accenture

Dominic Delmolino:	The Exploring AI in Government podcast is brought to you by Accenture Federal Services and produced by Government Executive Media Group's Studio 2G. Accenture Federal Services combines the power of artificial intelligence, automation, and advanced analytics with deep client, industry, and technology expertise to help agencies reimagine how they achieve their mission, serve citizens, and manage their organization. Learn what Applied Intelligence can do for your agency at "Accenture-dot-com-backslash-federal-AI."
	[Intro Music]

| Dominic Delmolino: | Hello and welcome to another edition of exploring AI in government, a podcast series dedicated to interviewing leading global minds in the artificial intelligence ecosystem and getting at the insights that drive adoption across key industries and the federal government. I'm your host Dominic Delmolino, chief technology officer for Accenture Federal Services and I'm here today with our guest analyst Kathleen Walch, AI and machine learning expert and managing partner at Cognlytica.

While we were exploring many of the thematic issues facing agencies and their artificial intelligence adoption, the question remains, how do you get started? In many cases, it's an individual with both the challenge that can be solved with AI and a vision for doing so. This week, we're bringing you a bonus episode showcasing one of those pioneers we're excited to have on the show. Alex measure economist at the Bureau of Labor Statistics. |
|---|---|
| Kathleen Walch: | His challenge was seeing if there was a way AI could augment existing processes in a high-quality manner so that his colleagues could shift their efforts to more challenging work. Alex, we're really excited that you're joining us today. |
| Alex Measure: | Oh, thank you very much for having me. It's a very exciting to be here. |
| Walch: | So I'd like to give a quick background on you, Alex, but I'm sure that you can do a much better job. Alex is an economist at the Bureau of Labor Statistics and he focuses on machine learning initiatives and natural language processing sometimes with neural networks. Most of his work deals with auto coding and text classification, but he also works on automatic record matching information extraction, error detection, and other applications of |

accenture

	machine learning. We'd love to have you explain that a little more.
Measure:	I can talk about uh sort of how I got started in, in machine learning. You know, obviously economists are not usually associated with machine learning. But I started out as an economist at the Bureau of Labor Statistics and one of my initial jobs working on the survey of occupational injuries and illnesses was to look through written descriptions of work-related injury and illness and look at how they're classified. In other words, how we pull data out of these written descriptions to generate statistics about the workplace safety, such as, you know, the most dangerous occupations, the most common causes of injury in those occupations.
	And so that's one of the things I did as a new economist, not something I particularly glamorous. And so that got me interested in ways of, of automating these sorts of tasks. And so that's how I became involved in machine learning. And now I was fortunate at this time that we're now a bunch of free online materials resources like Coursera and EdX, and they had some excellent machine learning classes taught by leading experts available for free. And so that's how I as sort of acquired the initial skills. And then of course I had many opportunities to apply those skills at the Bureau of Labor Statistics to automate tasks that I was performing myself.
Walch:	You know, it's, it's funny that you say that these aren't glamorous tasks that you were automating and those are sometimes the best use cases for, for AI because it, it's mundane, routine work that's, you know, mind numbing in some sense that it's great. But it's not glamorous and so people always want to start with a glamorous project. So, I know that one of your first tasks was to classify hundreds of thousands of job titles into more than 800 occupation codes to help decrease the error rate that humans were doing. And this in turn ended up saving many hours, many man hours. Can you talk to us about that project?
Measure:	Yeah. So, we collect a variety of information through the Survey of Occupational Injuries and Illnesses. We get the job title of for the injured worker, we get a few sentences describing what they were doing when they were injured, what happened, what part of body was injured, and then to generate statistics from that data we read through it and we assign classifications by hand, which we can then add up into these statistics that answer questions like what are the most dangerous occupations? What are the most common injuries to those in those occupations? And so on. So, you know, occupation was part of that and that was a part that I particularly worked on.
	So, you know, I learned about these machine learning techniques online and I

