

Science Will Win
Season 3, Episode 2
Transcript

Imagine this hypothetical scenario. You're making your way home on the subway after a particularly challenging day. You're a physician at a downtown city hospital. You're tired. And inside the train car, it's sticky and humid. The air conditioner is broken. But that isn't what's bothering you the most.

As the train jolts the crowd of straphangers around you, you can't help but stare into your own reflection in the window, thinking about the patient you just spoke to. They only had a cut from some broken glass. Getting the call from diagnostics that the patient's wound was infected with MRSA, a dangerous and antibiotic-resistant strain of staph infection, was difficult. But it's even more difficult to look a patient in the eye and deliver the news that their life may be at risk. You can't forget the shock and fear on their face.

You hope their illness is treatable.

That night, when you're brushing your teeth for bed, you look at your reflection again and remember the patients you've had in the past who have died from other aggressive, drug-resistant infections. Nothing makes you feel more powerless than trying one medicine after another, only for all of them to fail. It feels like there should be a solution.

It turns your stomach to think about it, but the problem seems to be getting worse. You've seen more and more instances of ineffective antibiotics. It doesn't always result in death, but the prospect is frightening. How long until *no* antibiotics work?

Your only solace is how much you've seen technology improve over the course of your career. You know exactly what you need to do when you get back to the hospital tomorrow.

The next morning you hurry back to work and head straight for the lab. Your hospital recently joined an interconnected database. It combines anonymous patient data from thousands of other hospitals with data from analysis of wastewater. It's one way to track the spread of antibiotic resistant MRSA throughout the country. You open up the spreadsheet on the lab computer and start typing.

The last time you added to the spreadsheet was a month ago—another MRSA patient. Today, you're filling in another line, with de-identified, anonymous data from the latest infection. Each new addition to the database worries you even more.

At the lab, biomedical researchers have recently been using artificial intelligence to find trends in all this data, and even predict the rise of antibiotic resistance in real time.

In a few clicks, you access the predictive chart. A line graph blinks in front of you, and you immediately notice the upward trend in infections caused by antibiotic resistant bacteria. There are ebbs and flows in the disease, but the predictive model indicates that a spike could be coming in your region.

Knowing that this spike is coming, the hospital can make preparations for potential incoming infections. That means strengthening infection prevention protocols, educating the medical staff about the problem, and shoring up more specific guidelines for antibiotic use during outbreak situations.

But even though you can understand and predict the spread of this disease better than ever, it still makes you worried to see individual patients suffering.

Looking at the seemingly endless stream of data in front of you, you feel more determined than ever to *do* something about this. Now that your hospital is part of this database, you might have some options this time. You've heard there are other artificial intelligence models that can use data just like this to help with treatment. It could be worth a shot.

Jeremiah:

This scenario is fictional, but the database described is based on real efforts that researchers are building and using today. This scene is a glimpse into a potential, not-too-distant future where superbugs are more pervasive, but researchers can stay on top of the problem with the help of data.

That's what we're talking about today: data, and the artificial-intelligence-based tools that can put it to use in the fight against AMR.

[THEME MUSIC]

Welcome back to season three of Science Will Win. I'm your host, Jeremiah Owyang. I'm an entrepreneur, AI investor, and tech industry analyst. I'm passionate about emerging technologies and the ways they can shape our world.

That's what we're talking about this season – specifically, artificial intelligence, and how it can help the scientific community overcome one of the greatest challenges facing humanity: antimicrobial resistance, or AMR for short.

In our first episode, we went back in time to learn about the history of antibiotics, and to understand how we got where we are now—a time when antimicrobial resistance is a global and existential threat.

For the next two episodes, we're going to be talking about how artificial intelligence and the latest technology are currently helping us to understand and address the problem of AMR.

Today, we're focusing on analysis. How can AI help scientists understand the numbers, track the problem, and potentially find new pathways toward solutions?

The scenario you heard at the top of the episode is fictional, but it's grounded in work that researchers are doing today—collaborating on sharing information across entire countries and developing more sophisticated predictive and trend-finding models than ever before, all with the help of artificial intelligence.

Before we talk about that, we need to start on the ground level.

