



THE DATA SCIENCE

INNOVATOR'S

PLAYBOOK

Today's top leaders reveal industry trends, career advice, and how to meet the world's most important challenges.

MEET THE INNOVATORS

Top data science leaders share insights on their **work**, their **careers**, and the **data science profession**.

The genesis of this ebook was a poll that Domino Data Lab took at its 2022 Rev 3 Enterprise MLOps conference, where we asked the more than 800 attendees who they thought was driving innovation in data science. Respondents wrote in the names of people they thought were advancing business outcomes and the profession in particularly innovative ways. We've interviewed their top nominations on everything from the direction of AI and machine learning to how data science is changing, and along the way, even solicited career advice for data scientists from these people who are at the top of their game.

A few things are clear from these interviews. Data science departments have seen tremendous growth, both in terms of the size of their teams and their centrality in answering critical questions for their businesses, and this growth is only expected to increase in the near future. This trend is accelerating, and these innovators expected even larger gains from the smart application of data science, especially AI and ML. This is entirely

consistent with other recent surveys, as well as with the other [REvelate survey findings](#), such as the belief held by 79% of Rev 3 attendees that data science, ML, and AI are critical to the overall future growth of their company, with 36% calling it “the single most critical factor.”

In life sciences, for example, ML models are no longer confined to the chemistry of a cure, but are used to investigate all facets of healthcare delivery—from the impact of COVID-19 on the ability to do clinical trials, to finding the best subjects for those trials, to investigating wait times and supply chain impacts on hospitals.

As Mona Flores, Global Head of Medical AI at NVIDIA, puts it, “There is a lot of opportunity to apply AI to the business of healthcare, because predicting and analyzing needs ahead of time can be critical to the work of doctors, nurses, and other caregivers.”

Similarly, in insurance, data science is now central in most business decisions, and data science departments have grown from a handful of people to teams of hundreds, including data scientists and data science leaders, data engineers, MLOps experts, and IT professionals, all working together to help their companies use data, AI, and ML modeling to make better decisions.

Enterprise MLOps is a critical factor anywhere data science drives innovation.

All of these data science leaders also mentioned the critical, growing role that MLOps plays in their operations, and how improvements in automating many critical infrastructure tasks are dramatically increasing their ability to do more innovative work.

As Google's Chief Decision Scientist Cassie Kozyrkov told us, "parts of data science are hard because they're annoying, but the amount of time spent fiddling with the tools is going down. This uncovers more opportunities in the parts of the work that are intellectually hard, not fiddly-hard. This means that data professionals will have less reason to define themselves by the tools they use, and more reason to focus on the intellectual essence of making data science useful."

For example, many pharma, health, and life sciences companies are now using Enterprise MLOPs beyond the realm of exploratory research, where it has long been important for its ability to boost collaboration to unlock life-changing breakthroughs and to see them through their lifecycles. Enterprise MLOPs is also becoming increasingly valuable for the easy traceability, reproducibility, flexible tool access, infrastructure management, and data-sharing framework that make model development, deployment, and maintenance possible at a large scale.

Finally, all of the top innovators we profiled were enthusiastic about the future of data science and the opportunities for people who love to use data to answer key questions about everything from cures for disease to the best ways to run a business. They were extremely generous in giving insightful career advice that many will find helpful.

Data science is a profession enjoying a meteoric rise, and many will find great success and satisfaction chasing the answers to some of the most important problems the world faces today. We hope you enjoy the thoughts they've shared with us, with you, and with the wider data science community.





WE INTERVIEWED...

01 Cassie Kozyrkov, Ph.D.

Chief Decision Scientist, Google

02 Glenn Hofmann, Ph.D.

Chief Analytics Officer, New York Life Insurance Company

03 Najat Khan, Ph.D.

Chief Data Science Officer and Global Head, Strategy & Operations for Research & Development at the Janssen Pharmaceutical Companies of Johnson & Johnson

04 Robert Nishihara, Ph.D.

Co-creator of Ray, and Co-founder & CEO, Anyscale

05 Mona G. Flores, M.D.

Global Head of Medical AI, NVIDIA

06 Andy Nicholls

Senior Director, Head of Statistical Data Sciences, GSK

07 John K. Thompson

Analytics thought leader, best-selling author, innovator in data and analytics

01

02

03

04

05

06

07

CASSIE KOZYRKOV

Cassie Kozyrkov, Ph.D., is the Chief Decision Scientist at Google.