	started applying them to the tasks that I was facing. And what we found was that they worked much better than expected. Not only did it allow us to teach the computer how to perform these tasks, but in fact it allowed us to teach the computer to perform these tasks as accurately, initially, as our trained human staff. It was very surprising that we would be able to do that as a sort of initial step. And it demonstrates the power of these machine learning systems. That's great. You've got this really nice use case that you're able to start with.
Delmolino:	You obviously had success with that, you know, how do you see AI impacting other areas at the bureau? Do you think there's going to be some broader use cases or what do you see there?
Measure:	Yeah, so the particular technique that we're using is something called supervised machine learning. And the basic idea is that if you want to automate a task, instead of trying to hand program in all the information the computer needs to know to perform that task, instead you will gather up a bunch of data showing the inputs to the task and the desired outputs from that task. Uh, you'll feed that into a learning algorithm and the computer will learn how to perform the task on that data. Now this is a very general recipe. You may have noticed all you really need is uh inputs relevant to the task and the desired output. And it turns out that we have tons and tons and tons of examples of this, not just in the bureau of Labor Statistics, but you know, in, in every industry. And that's why you see machine learning, transforming so many industries right now.
Delmolino:	So, we all have projects that we hope to start really well and finish where we want them to go. But always there's some challenges we encounter along the way. What challenges did you face? What potential challenges might other agencies face and how would you go out planning to address those? How did you handle those? Tell us a little about the challenges you've encountered.
Measure:	Well, so for the survey of occupational injuries and illnesses, we were fortunate not to have so many challenges. We had a lot of great training data, which we had collected over the years. And it was stored in a nice, clean, a database that had been very carefully in uh reviewed and validated. And so, it was relatively easy. In fact, we were able to build these initial machine learning systems in a couple of weeks. If you don't have that data, the project is much, much more difficult. And since then I've worked on some projects where we don't have that nice clean database with all of our training data. You know, it can completely prevent

	you from using the tools in the first place. So, I think if you're interested in applying machine learning in your organization, it starts with data. In many cases you already have a manual process that is producing the data. Make sure if you have that process, make sure you're saving that data because that data then becomes the input into your machine learning systems.
Walch:	It was great that you were able to identify this as a good use case for a successful project.
Measure:	And I think an important point to be made is that many of organizations already have manual processes that are sort of as a side effect, generating the data. They're just not saving that data. And so, I think that if you want to start applying these techniques inside your organizations, I think that's a critical starting point. Start saving the data that's being generated by your manual processes because that will allow you to apply these techniques down the road.
Walch:	Yeah, and we know that the government produces a lot of a lot of data. So, I know that that you were very early to embrace AI. You've been at the Bureau of Labor Statistics for about 10 years now and not everybody was as forward thinking as you are working on AI projects 10 years ago. Can you talk to us about how you got started, why you thought AI was the right solution and why you didn't go another route?
Measure:	Well, I got started out of personal interest in whole as a new economist, it was my job to review some of this data by hand and to classify it by hand and it's not the most glamorous thing to do. So that got me interested in machine learning and then that got me interested in learning about these techniques and applying these techniques. And so, it sort of went from there.
Walch:	And I'm sure in the past 10 years you've seen a lot of tools change and actually become a lot easier to use. So, have you found that more people are using these, these tools and creating AI projects within the Bureau of Labor Statistics? Because of that?
Measure:	The advances in the tools has been amazing over the last 10 years. And I think if you went back 15 to 20 years, it was a, it was actually very hard to implement these systems. And now you have a libraries like the psychic learn library and Google tensor flow, AWS' Pytorch that make a lot of these things much easier to build and implement. So, one of the things I spend a lot of time on in my current role is actually teaching my colleagues how to

