide guardrails to ensure innovations align with values, and to anticipate and inform regulatory trends.[62] Putting values into practice requires a "by-design" mindset, infusing privacy, security, and ethical considerations into our engineering and technology development—from the very outset. Such a governance framework must evolve with technology development while engaging multiple stakeholders throughout the technology lifecycle. Values must be validated and continuously assessed and enhanced through partnerships across industry, academia, research organizations, and other stakeholders. Through the implementation of innovation guidelines, along with assessment and oversight processes, this framework will anticipate potential impact and align innovations to values, aided by tools to mitigate risk and support alignment to values and regulations. One such tool is AI Fairness 360,[K]software that measures and reduces bias in datasets and AI models.

We can also apply scientific thinking to to the development of regulations. One way is to experiment, using regulatory "sandboxes," introducing policies locally and measuring the results before adopting them more widely.[63, 64] Data from these experiments and simulations can facilitate collaborative rulemaking between industry and government, leading to effective dialogue with advocates and influencers and resulting in precision regulation based on trusted information. This can shape the responsible use of technology and pre-empt unwanted over-regulation.

[K] AI Fairness 360: developer.ibm.com/technologies/artificial-intelligence/projects/ai-fairness-360/

References

Urgency of Science:

1 UN Sustainable Development Goals https://www.un.org/sustainabledevelopment/sustainable-development-goals/.

2 E. Duflo, “Field Experiments and the Practice of Policy,” *American Economic Review*, 110(7): 1952–1973, 2020.

3 T. Kuhn, *The Structure of Scientific Revolutions*, 2nd Edition, 1970.

4 K. Popper, *The Logic of Scientific Discovery*, 2005.

5 S. Leonelli, “Scientific Research and Big Data”, *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), May 29, 2020.

6 T. Hey, S. Tansley, K. Tolle (eds.), *The Fourth Paradigm, Data-Intensive Scientific Discovery*, Microsoft Corp, 2009.

7 L. Ruddigkeit, R. van Deursen, L. C. Blum, and J.-L. Reymond, “Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17,” *J. Chem. Inf. Model*. 2012, 52, 2864- 2875.

8 R. Stevens, V. Taylor, J. Nichols, A. B. Maccabe, K. Yelick, D. Brown, “AI for Science,” *Dept. of Energy report*, Feb. 2020.

9 H. G. Gauch, *Scientific Method in Practice*, Cambridge University Press, 2003.

10 Y. Gil, M. Greaves, J. Hendler, H. Hirch, “Amplify scientific discovery with Artificial Intelligence - Many human activities are a bottleneck in progress,” *Science*, October 2014.

11 M. Baker, “Is there a reproducibility crisis?,” *Nature* 2016; 533: 452–54.

12 S. Spangler, et al, “Automated Hypothesis Generation Based on Mining Scientific Literature,” Proc. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2014.

13 A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller, T. Laino, “Automated extraction of chemical synthesis actions from experimental procedures,” *Nature Communications*, 2020.

14 P. Kohli, “AI Will Help Scientists Ask More Powerful Questions,” *Scientific American*, April 16, 2020.

15 T. B. Brown, et al, “Language Models are Few-Shot Learners,” *arXiv:2005.14165*, May 2020.

16 P. B. Jørgensen, M. N. Schmidt, O. Winther, “Deep Generative Models for Molecular Science,” *Molecular Informatics*, 2018.

17 D. Schwalbe-Koda, R. Gomez-Bombarelli, “Generative Models for Automatic Chemical Design,” arXiv:1907.01632v1 [cs. LG] 2 Jul 2019

18 T. Maziarka, A. Pocha, J. Kaczmarski, K. Rataj, T. Danel, M. Warchoł, “Mol-CycleGAN: a generative model for molecular optimization,” *Journal of Cheminformatics*, 2020.

19 L. Himanen, A. Geurts, A. S. Foster, P. Rinke, “Data-Driven Materials Science: Status, Challenges, and Perspectives,” *Advanced Science*, 2019.

20 C. W. Coley, N. S. Eyke, Klavs F. Jensenz, “Autonomous discovery in the chemical sciences part I: Progress,” *arXiv:2003.13754v1*, Mar. 30, 2020.