THE INNOVATION

Democratizing artificial intelligence and inspiring business leaders to harness the transformative power of machine learning. Through research, education, and public advocacy for new applications of data science, Cassie Kozyrkov is advancing the causes of better leadership and decision-making in private business and government.



“There’s a **revolution** going on in the space of data science **tools**.”

CASSIE KOZYRKOV

“Domino Data Lab frees data scientists to focus on the most impactful and exciting parts.”

What's changing about innovation in data science as the profession matures?

Parts of data science are hard because they're annoying, but the amount of time spent fiddling with the tools is going down. This uncovers more opportunities in the parts of the work that are intellectually hard, not fiddly-hard. This means that data professionals will have less reason to define themselves by the tools they use and more reason to focus on the intellectual essence of making data science useful. It'll be easier to see that even though data professionals use the same tools, they're using them in very different ways for different reasons. It'll be more obvious that skills don't fully transfer within the data science professions.

You can't do data science yourself. You have to embrace it as a team sport. There are many different types of makers in our ecosystem, and there are very few people who can do every aspect of data science themselves. Those who can tend to be very, very expensive to hire. But in practice, you'll want to hire more than one data professional because of the sheer volume of data work in a modern organization...and if you're hiring multiple people in any case, I'd recommend embracing specialization instead of seeking several carbon copies of the impossible hire.

What are some of the factors driving the rise of data science as a source of innovation?

There's a revolution going on in the tools space, and this revolution is why the data science universe is expanding more and more rapidly. The concept of the data scientist as a tool user was a completely revolutionary thought just ten years ago.

For years, we data scientists have lacked real advocates. Because real advocates don't just sell the glamor and the buzzwords. They fight to improve the user experience and effectiveness of data scientists. They ensure that data scientists have the support, resources, and tools they need.

Today, we finally have real advocates, and the tools keep getting better. For example, Domino Data Lab has really taken our needs to heart. Their entire existence is about making the data science user experience lovable so data scientists can skip the chores and get to their real work. Data scientists shouldn't have to think about punchcards and vacuum tubes—and they shouldn't have to

think about spinning up clusters or setting up their data science environment either. Domino Data Lab frees data scientists to focus on the most impactful and exciting parts of their work. That's what advocacy for our data science community looks like, and I'm really excited by this; it's incredible.

The world is finally on board with understanding that "knowledge is power" belongs in the same bucket of ideas as "data is useful." And if data is useful, then data scientists are useful, and their work needs to be streamlined to extract value and accelerate innovation.

What advice can you give to other data scientists who are getting started in the profession?

Some data scientists worry that when all of the fiddly stuff goes away, the need for their labor goes away too. What's left of our jobs? Here's what I'd love for data scientists to grasp: What's left over after the fiddly stuff goes away...is a lot.

In the same way that a writer's work isn't the papyrus or pencil or typewriter, but something more essential, our work will boil down to patterns of thinking and mindset - what some people call "soft skills." I don't like that phrase, I prefer to say "the skills that are hardest to automate." These skills hide behind all the time spent with technology, but they're critical to success and innovation.

So I have two bits of advice. First, don't overfocus on the tools; they're a mayfly. Focus on the essence of what you do and, most importantly, why you do it. Lean in and embrace the version of data science that you do best. Second, your work is important - before accepting a job, insist that your intended employer must prepare an environment in which you can be successful. You're too valuable to be wasted on an employer who's not ready for you. Before saying yes, check that they have the resources, the environment, and the tools to allow you to do what you do best: fuel innovation and make data useful.

What do you consider your biggest success in data science? How about in life?

Changing Google's organizational culture to be more data-driven when needed, and less data-driven when rigor would be a waste of resources, helping the company be more efficient and thoughtful in its approach to data. It's an incredible challenge and honor to be able to improve the practice of data science at Google scale.

As for my biggest success in life, it's the same as what I'd wish for every person on the planet: taking the time to understand oneself well enough to design personally fulfilling goals rather than pursuing a life that would be a better fit for someone else. And that also means saying no to some things that everyone around you seems to want. Every single person is unique; we're all weirdos in our own way, so real success has to start with designing our own lives instead of copying the lives of others.