	implement these systems and many of them have gone on to implement various successful systems.
Delmolino:	So, Alex, clearly you had a strong personal interest in learning how to use these tools, applying into the job that you are confronted with and that's really awesome. But how did you get management buy-in? What kind of support did you have? How did you convince people? Did you have to create a business case? Tell us a little about the, you know, great idea, but how do you start executing on it?
Measure:	Yeah, so I, I think the business case was sort of there from the start. In the case of the survey of occupational injuries and illnesses, each year we're collecting 300,000 written descriptions. We have dozens of people around the country that are manually reading through each of these descriptions. And obviously that takes a lot of effort. We estimate it takes about 20,000 hours of labor. Each year is equivalent to about 10 full time employees working on nothing else. So, you know, clearly there was a, a very real resource costs there. There were also quality considerations. One of the things we found when we started investigating this task was that if you give two different trained humans the exact same narrative, they will only choose a, in this particular case, the same classifications for that narrative about 70% of the time. And so obviously this was not ideal from a measurement standpoint. And so, it was one of the reasons why management was very interested in investigating this early on. So, I was very fortunate that, you know, senior management was very engaged and very focused on this uh from the start, they had seen research happening in academia where people who were applying similar techniques. And so they wanted to see this applied inside the bureau of Labor Statistics. And you know, so that got this project started. I think one of the things that helped us get it from sort of research to production was to evaluate the project not only in isolation but to measure also the quality of our existing manual process. And so what that allowed us to do was directly compare coding quality that was coming from our, our manual process to what was coming from these machine learning systems. And we found uh, you know, even with these initial systems that the quality was very similar to that made this the business case very strong. And so of course if you go to management and you say, look, we can do this very difficult time consuming and you know, get similar or even better quality, they're very receptive to that.

Walch:	Yeah, it's great that you were able to actually measure it that way and it had positive results because some people do AI just to do AI and that's not, that's not great, you know, do it because it's going to help reduce labor or it's going to help make people more efficient. Not because you want to do it because it's cool and you heard that you should do it. So that, that's great that you were able to measure it and see real impact. Now I know that you've been working on AI and doing AI for a long time. Can you tell us, you know, how you decide what tools and algorithms you should use and how you go about deciding on which projects are the right fit to move forward with?
Measure:	For deciding on projects, I think the key consideration is the data. So, the most powerful, you know, tool of AI right now is something called supervised machine learning to provides machine learning requires lots of data, typically showing the inputs to some task and the desired output to that task. If you don't have that, it's very hard to apply supervised machine learning. As for figuring out the tools, what we did was we looked at what the experts are, we're using, we looked at what they were teaching and Coursera, we looked at what the researchers are using to conduct their research. We looked at what Google and Facebook are using in their production systems. And it turns out that a lot of that software is available for free. And so that's how we decided to go that route. And I have to say we're very happy with what we've been able to do with the open source software.
Walch:	You're right, there is a lot of software out there that's very powerful and we always say don't reinvent the wheel if it's out there, use it. You just sometimes need buy in from management and also for security reasons. People can sometimes be a little concerned about that, but a lot of times you don't have the resources and the skill set that you need to really build these tools from scratch. So, don't do it. It's going to take a lot of time. Now I know you talk a lot about data and for supervised machine learning its very data hungry and for some projects you've had clean, well labeled data but not for all of them. So, can you talk to us about some of the challenges of getting that data cleaned and labeled and prepped and then how do you validate that your AI project is working and successful? How do you measure that?
Measure:	Yeah. Well yeah, in some ways getting the data is both the easiest and the hardest part. Takes the most work to get the data. And it's sort of remarkable. You look at the work that goes into building one of these machine learning systems, the actual writing, the code to implement the algorithm takes. Maybe, you know, if you know what you're doing, maybe a few days, but all

accenture

	the work to generate the data that goes into that can take years and years. I think it's really important that organizations start, start saving the data that they're already generating so they can use it down the road. Now how should you save that data? Well, it's sort of hard to predict what kind of data is going to be useful. So, I, you know, I suggest and hard drives, they're pretty cheap. So, I suggest saving as much data as you can. Also use modern tools for storing data. Things like relational databases. They make it very easy for computers to work with large amounts of data and that makes it much easier to implement these systems down the road.
Delmolino	So, Alex, you work in an organization whose very title includes a description of statistical expertise, right? Your labor statistics. And so I'm guessing a lot of your colleagues have some point of view on understanding, you know, statistical modeling and whether or not you can tell if data is relevant for a problem to that make it easier or harder to implement AI in an environment where you have so many people who are so familiar with statistical analysis data.
Measure:	That's uh interesting question because in some ways it makes it easier. In some ways it makes it harder. It certainly helps to have lots of people very familiar with the techniques and the tools of data analysis. Many of the techniques that we use in AI, that we use in supervised machine learning, are very closely related or in some cases identical to techniques that are used in the statistical community. Now that said, there are some important differences in the way that statisticians and machine learning people approach projects and that does lead to some confusion. I think when you're working in a statistical organization, you know, one of the things that statisticians like to spend a lot of time on his theory and uh understanding the sort of theoretical justifications for the modeling decisions that you're making. The machine learning approach is to not focus so much on theory but rather to focus on evaluating that data sort of empirically. So that leads to two very different approaches to tasks even when you're using the same data. And that would occasionally lead to some confusion about how we're doing things. I think overall, however, it's very helpful to have lots of people with data analysis, backgrounds implementing and using systems like this.
Delmolino	Yeah, it sounds like data literacy is certainly an important component of understanding what you're going to do with machine learning and how to use