What does artificial intelligence actually *mean*?

Ranjit Kumble:

Artificial intelligence, generally speaking, is teaching machines, computers, how to do tasks that would normally require human intelligence to do. Things like recognition, classification, translation, these are normally areas where, you know, as human beings, we acquire the skills and the capabilities through learning.

Jeremiah:

That's Ranjit Kumble. He leads the Enterprise Data Science Team at Pfizer.

Ranjit:

We are essentially a, a group that works on advanced analytics capabilities with essentially all parts of the organization from early discovery through development, through manufacturing, through commercial and medical, and all enabling functions.

Think of us as kind of a backbone for, um, advanced analytics capabilities across the organization, essentially, when you think of Pfizer's mission of bringing breakthrough, um, medicines to patients, um, to improve health outcomes, our value chain is empowered and accelerated through AI and machine learning.

Jeremiah:

So, AI is basically just technology capable of taking in data and making sense of it.

That may seem like technology fit only for computer scientists, huge supercomputers, and science fiction, but as everyday consumers, we're interacting with AI more and more every day. A prime example: generative AI. Here's how that works:

Ranjit:

The algorithm uses the relationships, it's learned to create new types of content. Um, it could be creation of new types of text, it could be images, it could be video, many different modalities. But machine learning is what helps teach an algorithm—where an algorithm understands relationships and generation and generative AI is situations where something new is created from those relationships.

Jeremiah:

One example of generative AI has been appearing in the news a *lot* lately.

Ranjit:

Generative AI as a, as a specialized kind of capability has been around for some time, but I think what really unlocked very, very broad, not just understanding of it, but also appreciation of the power of it, have been capabilities like ChatGPT, where essentially, you know, the algorithm is pre-trained on just vast amounts of information from the worldwide web, from the internet. And essentially, now are set up in a way that questions can be asked, vast amounts of information can be summarized and synthesized as well. And I would say probably ChatGPT might be the most familiar example, um, for everyone at this point.

Jeremiah:

However, not all AI is “generative.” Remember, artificial intelligence is essentially a computer system capable of analyzing data and solving problems, and this technology is already popping up everywhere.

Even if you don’t use ChatGPT, you likely still interact with AI-based technology on a near-daily basis.

Ranjit:

That’s how our phones essentially would perform facial recognition to, let’s say, unlock a phone. The very first time somebody purchases a phone, for example, and sets it up, the phone has an algorithm that recognizes certain or stores certain features of that person’s appearance. And then the next time around when they are presented with that person’s image, if they identify those same features, the phone unlocks. And so it’s essentially almost an exact replica of recognition.

Jeremiah:

The facial recognition technology that many phones use to unlock uses machine learning and computer vision to analyze thousands of points on your face, thousands of points of data to determine whether you’re... you.

Ranjit says that artificial intelligence, and its ability to take in data and break down the problem, can also be applied to *great* benefit in the healthcare field.

Ranjit:

Broadly speaking, there are many, many different conditions that patients might have but doctors might take some time to be able, you know, to suspect or potentially diagnose. There could be conditions that are rare.

There could be conditions that are known, but they're very in asymptomatic early on and doctors only really catch them after they progress to a point where the symptoms are more, more observable.

What artificial intelligence and machine learning allow us to do is they allow us to take very vast data sets that cover just a number of different dimensions, and they're able to find predictors of certain kinds of outcomes, and they're able to find leading indicators and signals of emerging trends.

Jeremiah:

The use of artificial intelligence in the medical field has evolved quickly and significantly over the last few years. And Marinka Zitnik has been a part of that evolution.

Marinka Zitnik:

My name is Marinka Zitnik. I'm an assistant professor at Harvard Medical School with additional appointments at the Broad Institute of MIT and Harvard, the Kempner Institute and Harvard Data Science.

Marinka has been working on the cutting edge of artificial intelligence, machine learning, and the medical field for about a decade.

Over the course of her career, she's gotten to know the history of the field.

Marinka:

Medical applications were one of driving examples for earlier generations of AI algorithms back in eighties where AI algorithms were mainly expert systems.