You have been very active in promoting education in data science and adjacent disciplines. What should we do to better educate and prepare people to work in the data science ecosystem?

Well, given that I do spend a lot of my free time creating resources exactly for this purpose, I'd say a great place to start is [my course on YouTube](#) and [my blog](#). But where we all need even more effort and resources is data leadership - there are relatively few experts in the data science decision-maker's craft, so I hope they will join me in recognizing it as a discipline in its own right and sharing our wisdom as widely as possible.



01
02
03
04
05
06
07



GLENN HOFMANN

Glenn Hofmann, Ph.D., is the Chief Analytics Officer at New York Life Insurance Company.

THE INNOVATION

A recognized thought leader on developing effective data science teams, Glenn built a successful, at-scale data science organization within New York Life, a Fortune 100 company that is America's largest mutual life insurer. New York Life's Center for Data Science and Artificial Intelligence (CDSAi) is fully immersed in the core of the company's business. Through the team's breadth of statistical and machine learning expertise, CDSAi enables real-time, model-based decision making across the company and helps New York Life thrive even in times of great uncertainty, including the COVID-19 pandemic.

“Domino enables flexible, real-time deployment of models.”

GLENN HOFMANN

What are some of the factors that are fueling the rise of data science as a force in solving important business problems?

Several developments have fueled the rise of data science. First, algorithms have gotten better over time, particularly in areas like natural language processing. Second, both the infrastructure and software to deploy models into production have vastly improved, and that drives the impact our models can have on the business. For example, at New York Life, we have both an on-prem and cloud-based computing environment that is used for all aspects of the data science lifecycle. That includes Domino, which enables flexible real-time deployment of models in any of the popular programming languages, along with monitoring and reporting of our models in production.

The algorithms, infrastructure, and technology have all helped tremendously in encouraging adoption of data science and thereby generating value. Through operationalizing data science and producing results more quickly, others are able to see tangible improvement in their parts of the business, and they're more willing to support and recognize the value of data science moving forward.

What's changing about innovation in data science as the profession matures?

There is the idea of the 'full stack' data scientist who executes all aspects of the data science model lifecycle, from building the data pipelines to developing the machine learning model to deployment and subsequent

monitoring of the model. That may work very well in a smaller team or one with a narrow focus. For an enterprise data science team operating at scale, however, defined roles and responsibilities are important, and it is not efficient or feasible for a single data scientist to manage all aspects. At New York Life, we have dedicated teams for MLOps, operations, change management, model monitoring and governance, and program management who work alongside the data scientists. These specialties have enabled us to achieve scale. The challenge we've experienced as we have grown is ensuring collaboration and ample communication throughout the model development process to produce successful products for our business partners. While this has taken some time and training, as we have grown so quickly, we're proud of the level of collaboration and communication taking place today.

Can you talk about how you evangelized the use of data science at New York Life in the early days?

We were fortunate to have courageous early advocates of data science in different areas of the business. Once we were able to showcase those early successes, other areas recognized the value that data science brings.

For example, we had some early successes with our marketing team. Then, within our finance department, we were given the green light by our chief actuary to implement the mortality assumptions model using a sophisticated statistical methodology that we developed in partnership with the actuaries in

that area. That led to higher engagement in other areas such as Agency, our sales distribution area, where we now have models deployed in production which enable real-time decisions to be made.

We also foster an active data science community through events, speakers, training, and education, which helps bring a broad understanding to all levels of employees at New York Life on how data science models work and how the models can provide value. We taught a popular data science workshop specifically targeted to New York Life's executive officers. Through 15 workshops thus far, we have been able to reach half of this group and provide a background on what data science is at New York Life, what the executive's role is in signing on to a data science project, how data science can impact their bottom line or simply help with make better and faster decisions, and showcase examples in different business areas. The workshops have been very effective in ideating new data science projects.

What's the biggest obstacle to using data science to solve big problems?

The biggest obstacle is showing a business partner that it is worth investing the time to develop a rigorous model based on historical data to make model-based decisions, as opposed to a decision based on intuition and consensus which can be made faster but far too often proves to be lacking. We as the data science community need to help our business partners shift their mindsets that this upfront investment in time and planning to build a machine learning model will reap benefits in the future as the predictions are more accurate and thereby enable better decisions to be made. Models are monitored to detect data drift and are more easily updated to reflect changes over time. We need to show that it is a valuable use of time for future benefit.

