Demystifying AI – Episode 4 - Transcript

The butterfly effect of digital twins

a podcast by Deloitte AI Institute Canada

Aisha Greene

Well, hello and bonjour. Welcome to our next episode of Demystifying AI. I'm your host, Aisha Greene, the Senior Manager of the Deloitte AI Institute. For this podcast, we ask the question of what is a digital twin? But before we delve into that question, I want to paint for you a picture. Imagine a ripple effect. We want to understand how changing even your monetary funds or even your mortgage might have implications to not only your carbon footprint, but also other aspects in the world. Believe it or not, items like digital twins, scenario-based testing can make a big difference in how we plan and plan for future investments. I'm honored to have our guest and our esteemed guest today.

Arthur Berrill is CTO of RBC Bank of Canada and data analytics team. He's the technology leader and voice of RBC technology and innovation within the commercial partners, government and open source communities, and academia researchers and domains relevant to RBC's vision and strategy. In service to this responsibility, Arthur's involved in most of the data science disciplines involving location intelligence, data content, artificial intelligence, ontology, graph analytics, and climate change studies. In particular, Arthur has a long history of location intelligence field. Arthur is an RBC distinguished engineer. With that, I welcome Arthur to this podcast. Welcome, Arthur.

Arthur Berrill

Thank you very much.

Aisha Greene

For the first question. Let's delve into that question of what is a digital twin?

Arthur Berrill

It's one of those great questions, isn't it? A digital twin can be many things according to its application. I was just talking to one of our other guests here about a digital twin being used to model electric motors to see what's going on inside the electric motor, or it could be modeling what's going on in a

country or a town. The bottom line is that it is a virtual representation of a combination of physical, virtual, and social elements. Those are combined in a manner to allow you to do prediction and simulation, particularly the simulation. It's a magnificent tool for doing what-if analysis.

Aisha Greene

What are the foundations of putting together a digital twin and the right elements that we need to have as components to even get started?

Arthur Berrill

The digital twin technology that we work with is variously labeled as a digital twin architectural style or a digital twin technology stack. Both of them adequate explanations of what makes up a digital twin. But what's really, really cool about it is that it is a set of technologies, not all of which are brand new, who cooperate in a brilliant way. That's really the magic of the framework, is that they all cooperate neatly together. If I run down a list, the base of a digital twin is usually, not always, but it's usually a knowledge graph which gives you incredible flexibility in the manner that you hold the data, but it also allows you to look for patterns in the data that you might not otherwise be able to detect. Build on top of that, we use models of various descriptions. Quite complicated machine learning models can become resident inside a digital twin.

What's particularly cool about that is that the connection into the knowledge graph is here's a knowledge graph node. That knowledge graph node, if I perturb it in some fashion and there's a machine learning engine resident in the twin that depends on that variable, then it will recompute and it will predict something. But it gets better than that because we also include ontology and all of the ontology structures. The structure in the knowledge graph is not random and it's not chaotic because it is organized by the ontology. Once the ontology, one or more, is in place, you can do ontological reasoning. That means that that variable I was talking about before that I perturb not only triggers off the machine learning engine or model that's behind it, but it can also trigger ontological reasoning and produce new inferences which now become resident in the digital twin.

If you are not squeamish about categorizing learning, it is a learning engine. There's a lot more to it. There are active agents, there are passive agents, there are various aspects of the data that can be made to be real time. In the model that we work with, for example, it's relatively straightforward to hook up the IoT sensors that are in the phrase of river at west. That, all of a sudden, gives you a live predicting engine that you can also perturb. What if there's 12 inches of rain in Toronto tonight, what are the effects of that?

Aisha Greene

Do you mind just giving us an example of what is ontology?

Arthur Berrill

Oh, that's a great... I'll give you the five bucks later. That's a great question. Ontology is a way of understanding the meaning of data. It's not metadata. It's actually an active machine executable understanding of the data. What becomes very nice and cool about this is that it is typically expressed as

triples. A knowledge graph is expressed as triples. Here's a node and the relationship to another node. That's all you really have in a knowledge graph, but you can also have a node and here's the relationship to another node as a part of your ontology. The definition of the structure of the knowledge graph is part of the knowledge graph. That's a particularly sweet solution. That's something that doesn't happen very much yet because you really need to prioritize the mechanism of ontology. That's a particularly sweet combination. When you put that into a digital twin and you have your active machine learning agents, you have your ontology reasoning, your live feeds into the engine, it suddenly becomes an evolutionary engine effectively.

Aisha Greene

I want to talk about how those relationships actually build essentially a neural network, because that is the relationships that also people want to see what is the practical applications. Before coming to RBC, your history is phenomenal in terms of using digital twins for actual practical applications. Can you walk us through how people could actually see this and practice and what this looks like?