Jeremiah:

Expert systems are computer systems that mimic human "experts" by making deductions, solving complex problems, and proposing decisions based on information.

Marinka:

And there was a big driving example for how to develop those models in the context of certain healthcare applications, such as for detecting cardiovascular events that patients might be experiencing and do that early on.

Jeremiah:

These early expert systems were far from perfect. In order for them to work, researchers had to painstakingly write every rule that dictated how the computers made decisions, reflecting the knowledge of real human experts.

As technology improves, the potential applications for AI also expand.

For one thing, we have better computers than before – in the 80's, the most common home computers had about 64 kilobytes of RAM. Today's home computers are more than 100,000 *times* more powerful than that. Even your smartphone likely has about 6GB or more of RAM.

As time went on and computer technology improved, scientists gained the ability to create more sophisticated algorithms, or problem-solving procedures. But one major factor has driven the acceleration of artificial intelligence: data. And a lot of it.

When researchers started using AI, there wasn't much data available to inform the algorithms being used. Humans were often the ones actually looking at medical data and trying to make sense of the trends. But now, we have more data than any human could ever break down alone.

We need that analytical help. Which is why researchers are turning to another subset of artificial intelligence known as machine learning.

Marinka:

It's an umbrella term that we use to refer to a toolbox of techniques that—where, those techniques, what they do is look into a large dataset and try to extract patterns from the dataset, and do so in such a way that they can extract those patterns automatically. And these techniques can find patterns that humans cannot see. Those patterns are generally quite complex and you could not distill them by simply looking as a human into— through a large table because the dataset is so large, you cannot easily visually distill and extract a pattern from the dataset. And machine learning algorithms can do that very effectively.

Jeremiah:

To us humans it looks like an insurmountable sea of numbers and figures. But with a machine learning system, it can turn into potential answers, solutions, and suggestions.

Marinka:

And now that the datasets that are being collected and generated by biological and medical research are getting larger and larger, these algorithms that extract and find patterns in the data are getting better because they can extract more complex patterns that are more realistic and indicative of real-world phenomena.

Jeremiah:

And beyond the sheer scale of data, researchers are sourcing data from different places – where you might not expect there to be medically-relevant information.

Marinka:

A very well-known example of that are digital traces that one can extract from social networks or from internet-scale data that was not collected with any biomedical application in mind.

And so across all these, across these different scales, going from molecular biology all the way to what is going on at the level of individual cells, tissues, human body, an entire individual person, and then these broader communities and ecosystems, um, of how a person interacts with other people in their neighborhood, in their family, and in their broader social ecosystems. Across these very different scales, we have now new data types, new data resources, and the data that are being collected, once collected, are potentially amenable to, uh, analysis with advanced computational methods, particularly machine learning techniques.

Jeremiah:

So, artificial intelligence, and more specifically machine learning, is helping scientists build this bigger picture of the health of people, but also of communities as a whole. And with that more detailed image, patterns and potential solutions can begin to emerge. One common example comes from a surprisingly ubiquitous source: Google.

Adrian Egli:

Google Flu was an algorithm which was in place by Google when people had the flu.

Jeremiah:

You may remember Adrian Egli from episode one. He's a professor of medical microbiology at the University of Zurich in Switzerland, and he's been studying AMR for more than a decade.

"Google Flu Trends" was a project that ran from 2008 until 2015. It used automated systems to analyze data and track the spread of the flu in near-real-time. The data source? The search terms users were Googling.

Adrian:

When they were coughing, they had fever, all signs of an upper respiratory tract, then they googled these terms. So for example, "fever," "coughing," and so on. And if many people actually look for the same search terms, you can actually estimate that there's an outbreak somewhere or an endemic wave coming because you have more and more cases of people who look for the same symptoms. And so at the end, what Google Flu does, it, it looks for a data pattern.

Jeremiah:

Google collected common searches that corresponded with previous CDC data on the rise and fall of the flu. Those searches all became data points that contributed to predictions about when and where the flu was spreading.

It wasn't always perfectly accurate. At times, Google Flu Trends would overestimate the incidence of the flu. Because, of course, not all flu-like symptoms mean that someone is *actually* sick with the flu. Nonetheless, the trends data proved helpful in enhancing the accuracy of CDC predictions.