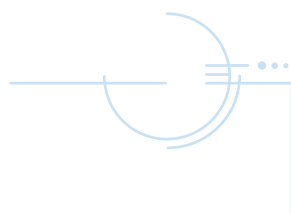
Do you have advice for other companies that want to grow their data science departments and impact?

First, recognize that it takes a certain critical mass of infrastructure, resources, and people to make a successful data science team. That means you must scale up your efforts and build the infrastructure needed to deploy models. This way, you avoid getting stuck in a loop where real production work never takes place and models are only scored on an ad hoc basis or used only for insights.

Second, remember that data scientists and everyone else within a data science team—including the MLOps engineers, operations, and program managers—want to see that their work is making a difference. Ensuring that models are deployed and used by the business will help with both attracting talent to your team and retaining talent. Also, to attract talent, you need to provide the right resources, career paths, training programs, and learning stipends.

Third, educate your business partners on the data science process. Remove the mystery and clear up misconceptions. The more your business partners understand and take ownership of the models themselves, the more interested they will be to work with your data science team. If, on the other hand, you communicate infrequently and only present a complicated final model they don't feel ownership in, you may not see the type of growth you'd like.

The supportive environment within New York Life has allowed us to develop a highly functioning data science team that we're proud to say now operates at scale.



01
02
03
04
05
06
07

NAJAT KHAN

Najat Khan, Ph.D., is the Chief Data Science Officer and Global Head, Strategy & Operations for Research & Development at the Janssen Pharmaceutical Companies of Johnson & Johnson; and Co-Chair of the Johnson & Johnson Data Science Council

THE INNOVATION

Recently named one of Endpoints' 20 Under 40, Najat Khan, Ph.D., is responsible for driving the overall R&D strategy for Janssen and identifying the areas where data science and digital health can have the greatest impact. She has grown the R&D data science program at Janssen from a handful of models and a small team to more than 100 projects that are accelerating, supporting, and expanding the company's R&D pipeline—everything from targeted cancer therapies to vaccines.

"A smart idea can come from anywhere. Ultimately, it's all about bringing together the best ideas, skills, and talent together to generate maximum impact for patients."



“ A smart idea can come from anywhere. ”



NAJAT KHAN

“When people start to see what is **possible**—it can generate huge **momentum**.”

What are some of the toughest challenges of innovating in pharma through data science?

Data science is fundamentally upending traditional ways of doing things across all sectors, including the pharma space. And change is always difficult, in any field. But when people start to see what’s possible—how data science can transform how we innovate and, most importantly, how we deliver for patients in need—it can generate huge momentum. That’s why my team has focused on looking across our portfolio and identifying areas where we can generate quick wins—to really move data science from theory to action to, most importantly, patient impact.

As you’ve grown your team so rapidly, what are some of the lessons you’d like to share about innovation on a large scale?

Individual expertise is important, but the team effort is even more important. It doesn’t matter if you’re an SVP, an analyst, inside the walls of J&J or outside the company. A smart idea can come from anywhere. Actually having the humility and open-mindedness to accept that is something I push a lot in my teams, because that’s the only way you truly collaborate—not just with the person who’s across the table, but being open to all kinds of different ideas from various sources.

Ultimately, it’s all about bringing together the best ideas, skills, and talent to generate maximum impact for patients. This is also one of the key reasons why

I’ve prioritized diversity on my team. My team comes from all over the world, and is nearly 50% women, in a field that is traditionally male-dominated.

How is having a ramped-up data science program improving innovation?

Let me try to make it real with an example. Imagine a patient that has unfortunately been diagnosed with cancer. As we drive towards precision medicine, therapies are being developed that have better outcomes for patients with a specific mutation in their cancer. Now imagine that this patient has a specific mutation for which there is a targeted therapy.

Being able to deliver the right therapy for the right patient is the holy grail in medicine. However, currently, these mutations are not picked up in a routine way because they are new and therefore are not sequenced on a regular basis. But every patient gets a tumor biopsy. So what we have done is used data science to digitize those biopsy slides, and then we use machine learning models to predict which mutation a patient might have, just using those images.