21 C. W. Coley, N. S. Eyke, Klavs F. Jensenz, “Autonomous discovery in the chemical sciences part II: Outlook,” *arXiv:2003.13754v1*, Mar. 30, 2020.

22 S. Thomke, *Experimentation Works: The Surprising Power of Business Experiments*, *Harvard Business Review Press*, 2020.

Emerging Discovery-Driven Enterprise:

23 "COVID-19 and the Future of Business", Institute for Business Value, https://www.ibm.com/thought-leadership/institute-business-value/report/covid-19-future-business.

24 "Global Research and Development Fact Sheet", *Congressional Research Service*, R44283. https://crsreports.congress.gov, April 2020.

25 M. A. Aftab, Q. Yuanjian, N. Kabir, and Z. Barua, "Super responsive supply chain: The case of Spanish fast fashion retailer Inditex-Zara," *Intl. Journal of Business and Management*, 13(5), p.212, 2018.

26 S. Thomke, "Building a Culture of Experimentation," *Harvard Business Review*, March – April 2020.

Hybrid Cloud is Essential for Discovery:

27 JJ. Hellerstein, K. Kohlhoff and D. Konerding, "Science in the Cloud: Accelerating Discovery in the 21st Century" in *IEEE Internet Computing*, vol. 16, no. 04, pp. 64-68, 2012.

28 *European Open Science Cloud (EOSC) Strategic Implementation Plan*, EOSC Executive Board, June 2019

29 J. Fernandes, B. Jones, S. Yakubov, Andrea Chierici, "HNSciCloud, a Hybrid Cloud for Science," EPJ Web of Conferences 214, 09006 (2019).

30 P. Ivie and D. Thain, "Reproducibility in Scientific Computing," ACM Comput. Surv. 51, 3, Article 63 (July 2018).

31 M. Wilkinson, M. Dumontier, I. Aalbersberg, "The FAIR Guiding Principles for scientific data management and stewardship," *Nature: Scientific Data*, Vol 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18.

32 S. Hodson, "Turning FAIR into reality," *Final Report and Action Plan from the European Commission Expert Group on FAIR Data*, 2018.

Accelerated Discovery:

33 E. O. Pyzer-Knapp, T. Laino (eds), "Machine Learning in Chemistry: Data-Driven Algorithms, Learning Systems, and Predictions," Volume 1326, Nov. 2019.

34 P. Schwaller, T. Laino, T. Gaudin, P. Bolgar, C. A. Hunter, C. Bekas, A. A. Lee, "Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction," *ACS Cent. Sci.*, 5, 1572–1583, 2019.

35 P. Staar, M. Dolfi, C. Auer. "Corpus Processing Service: A Knowledge Graph Platform to perform deep data exploration on corpora," *Authorea*. September 16, 2020. DOI: 10.22541/au.160028088.82614622.

36 Ł. Maziarka, T. Danel, S. Mucha, K. Rataj, J. Tabor, S. Jastrzebski, "Molecule Attention Transformer," *arXiv:2002.08264v1* [cs.LG], 19 Feb 2020.

37 D. C. Elton, Z. Boukouvalas, M. D. Fuge, P. W. Chung, "Deep learning for molecular design—a review of the state of the art," *arXiv:1903.04388v3* [cs. LG], May 2019.

38 S. Takeda, et al, “Molecular Inverse-Design Platform for Material Industries,” *Proc. ACM KDD-2020*, August 2020.

Intelligent Infrastructure:

39 Y. Alexeev, et al. “Quantum Computer Systems for Scientific Discovery,” *arXiv:1912.07577v2* [quant-ph], Feb 2020.

40 D. Gil, W. J. Green, “The Future of Computing: Bits + Neurons + Qubits,” *IEEE International Solid- State Circuits Conference (ISSCC)*, Feb. 2020.

41 R. Chard, Y. Babuji, Z. Li, T. Skluzacek, A. Woodward, B. Blaiszik, I. Foster, K. Chard, “funcX: A Federated Function Serving Fabric for Science,” *arXiv:2005.04215v1* [cs. DC], May 2020.

42 P. Moritz, R. Nishihara, S. Wang, A. Tumanov, R. Liaw, E. Liang, M. Elibol, Z. Yang, W. Paul, M. I. Jordan, and I. Stoica, “Ray: A Distributed Framework for Emerging AI Applications,” *Proc. 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI ’18)*, Oct. 2018.