Arthur Berrill

Yes, I can. One of the headaches, I'll give you a little bit of a history next, but one of the headaches in this is that in North America, there's very little activity in digital twins, which is why I'm hoping this podcast will resonate a little bit. I mean, NASA made digital twins famous in 2010 when they started using digital twins to simulate spacecraft. Generally speaking, it's a bad idea to destruction test a $3 billion satellite. It's frowned upon. If you build a virtual copy of it and hook it up with IoT instrumentation, you can simulate in a virtual space what that spacecraft is going through and where it's going to fail. They made it famous in that respect. I did mention Siemens earlier as well. They've been very successful in modeling electric motors, particularly transformers, and there's various others.

The notion of simulation by a virtual copy of a device has been around for a bit, but the idea of applying that to the outside world is relatively new. In North America, that's virtually non-existent. There is a knowledge graph that is perilously close to a digital twin, which is used by FEMA in the U.S. It's called the NOAA graph. You can look it up. It's a 10 billion node knowledge graph, but the executable capabilities are a bit limited. It's definitely on the right path and we learned quite a bit from point of view of scale by looking at what they were doing there. But overseas, for example, Christchurch, when Christchurch went through those terrible earthquakes, one of the things that we looked at doing there was a digital twin of the ideal rebuilt state of Christchurch. That's a very powerful way of modeling how a town will work. Now, normally, you don't have to demolish half the town to figure out a new solution for it.

While we were doing that, the folks in Singapore decided to do a digital twin. The University of Cambridge got involved in the process of putting that together, and then there were spinoffs from that. In European terms, Australian terms, digital twins are ascending very quickly. Helsinki has one, Amsterdam has one. Run your way around many of the European cities, you'll find at least the beginnings of a digital twin. The UK is very advanced in digital twin technology. The Australians, I think, may actually be at the top of the pile.

For example, in my home state of Queensland in Australia, if you contribute open data, you must contribute it as part of the digital twin, which is a lovely idea because it now builds... The process of building a dataset is documented as a resident ontology and you know that the data is going to fit, so you don't have that horrible mismatch between multiple datasets.

Aisha Greene

What is RBC doing in that space to make digital twins something that we can look towards as an example? Two, how do we get to the point of Canada as a nation replicating what's happening in Australia? Because we have a lot of areas where data isn't always as free and accessible, and there's a lot of other privacy concerns that are added onto it as well.

Arthur Berrill

We're at step one, is the way I would characterize it with RBC. I'm very pleased that we're leading... We're certainly leading North America and I think leading the world in terms of an FI investing in digital twin technology. One of the things that we've done inside RBC, with respect to the digital twin, is in preparation for the digital twin, we went to the extent of figuring out how we can manage privacy for the digital twin. We have two mechanisms that are now resident in our suite of capabilities.

One of them is differential privacy. That's probably a job for another podcast. The second one is basically simulated data where you take an original data set and you create something that will give you the same analytics results but is not violating any privacy rules. We did that initially because we knew that we were going to need that for the digital twin. The ultimate goal of the digital twin we've built is that the outside world can use it. That's going to be a long way away because some people would regard that as fantasy. I'm not convinced it is a fantasy. I'm convinced it can be done. The idea being that if you are a commercial partner of RBC, this will be step eight or something or other in our series.

You're a commercial partner of RBC and you bring your SKU data, your point-of-sale SKU data, and you introduce that into the digital twin. The capabilities that are already there compliment the data that you've put into it, and now you can read out from there how you're going to move your stores around, which products you're going to put in which stores, and so on and so forth. That's a brilliant service. We could do that for the customer, but I would rather it was in a situation where the digital twin is accessible under certain controlled circumstances and that customer can benefit from that.

Aisha Greene

With all of these technologies because it is talking about sometimes private, confidential information, there's that layer of trust in order for people to have access and to utilize these tools and even to trust the SKU data that's coming out of these support models. Are we at that level of trust where people will readily trust what's being presented in these digital twins?

Arthur Berrill

It's an understanding thing, yes. We're definitely not at that level of trust. As you know, banks operate on trust, which is one of the reasons why I think banks will be the first to make this move because the trust is there. It's a matter of transferring that trust to the thoughtful use of the data. We would have to have particularly strong evidence that it is safe to put the data there, that no PII information is going to be released by accident, so on and so forth. It's risky. We would have to have really solid foundations under it before we can actually turn this into something very powerful. Yeah. It's a tricky space and I don't know that we've got all the right answers there yet. I think that's one that benefits from conversations with experts, but also conversations with the public.

Aisha Greene

What's the step one for people to actually get started in this space and actually realize the true value of these digital twins?

Arthur Berrill

Well, that's a good question and I think there's a reasonable answer to it. The reasonable answer is probably following the path that we followed. We started extremely small, really, really small datasets. When they started to balloon, we went small again. We cheated a bit in that we ignored scale. We just said, "Okay, we can make this work. Let's not worry too much about scale initially." You can put a knowledge graph together with relative ease. There are very, very good tools out there. We happen to use GraphDB as the base, but it doesn't have to be GraphDB. There are multiple options out there. You get some data loaded into your knowledge graph, that's one part of it. Another part is to make sure you've got an ontology. The really cool thing here is that you can put an ontology together without actually writing one.