This is precisely what we’re doing in the bladder cancer space, and we recently began a [collaboration](#) to deploy this innovative algorithm to screen patient data from 100+ clinical trial sites globally to help facilitate clinical trial enrollment. The goal of this is to find—and find earlier—patients who could benefit from a targeted therapy and may otherwise have been missed, which could translate to a significant impact for patients. To me, that’s the true impact of data science at J&J and beyond.

01
02
03
04
05
06
07



ROBERT NISHIHARA

Robert Nishihara, Ph.D., is the Co-founder and CEO of Anyscale and the co-creator of Ray.



THE INNOVATION

Co-created Ray, software for easily developing Python applications that scale across hundreds of nodes or GPUs, making possible new innovations that were previously infeasible or even unthinkable. Co-founded and leads Anyscale, which has created a managed platform for easily running distributed applications with Ray.



“With Ray, you could build an innovative new search engine in a day.”





ROBERT NISHIHARA

“A lot of innovation
is in simplifying the
user experience...”

What do you see as the most interesting trends in data science?

The proliferation of deep learning in areas like language and vision will allow a different kind of user interface to arise. We'll be able to interface with applications by talking to them. That will open up an entirely new set of possibilities.

How will that change the practice of data science?

Today, to do AI and data science, you have to build a lot of infrastructure for scaling. AI is incredibly computationally intensive, and that is only going to grow. To enable more people to succeed with AI, we're going to need to make it easy to build scalable applications without becoming an expert in distributed systems. That's what we're working on with Ray and Anyscale.

One of the trends we're seeing is that more and more data and applications will have to be distributed and parallelized, and there is a lot of infrastructure work on the critical path.

For example, today if you want to build an innovative new search engine, it's a huge project. There's a lot of data processing, you have to do web crawling

to get the pages, you have to do data processing to extract the key words and build the search indices, you need to train ML models to rank pages, as you need to do serving to handle queries. Every single one of these components needs to be scalable, and it's a tremendous infrastructure lift. One of our goals with Ray is to enable developers to build scalable applications like that in a day without any knowledge of distributed systems. We're going to enable developers to reason only about their application logic.

What innovations would have to happen to realize that vision?

One major area where more innovation is needed is serverless computing, where developers don't have to think about the hardware they're running because it's been abstracted away for them. Some work has already been done in this area; AWS Lambda has done it for functions, and BigQuery has done it for SQL, but it needs to be done for all applications.

You also need to hide machine failures, so people can write code as if there won't be any hardware failures. They need to be handled invisibly to the user. A lot of the innovation is in simplifying the user experience.

And one more innovation that's really important is a seamless multi-cloud experience. When you build on top of Windows, you don't have to think about the underlying PC manufacturer, but if you build a scalable AI application in the cloud, you absolutely have to think hard about the cloud provider that you're running on. That will change.

Do you see a trajectory of convergence between parallelization frameworks such as Spark, Ray, and Dask, and traditional databases?

There will be some convergence, like we're seeing now, but it won't go all the way. The generality of apps people will build will still require general purpose programming frameworks. I don't expect everything to be a database at the end of the day.

“...That's not just one **breakthrough**;
it's a lot of little ideas.”

01

02

03

04

05

06

07

MONA FLORES

Mona G. Flores, M.D., serves as the Global Head of Medical AI at NVIDIA.



THE INNOVATION

As a surgeon, did groundbreaking work in applications of medical AI, including a pilot at 20 hospitals that proved the advantages of using federated learning with privacy-protected data to train models to recognize diseases such as COVID-19 and predict which treatments will be needed.



“Monitoring is critical to keeping models working as they did in the beginning. That feedback loop is essential.”

MONA FLORES

Which applications of AI do you think will have the greatest impact on our healthcare system in the near future?

Medicine today is an art, not a science. Today, the way we treat health is that we often have to throw the kitchen sink at a problem. Then we titrate, we try different things, add new treatments, and try to help. The reason we have to do that now in medicine is because we know very little; it's as if we're looking at a puzzle, and we have so few pieces. Some pieces, we understand what they look like very well, but there's so much missing, and data science can help us construct those missing pieces. What we want to do is to personalize and target treatments at specific individuals with specific needs, based on data.