Though that program concluded in 2015, Google sent that data to healthcare and research organizations like Columbia University, the Boston Children's Hospital, and the CDC. That way, the flu search data can continue contributing to predictive models.

This same principle can be applied to the spread of antibiotic resistant bacteria, using data from doctors' inquiries in their health system, or their prescriptions for patients.

Adrian:

If you think about, um, physicians looking up all of a sudden, you know, antibiotics, which are very unusual because they are second line drugs, for example, one could estimate are there, there must be something around in the community which is hard to treat, and this is why they look for unusual, uh, drug medications.

Adrian:

So one thing is the outside world on how people would actually look and search for information. And if many people look for the same information you, you could estimate that there's a problem. But on the other hand, inside a closed system, like a healthcare system, a hospital, for example, you could also use data analysis to find out if there's, uh, an outbreak, uh, going on.

Jeremiah:

So data like this can help researchers or algorithms identify and isolate trends, but more sophisticated discoveries require more sophisticated data than just Google searches. Unfortunately, not all countries have the infrastructure, funding, or support to build data sets that can be used to track and understand these big problems.

Adrian:

At the moment, the biggest data producers are technologically advanced countries such as the US, European countries, some Asian countries, which produce a lot of data. But if you look at the global data sphere, there are clear gaps. So South America or African countries, India, I mean, a lot of people live there, but the amount of data they produce and contribute to the global data sphere is really minimal. So I think at the moment that data is absolutely dominated by a few countries, which produce a lot of data and then this has an impact on the algorithms.

Jeremiah:

Think about it: If you're a doctor in Sub-saharan Africa, getting data from the U.S. is only going to be so helpful. That is one element of a healthcare and technology divide that can prevent AMR from being adequately and fairly addressed.

Bridging those gaps is one primary concern of researchers. That means providing infrastructure and opportunities to gather data and create the kind of technology that can use it.

In 2020, for example, Pfizer and the charity organization Wellcome launched the Surveillance Partnership to Improve Data for Action on Antimicrobial Resistance, or SPIDAAR, in several Sub-Saharan African countries. The initiative launched in 2020 and partners with the governments of Ghana, Kenya, Malawi, and Uganda to gather and develop region-specific data.

Adrian's home country, Switzerland, also faced the problem of needing more local data.

Adrian:

This is actually one of the reasons why we have invested in Switzerland into a network for data exchange and data generation, because we were afraid, I would say almost that all of a sudden we will have algorithms which are used in our patients, which were trained not by Swiss data sets, let's say. And, and then it might not reflect and be as efficient as a data set should be.

Jeremiah:

To tackle that, Switzerland *did* create a democratized database of medical information.

We're going to take a closer look at that database to learn more about how medical data and AI can come together to help researchers and doctors track and understand the problem of AMR.

The first step is creating the infrastructure, making it possible to even house all that data.

Adrian:

So in Switzerland we have an incentive called Swiss Personalized Health Network. So in each university hospital, they have built a so-called "clinical data warehouse" where data can be stored, where data can be quality controlled, and then also be exchanged between the different hospitals. And this is done over a data highway, sort of, and this data highway is a very secured way how data can be exchanged between the Swiss data centers.

Jeremiah:

The Swiss Personalized Health Network, or SPHN, tracks thousands of points of data. There are databases of patients' height, weight, and blood pressure. The specifications of patient illnesses—whether it's a tumor, an allergic reaction, or something else entirely. And there are even databases that track the methods doctors use to measure and treat patients' illnesses.

These are all really personal details, which is why data protection was an important consideration from very early on.

Adrian:

What we have in place in Switzerland is a so-called general consent. And so basically you are asked if the data you have produced during your stay in the hospital can be reused for health research purposes. You can say yes, no, or you can also say, I don't know. And then, uh, you basically, that's almost like a no. But if you say yes, you agree,

then the data can be reused for research purposes. If you say no, it's absolutely clear your data is not gonna be used for research.

Jeremiah:

This puts the patient in the driver's seat. They have control over whether the data becomes a part of this important research. Additionally, these healthcare databases usually use anonymized data that can't be easily traced to each individual.

