As Seen In
THE STANDARD
NEW ENGLAND'S INSURANCE NEWSMAGAZINE
Published Since 1865 Boston, Mass.
September 5, 2025 Vol. 296 No. 18

# Building a Data Science Culture that Ships Models and Meets Regulatory Demands

*Data Scientists, Actuaries Working With Data on a Scale Never Before Seen*

Data science in the insurance industry has experienced a dramatic acceleration from traditional actuarial methods to a modern approach, defined by large amounts of granular data from numerous new sources. The shift is creating a convergence of skills: Actuaries are expected to adopt data science tools, and data scientists need to develop a deep knowledge of insurance risk. Panelists in a recent Posit webinar discussed how to create models that meet regulatory demands without getting stuck in the research stage, how to use generative artificial intelligence (AI) to gain efficiency and how to stay ahead in the "arms race" against AI-driven fraud.

In the past 10 years, actuaries have shifted from relying on tabular, summarized data sets to working with much more granular data sets. There has been a technical shift from using purely statistical methods to a blend of machine learning, predictive modeling and — to a lesser extent — deep learning in some cases, according to Kshitij Srivastava, director of technology, Milliman.

The cultural shift has been just as big. "Actuaries are being asked to code, and data scientists are being asked to understand the domain deeply to understand insurance risk. That's not always comfortable. You're trained in a certain area … but traditional boundaries are merging," said Srivastava.

Actuaries and data scientists are working to bridge the gap between highly regulated, reproducible actuarial processes to more experimental, iterative machine learning workflows, noted Srivastava. "Culturally, I think the actuary versus data science mindset is breaking down. I feel like both are doing the things that the other used to do."

One of the biggest challenges is that neither data scientists nor actuaries are data engineers, and both are handling data on a scale and in formats that they have not had to before, according to Jamie Warner, managing director, data science and pricing, Plymouth Rock Assurance.

"When we think about machine learning and other tools — even the bolt-on tools that you can buy — there's a bigger question of data itself as its own product that we haven't had to deal with historically. With the generative AI improvements and how we can pull data out of things like PDFs, there's a whole employment segment in the insurance industry (in terms of getting and placing data in people's hands) that didn't exist five years ago," said Warner.

## Current Technology Revolution Versus Prior Shifts

How does the current technology shift in data science and insurance compare to previous waves, such as moving to the cloud? For Warner, the exciting part about the most recent set of tools is the space they apply in. The biggest barriers to adopting new technology are regulatory and ethical concerns. "It's really hard to make changes quickly because of the regulatory catch-up. When I think about the use of the newer tools coming out, like generative AI large language model tools, they are applied

S·P 1865 Standard Publishing Corporation
www.spcpub.com

in places where we don't need to worry as much about regulatory oversight or other complications," said Warner.

There is not a lot of regulatory concern about pulling data out of PDFs, for example. A lot of advances target getting information more quickly and helping employees who do frontline work, such as claims adjusters.

"Obviously, we have to worry about hallucinations and [having a] human in the loop, but I think insurers have been able to adopt [technology] much more quickly, because it is not hitting those core areas of pricing or regulatory boundaries," said Warner.

"The foundation of a lot of regulated industries is making sure you have good, solid statistical evidence for what you're doing. And statistics are still extremely important," said Adam Austin, director of data science, price and risk analytics, the Hartford Insurance Group. He described the early conception of data science as a Venn diagram with statistics, computer science and business knowledge overlapping. In the last several years, the computer science part of the job has changed the most, according to Austin. He explained that the field has evolved from doing one-time analyses to building models that answer a single business problem.

As data scientists build digital products, they have to think about the distribution of the insights and how systems work. "It's not just focusing on what data we have, what insights we generate, what value they add, it's all of these operational questions: What systems can generate the insights? Which systems are going to consume them? What scale do I need to produce and share these insights?," said Austin. "It's really much more of a software engineering mindset that sets you apart as a data scientist in insurance these days, and it really is about enabling the right decisions."

## The Constant Battle Against Insurance Fraud

As insurers adopt more sophisticated, data-driven processes, fraudsters are leveraging new technology to orchestrate more complex scams, creating what panel moderator Nick Rohrbaugh, senior product marketing expert, Posit, referred to as "an arms race." Insurers are using data science to deploy advanced analytics to identify fraud patterns.

For example, fraudsters could use AI to generate an image of vehicle damage that did not occur and submit a claim for it. "That's a big danger for us," said Austin. One solution would be to use AI to detect if an image is real or AI generated. However, Austin is concerned that as AI-generated images improve, detecting them may become too difficult.

Data science may not hold the answer to the problem. Austin suggested that taking a human and product-centered approach might work better. A solution might be insurers sending adjusters to auto body shops for damage assessment as well as incentivizing insureds to use telematics programs, which can detect accidents, noted Austin.

> *If someone thinks they have a great data scientist but that person is not willing to document their work, they're not a great data scientist.*
> *— Jamie Warner, Plymouth Rock Assurance*

When discussing insurance fraud, people tend to think about one individual taking a fraudulent action; however, insurance fraud often happens on a very large scale. Rather than one person or one physical therapy center committing fraud, it is multiple physical therapy centers taking part in a fraud ring. Having more data has allowed insurers the ability to detect network patterns better, noted Warner. "Fraud is very tricky to predict," she said.

## Regulators and the Modern Tech Stack

The modern data stack has become increasingly complex with open source tools like R and Python, major cloud platforms and other specialized vendor solutions. The mix of tools can present IT and operational challenges, ranging from reproducibility for regulatory review to hindering collaboration among teams. The platforms, languages and frameworks that insurers rely on to shift a model from an idea to production are key.

Data scientists are increasingly moving to the Python programming language as part of their collaboration across the end-to-end stack, according to Austin. Interestingly, he thinks the choice of tools is less relevant for data work. "In terms of model development, major programming languages [like Python] are entry points into more powerful open-source libraries. Successful teams are not going to worry as much about specific technology brands as they are about ensuring that they have the right data, that they understand the domain of the problem, and they're building reproducible work."

Process culture is as critical as the technology and workflows. Srivastava noted that actuaries, data scientists and IT teams have different timelines and risk tolerance. "Data scientists want to build models quickly, iterate quickly. IT wants stable systems. Actuaries want explainability. The governance frameworks take into account the inter-collaboration between these three different business units. That's the tricky part rather than the technology itself."

As models become more complex, it becomes more difficult to make sure they are transparent and explainable to regulators, other stakeholders and customers.

One of the biggest issues Warner sees is problems with a lack of documentation. She noted that projects can switch back and forth from different programming languages, like R and Python, but if the documentation is good, it should not be a problem. Warner explained that there are a lot of coding tools that can convert a high percentage of code from one language to another fairly easily.

"We're not just moving model to model. You might be iterating on a model, and you will iterate on it next year, but the packages you used before aren't the best ones to do it now. Or the cloud infrastructure you have now allows you to pull in much more data than before." Because technology changes so quickly and people forget certain aspects of projects once they are finished with them or if the project is paused, Warner is adamant that her team uses proper documentation to describe the different steps they are taking and why they are using one package versus another.

Warner described documentation as the most critical part of a data scientist's job — and the key to avoid being the only person who understands how a model works, and therefore, being completely tied to that one project.

"Something we see as we take over some old actuarial models is that no one knows who built the spreadsheet or what a formula means and why it was used. To me, the core of what our job is is to make really good decisions, be able to explain them and have curiosity about why things work. If you can't do that, you're not doing your job. If someone thinks they have a great data scientist but that person is not willing to document their work, they're not a great data scientist," said Warner. ■