accenture

	select the right data to train the machines and how to do the task. I'm also curious, you know, you, you get a lot of data coming in and I'm certain that you've got people who've been trained on dealing with sensitive data, how they handle private information and the ways to address bias and understanding when they do their actual activity. How do you translate that to the machine? How do you take that same kind of ethical handling of data that the humans had been trained on? They show that they can do that. How do you translate that to the algorithms so that we make sure that there is as compliant and ethical as the humans looking at the data?
Measure:	Well, so in some cases we can translate it directly. So, you know, one of the things we do to make sure that humans are behaving ethically is we review the, the work that they're doing. We can, and we do the same with machine learning. And I think that's a critical component in maintaining these machine learning systems. You can't just build it and then forget about it. You really do need to continue to monitor it and make sure that it's doing what you want it to do.
Walch:	That's a really important point because it's not a set it and forget it. It's not due at once. And then revisit it in a year or five years, constantly go back and make sure that you're continuing to get the results that are expected and if not, you know, retrain your systems. Also, I'm sure that it's continuously being fed new data and you want to make sure that that doesn't misalign anything.

So, Alex, thank you so much. You have really provided a lot of insights into how you use data, the ethical considerations and also some successful projects that you have already implemented. But can you talk to us about your goal for the future and where the Bureau of Labor statistics wants to continue to use AI and automation going forward? |
| Measure: | I think it's a very exciting time to be at the Bureau of Labor statistics because we have a ton of data and have data about very important things, and we have a lot of people with good data analysis skills. And so now that we have access to these machine learning techniques, I think there's a lot of places where we can start applying these tools and techniques to do a lot of interesting and important research and analysis.

So, one of the things that I'm focused on right now is spreading knowledge about these tools within the bureau of labor statistics so that other people can apply these techniques. We've seen already a number of exciting projects inside the bureau led by people who are now applying these techniques to their own challenges.

I think another interesting area that's sort of developing in the machine |

	learning world right now is in the area of transfer learning. So, one of the bottlenecks to implementing these machine learning systems is the training data. You need the training data and then you'll typically need a huge amount of it. What's happening over the last few years in the transfer learning research is that researchers are finding ways of allowing us to create models that do not need so much training data in our specific domain. Though I think that has the potential to open up even more opportunities for applying these techniques and I'm very excited to see what that will allow. But even without that, there are already tons of opportunities and we're working very hard to explore those.
Delmolino:	So, Alex, it sounds like the future is really promising for AI and automation at the bureau, so thank you so much for sharing your perspective and lessons about how things have gone for you. We're really excited to see where you take this in the next few months.
Measure:	Thank you very much for having me.
Delmolino:	Thank you for joining us for this bonus episode. If you've liked what you heard, tell a friend or share socially using the hashtag explore AI. Stay tuned for another full episode with federal CIO, Suzette Kent and Courtney Winship of USCIS talking artificial intelligence and customer experience in two weeks.
	[Exit Music]
Delmolino:	Thank you for listening to exploring AI in government brought to you by Accenture federal services. To continue the conversation, visit us at accenture.com/explore AI where you can listen to other episodes and download relevant research, or you can connect with myself, Dominic Delmolino, and our guest analyst Kathleen Walch on social media. We look forward to seeing you at accenture.com/ExploreAI.