For example, if we get negative results from a study on a drug, how do we know if the drug showed no effect because it really doesn't work, or if we just studied the wrong population because the drug only works on people with certain traits? We are at the dawn of understanding how to innovate with better treatments, drugs, and even changes in healthcare administration. AI models will help us answer these questions.

There is a lot of opportunity to apply AI to the business of healthcare, because predicting and analyzing needs ahead of time can be critical to the work of

doctors, nurses, and other caregivers. As a surgeon, I used to sometimes get very frustrated because we were ready, but an entire surgical team had to wait because an instrument was still being sterilized or a piece of equipment wasn't available. AI can be used to make sure that things like that happen very rarely.

Also, some of the frameworks being used to develop "digital twins"—building realistic replicas of hospitals or operating rooms so you can do simulations—could have a big impact on innovation.

What's the most interesting or important medical AI project you've worked on?

Federated learning, which helps us when data is distributed in many different places or when there isn't very much data we can use. We can find and use that data to train our models—even then it's protected by privacy restrictions—because nobody actually sees protected data. We had to show our proposal to many different people and ethics boards, and educate them on its inherent privacy-preserving features. In the end, we were able to use data from 20 hospitals across the globe to train an FL model called EXAM.

Following a few proof-of-concept projects, we leveraged the NVIDIA Federated Learning SDK to do the first real-life federated learning training of a clinical model at a large scale. We retrained a COVID-19 triage model that was originally developed by Mass General Brigham. The result was a more performant model that was also more generalizable. The EXAM model predicts the future oxygen requirements of COVID-19 patients using inputs from the Electronic Health Record (EHR), such as lab values and vital signs, along with chest X-ray images. If we know what equipment and treatments people will need, we can often have it ready to help them when they need it. It was the first federated learning model of its kind and size. Being able to show that we can train AI models, while still protecting privacy by using federated learning, will change what we can do with AI and ML in medicine, and will spearhead many advancements to come.

How important has MLOps been in your work?

MLOps is absolutely important, if you want to apply models in real life. It can help with measuring model and/or data drift. You don't just train a model and then you're done. It's a very iterative process. MLOps helps with monitoring your models to ensure that they keep working as intended so that they do not become obsolete. This feedback loop is essential.

MLOps also helps provide an audit trail of sorts, for when you might need to reuse parts of a previous model. Or, you might need to show someone how the model was constructed or what data it used. This is especially important when applying for regulatory approval.

MLOps is really a science by itself—to manage that process—so you can take a model to a regulatory body and show how it was developed, to get approval, and continuously measure model performance.

“Showing that we can use federated learning and still protect privacy will really change what we can do with AI and ML in medicine.”

01
02
03
04
05
06
07



GSK

ANDY NICHOLLS

Andy Nicholls is the Senior Director and Head of Statistical Data Sciences at GSK.

THE INNOVATION

Uses data science to speed clinical trials and approvals at one of the world's largest pharma companies, and contributes to a standard set of tools called "The Pharmaverse" that can be reused to help streamline the drug approval process.

“For data science to have an impact, you have to **change a decision.** Otherwise your work isn't data science, it's **just a report.**”

ANDY NICHOLLS

Where do you think data science stands to have the most impact in the near future?

I'd love to be able to say that we have targeted medicine, targeted specifically to individuals or cohorts of individuals. If we could get to that, that would be fantastic, but it's hard to say if that's five years away or 15.

I also think data science is helping to change the way people think about how they collect and store information. Ten years ago, people thought that data science was a magic wand that would solve all their problems. They didn't understand that having hundreds of disparate data sources that are not well connected makes it difficult to answer business questions, even with data science. Now they are beginning to understand what needs to be in place to realize the potential for data science.

What's the biggest problem data scientists face today?

Data engineering is a big problem. Whenever you start a new project, the data and the data sources are almost never as good as you'd initially hoped, so there's always a need for good data engineering. I also think there's a growing recognition that there are lots of roles within data science. Data engineering is a critical one, and it's becoming more valued.

How has the data science profession changed over the years?

Many businesses have now come to understand that there are no "data scientist unicorns." Sometimes I even think there's no such thing as a data scientist. There's data science, and within that discipline there are many different roles that contribute to the successful practice of data science. You need data engineers and data scientists, as well as other roles such as visualization specialists. They all have valuable skills, and they all need to collaborate.

What's the importance of MLOps in companies that are leaning into data science?