Communities of Discovery:

43 T. Okada, H. A. Simon, “Collaborative Discovery in the Scientific Domain,” *Cognitive Science*, Col. 2 (2), 1997.

44 H. Chesbrough, “From Open Science to Open Innovation”, *Science Business*, 2015.

45 J. Furman, P. Gaule, “A Review of Economic Perspectives on Collaboration in Science,” *National Research Council Study for the Science of Team Science*, 2013.

Future of Climate:

46 D. Rolnick, et al. “Tackling Climate Change with Machine Learning,” *arXiv:1906.05433v2* [cs.CY] 5 Nov 2019.

Future of Health:

47 J.A. DiMasi, et al., “Innovation in the pharmaceutical industry: New estimates of R&D costs,” *J Health Econ*, (47 2016), pp. 20-33.

48 R. E. Sherman, S. A. Anderson, G.. Dal Pan, G. W. Gray, T. Gross, N. L. Hunter, L. LaVange, D. Marinac-Dabic, P. W. Marks, M. A. Robb, J. Shuren, R. Temple, J. Woodcock, L. Q. Yue, R. M. Califf, “Real-World Evidence — What Is It and What Can It Tell Us?”, *The New England Journal of Medicine*, December 8, 2016,

49 S. Klonoff, “The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics,” *Journal of Diabetes Science and Technology*, Vol. 14(2) 345–349, 2020.

50 D. Parisi, et al., “Drug repositioning or target repositioning: A structural perspective of drug-target-indication relationship for available repurposed drugs,” *Comput and Struct Biotech Journal*, 18 (2020), pp.143-155.

51 N. Nosengo, "Can you teach old drugs new tricks?" *Nature*, 534, (2016), pp. 314-316.

52 S. S. Francis, L. W. Riley, "Metagenomic epidemiology: a new frontier," *J Epidemiol Community Health*, 2014.

53 P. M. Gundy, C. P. Gerba, I. L. Pepper, "Survival of Coronaviruses in Water and Wastewater," *Food Environ Virol* (2009) 1:10–14.

54 W. Randazzo, E. Cuevas-Ferrando, R. Sanjuán, P. Domingo-Calap, G. Sánchez, "Metropolitan Wastewater Analysis for COVID-19 Epidemiological Surveillance," medRxiv preprint doi: https://doi.org/10.1101/2020.04.23.20076679, April 29, 2020.

Future of Work:

55 E. Brynjolfsson, J. J. Horton, A. Ozimek, D. Rock, G. Sharma, "COVID-19 and Remote Work: An Early Look at US Data," *National Bureau of Economic Research*, Working Paper 27344, June 2020.

56 J. I. Dingel, B. Neiman, "How many jobs can be done at home?", *National Bureau of Economic Research*, Working Paper 26948, April 2020.

57 L. Yang, et al., "How Work From Home Affects Collaboration: A Large-Scale Study of Information Workers in a Natural Experiment During COVID-19," *arXiv:2007.15584*, July 2020.

58 D. Ford, "A Tale of Two Cities: Software Developers Working from Home During the COVID-19 Pandemic", *arXiv:2008.11147*, August 2020.

59 S. Das, S. Steffen, W. Clarke, P. Reddy, E. Brynjolfsson, M. Fleming, "Learning Occupational Task-Shares Dynamics for the Future of Work," Proc. of the *AAAI/ACM Conference on AI, Ethics, and Society (AIES)*, 2020.

Utopias and Dystopias:

60 N. Bostrom, "The Vulnerable World Hypothesis", *Global Policy* Volume 10, Issue 4, November 2019.

61 E. Sandel, "The Case Against Perfection", *The Atlantic Monthly*, April 2004.

Policies and Prevention:

62 D. Chandler, "The Rise of Data-Driven Governance," *Current History*, Jan. 2020.

63 D. Huitema, A. Jordan, S. Munaretto, M Hildén, "Policy experimentation: core concepts, political dynamics, governance and impacts," *Policy Sci* (2018) 51:143–159.

64 B. McFadgen, D. Huitema, "Experimentation at the interface of science and policy: a multi-case analysis of how policy experiments influence political decision-makers, *Policy Sci* (2018) 51:161–187.

IBM **Research**

Connect with our IBM team by visiting
research.ibm.com/collaborate