But in the future, as our feelings about data evolve, Adrian sees the potential for even more healthcare data sharing.

Adrian:

So I think asking the patient to be involved in the decision making about sharing data, um, and healthcare-related data is a very, very important one. What is quite interesting, people used to store a lot of their data in their own, you know, computers in their, you know, smartphones, et cetera. But more and more data is shared in a cloud. And that could also happen to healthcare data at a certain point that people say, you know, the systems we have are so trustworthy, it's okay that, you know, my healthcare related data is stored in the cloud, and then potentially can also be accessed and used if I agree to it.

Jeremiah:

Today, the massive dataset that SPHN has amassed and stored can be used to track pretty specific health concerns. Adrian is the principal investigator for one project that uses funding and data from the SPHN to track antibiotic-resistant sepsis. Sepsis is a life-threatening response to infection in the body, caused when the immune system starts attacking itself.

Adrian:

So people with sepsis, they clinically look very different. Some people may have a fever, other people do not have a fever. Some people have, you know, let's say really difficulties with breathing. Others have other symptoms. So it's a very complex, very heterogeneous disease. And having AI analyzing the complexity of the data can tremendously help and support us.

And in our project, what we try to do is to really compare the different, uh, treatment approaches, the different antibiotic drug resistance we face across the country. And we exchange this information to also, um, use the data for digital biomarker discovery. So to basically recognize sepsis at an earlier stage and recognize antibiotic resistance also at an earlier stage, than using classical standard methods.

Jeremiah:

The data points involved in understanding sepsis are much more numerous and complicated than those involved in tracking something like the flu. When you add tracking and identifying antibiotic resistance on top of that, those trends become even more difficult to find. That's where artificial intelligence can come in.

While there isn't yet a final product that allows a doctor to, say, plug in their patient's symptoms and determine whether they have sepsis, Adrian's team is building those specific datasets with AI in mind. They're actively working on algorithms to help understand these trends, and these datasets are the foundation of truly understanding the problem.

Initiatives to build accurate and widely available datasets for understanding AMR are in motion all over the world. While Adrian and his team in Switzerland are looking at how patients respond to different microbes, at Pfizer, Ranjit and his team are looking at AMR from a different angle—the microbes themselves.

Ranjit:

It's a tool for providing real-time information to the medical community to understand emerging trends in, um, resistance to antibiotics.

Jeremiah:

In 2017, Pfizer launched ATLAS across 60 different countries.

Ranjit:

ATLAS stands for antimicrobial testing, leadership and surveillance system. It's essentially a website and it's designed to, to ensure that physicians have the very latest information globally on the latest trends with respect to resistance to particular medication.

So by understanding kind of emerging trends and resistance, what they're able to do is anticipate what they're upcoming trends are going to look like locally in terms of patients seeking care, in terms of the emergence of new diseases that are going to potentially not be controllable through existing medications and, um, allow them to, to design interventions in a way that gets ahead of that, in particular, sort of anticipating this from very, very early signals.

Jeremiah:

It's important to underscore why it's so vitally necessary to source and understand all of this data. One big step toward stopping this evolving threat of antimicrobial resistance, is having a grasp of how, why, and where it's happening. But the other step is using artificial intelligence systems to help researchers and doctors analyze the data and take direct action.

That's what we're getting into in the next episode—how AI can take this data and help humans find new pathways for solutions and drug discovery. AI can help researchers study the bacteria itself, enable appropriate diagnosis of patients, and even pursue the discovery of new antibiotics.

Marinka:

Drug discovery and development is an incredibly expensive and costly process. And so an actual question to ask is, is it possible to augment and accelerate, uh, various steps in this, in the drug development pipeline. And so that we can compress the timeline from years to months or even weeks in certain cases.

Jeremiah:

Science Will Win is created by Pfizer and hosted by me, Jeremiah Owyang. It's produced by Wonder Media Network. Please take a minute to rate, review and follow Science Will Win wherever you get your podcasts. It helps new listeners to find the show.

Special thanks to the responsible AI and anti-infective teams at Pfizer. And thank *you* for listening!