For data science to have an impact, you have to change a decision. Otherwise, your work isn't data science, it's just a report. The role of MLOps is to get us to the point where we can analyze our data and use it easily and effectively in everyday decision-making. So if your model starts to drift away from the truth and it refits itself and you've got a new, trustable model deployed—with the help of MLOps—then we're winning.

Could you talk about the pharmaverse?

The part of data science that people get excited about is machine learning, vision recognition, things like that. But for us, data has an impact in the approval process, where we have to prove that our drugs work and are safe. We reuse many of the same tools for analytics, visualization, analyzing safety data, and documentation over and over. But now with data science, we have visualization tools, we can combine sources of data, and we can do things like share the database with the regulator. The pharmaverse is an idea based on looking at how data can be reused, and creating a bunch of vetted tools and results that streamline many parts of the end-to-end process of getting drugs approved.

How important has MLOps been in your work?

COVID-19 presented huge difficulties for the world and for our industry. Many clinical trials effectively shut down or couldn't start, because of its impact on healthcare systems around the world. Unless it was a COVID study or cancer

study, it was almost impossible to proceed, so most of our black box models went out the window. We needed to find out what the impact on clinical trials—and our ability to develop new medicines—would be for different actions taken in the pandemic.

So we wanted to ask, “what if you take any study in our portfolio that’s doing trials in three or four countries, and you need to know the impact, for example, of what would happen if a country like France opened up all of its schools? How would that action affect our ability to do trials?”

We had to take internal data, combine it with external data—such as from health authorities and government sources—apply academic models, and come up with an app. Our clinical teams could then go into the app and understand a variety of what-if scenarios and their impact on our ability to complete clinical trials in the pandemic.

I’m proud of it because we did it very quickly, and it met an important need. And it demonstrates the power of the data science profession to have a real-world impact on people’s lives around the globe.

The opinions expressed in this profile belong to Andy Nicholls and are not necessarily those of GSK.

“There are many different roles that contribute to the successful practice of data science.”

01
02
03
04
05
06
07

JOHN K. THOMPSON

John K. Thompson is an analytics thought leader, best-selling author, and innovator in data and analytics.

THE INNOVATION

Built and led data science teams to produce award-winning applications in telecom, microbanking, and medicine, such as pioneering work in managing and treating disease. Developed a Center of Excellence of more than 400 members and 14 Special Interest Groups (SIGs). Has written extensively on how to build and manage data science teams.



“What you **become at work** will have a direct and dramatic impact on your **personal life.**”

JOHN K. THOMPSON

What's the most interesting trend in data science innovation today, in your opinion?

First, explainable AI, which is important because it could unleash more of the power and innovation that's possible in data science. Currently, data science is constrained in when and where we can utilize our most powerful analytic approaches, because we cannot explain how those algorithms, applications, and systems make decisions in the many variations of neural networks that we have at our disposal. We cannot yet use systems based on neural networks—nor develop applications using them—in regulated industries such as pharmaceuticals, finance, utilities, and many more specific use cases. But the progress is steady and impressive. Due to developments related to explainable AI, we'll soon be able to use the vast majority of our analytical arsenal, including neural networks, on any problem we want, as long as the reasoning and decision logic is transparent.

Second, we're starting to move beyond correlation to causality, thanks in large part to the work of Dr. Judea Pearl. In data science, we're very good at providing high levels of accuracy of correlation, but we are not very good at delivering reliable and accurate measurements of causation. The development of new approaches to deliver repeatable, reliable, and accurate levels of causation will be a revolution in analytics. We'll have new systems that will

automatically be able to search a high-dimensional space to determine the most relevant and powerful causal features or factors that can improve the efficiency and effectiveness of all analytics. This will change the landscape of how we can leverage and use analytics in a fundamental way.

What is the most important challenge data science has helped you solve?

Two of the most satisfying problems that some of my teams solved were:

- Helping a national telecommunication carrier expand their operations and offerings in the developing world and to offer micro-banking services to subsistence level farmers in multiple countries. The telecommunications firm won multiple humanitarian awards for the work and results.
- Helping a hospital system reduce the incidence of post-surgical sepsis by over 74% in a three-year period by developing a predictive analytic environment that surgeons could consult during surgeries to help them close incisions with the most effective closing procedures, to ensure a healthy recovery without complications such as sepsis.

Both applications made lives better for the target populations and the staff members serving those individuals.

What is the importance of MLOps in organizations trying to scale data science?

MLOps is a solid innovation in the field of data science. MLOps streamlines and eases the previously tedious and manual processes related to the development, testing, and deployment of analytical models, making the process more automated and reliable for data science teams. As MLOps continues to mature and as more functionality is subsumed into the areas of development, testing, automation, and implementation, MLOps makes data scientists' jobs easier and helps them focus on areas of higher value-add.

What career advice can you share with young data scientists, mid-career pros, and data science leaders?

For everyone: don't work for jerks. And one of my MBA professors taught us that what you become at work will have a direct and dramatic impact on your personal life. The two cannot be separated. If you are rude and unbearable at work, you'll be that way in your personal life.

Here's what I recommend for data scientists in different stages of their careers:

Young data scientists

- Be curious.
- Learn every day.
- Take chances.
- Fail in a grand manner.

Mid-career professionals

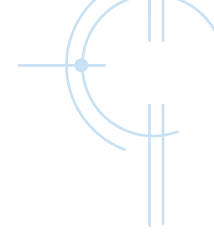
- Strive to take on ever-increasing levels of complexity, do not become complacent and solve the same problem over and over and over.
- Help younger staff members.
- Help your community.
- Be generous with your time and knowledge.

Seasoned data science leaders

- Be aware that you don't know as much as you think you do.
- Learn from younger staff members.
- Give back to the data science community.

“We'll soon be able to use the vast majority of our **analytical arsenal** on **any problem** we want, as long as the **reasoning** and **decision logic** is the same.”

Where to find **Enterprise MLOps** resources to promote data science **innovation** in your organization



As we've seen in these stories, data science is now a driving force for innovation at many of the world's most sophisticated companies. Clearly, ML models are helping companies to extract value from the huge volumes of data they collect, helping to solve some of the biggest problems in businesses, ranging from insurance to medicine and beyond.

All of the top data science leaders interviewed for this white paper cited the importance of Enterprise MLOps in helping them develop their innovative programs and data science departments. In particular, they emphasize the role of MLOps in helping data scientists develop their work in the best, most collaborative environments possible. As Google's chief decision scientist Cassie Kozyrkov puts it, Enterprise MLOps gets rid of a lot of the "fiddly stuff" involved in setting up and maintaining data science environments.

Video Resources

For further exploration, we have selected further video resources highlighting the innovators featured here:

- Andy Nicholls on [scaling R](#).
- Cassie Kozyrkov's insightful ML course on [YouTube](#).
- Glenn Hofmann on [leading data science teams to excellence](#).
- John K. Thompson on [data science teams](#) and [model-driven businesses](#).
- Mona Flores on [medical AI](#).
- Najat Khan on [unleashing exceptional performance](#).
- Robert Nishihara on [data science and distributed computing](#).

More Enterprise MLOps Resources

Domino Data Lab provides a number of resources for data science leaders and practitioners, and for IT leaders who support and promote data science programs, including the following:

- A [blog](#) featuring technical content, thought leadership, and strategic content on the effective use of data science in all facets of business.
- Specialized microsites for specific types of business, including [pharma and life sciences](#), [insurance](#), [financial services](#), and [more](#).
- Information on the [Domino Enterprise MLOps Platform](#).
- [Case studies](#) documenting how companies like yours have employed Enterprise MLOps to attack their biggest challenges and realize outsized returns.
- A [wealth of other resources](#), including white papers, an [ROI calculator](#), and free analyst reports.
- A white paper introducing [Enterprise MLOps](#), a holistic approach to scaling the production of models across modern enterprises..
- The [Data Science Leaders podcast](#), featuring in-depth conversations with executives across industries.

Domino specializes in helping some of the world's most sophisticated companies unleash the power of data science to address their most important challenges.





About Domino Data Lab

Domino Data Lab provides the Enterprise MLOps platform trusted by over 20% of the Fortune 100. Our products enable thousands of data scientists to develop better medicines, grow more productive crops, adapt risk models to major economic shifts, build better cars, improve customer support, or simply recommend the best purchase to make at the right time.

At Domino, our mission is to unleash the power of data science to address the world's most important challenges.