The way that's done is that you harvest from existing ontologies. Some of them are really, really good and some of them are pretty much rubbish. For example, in the agriculture solution that we put together, we found a soil ontology, a really, really good ontology. Very, very well done, very logically organized. They basically took the entire soil's manual, and pulled the whole thing apart, and turned it into an ontology. If you have that ontology, you now have the controls in the way that you load the data and you now have the schema for your data. You haven't got a digital twin yet, but you've gotten very close. Now, there are tricks that you can use to federate ontologies together. If soils' ontologies is not your bag and you want to do something altogether different, there's a fiber ontology, which is one of the biggest one that's the financial, I forget what it's called, but it's the one that the banks would use.

Once you have that altogether, the framework that we used permits active agents. We've built in active agents and that's a matter of... You can take a model you already have. Let's suppose you're doing something with, let's say, real estate. You're doing mortgages. One of the things that you need in your digital twin for the mortgage calculation is the property valuation. Well, we have a model for property valuation. It's not all that hard to unlatch it from

where it is now and connect it to the variables that are resident inside the knowledge graph. It's now a functioning part of the knowledge graph.

Something changes in the mortgage application and the digital twin, excuse me, knows by itself that it has to call this model to get an updated property value. You build it up in that fashion, but don't make it... You want all the features of a digital twin, but don't make it excessive. We said about this. We actually used eight separate ontologies and we federated them together. We will publish a paper on that because nobody's ever done that before. Believe me, it was painful, but we built the tools to do it so the people who follow the paper will probably be able to do it relatively easily.

Aisha Greene

I want to ask a question about the ripple effects of some of these connections, because I started off with talking about looking at climate change and the effect of changing one element and what does that mean for other things that are downstream. Part of your role is also looking at the impact of climate change, and how we've changed one element of financial modeling, and what that will mean for other citizens. Can you talk to us about how you're looking at climate change and some of those ripple effects to those nodes of ontology that you start to change?

Arthur Berrill

Yeah. It's another good question, but let me preface it by saying that climate change is pretty nearly the perfect application for a digital twin because it is complex beyond human comprehension. All of the climate models I've seen have all been, "Well, I think that's roughly it." Or, "I think this might have an effect on this," or, "Is this a factor in..." The cool thing about a digital twin used in this context lets you understand how those connections work by building them up piece by piece, but you don't build them from a human understanding of the model, nor do you build them from a machine learning understanding of the model. You build them piece by piece. I know that those two nodes are related to each other by this model. Okay, great. I can put that in. I know that these two nodes are...

Once you get that knowledge graph of connections, then you calibrate your projections. Say, "Okay. I'm going to use all the data from 2015 until now on an hourly basis for precipitation. I'm going to use that. I'm going to use half of it to set the model going and I'm going to use the other half to figure out whether the model was actually reasonably close." In there will be a thread of climate change, because we know the precipitation is changing due to climate change. That's an inference that no human has put into the model, but it has built an understanding of it as it went forward. That's how I think about the climate change challenge.

I think, just one more point on that, is I think it's really, really important that we drill deep on the climate change, because too many things are, "Oh, it's 1.5 degrees or it's 1.25 degrees." Those generalities mean literally squat in application. What you need is that farmer Fred Smith on his 2000 hectare farm in Alberta can do this thing and it will change climate change slightly. If you have 193,000 of farmer threads, which is roughly how many there are,

and every one of those 193,000 is doing that little bit, now you can have a generalization of a number that you can trust.

Aisha Greene

I like that you put it into context that anybody could understand. That also really encapsulates what is the value of AI. Bringing it back to how do we start to think of value proposition for digital twins across different sectors? We've talked about climate, we've talked about a little bit about financial services. How do people start to realize the actual value from these types of applications?

Arthur Berrill

I think your question is about applications, plural, and I think that's really the way to think about it. If you were to build a digital twin for the express purpose of predicting some individual factor right across Canada, it's going to be very expensive to put it together and it's going to be expensive to run. But if that same digital twin gives useful information to a commercial venture, that means the commercial venture is successful and it can show a couple more million dollars profit on the bottom line, and then you do it to another, and then another, and then you wind up doing hundreds and hundreds of them. Now you're talking about a significant value. In our particular case, we chose a very simple... Well, we thought it was simple. We thought it was a simple problem that we solved with our first run at the digital twin. It was valuable and it substantiated the effort.

Now we are looking at multiple applications of digital twins in various places and evaluating where's best to put the resources to build out all the... But the interesting thing is the digital twin we built, we're just going to build on top of that, because what we already have in there is a fully functional, well-disciplined, ontologically constrained knowledge graph that is working beautifully. Now we just need to... If you think about it, we have soil conditions right across Canada built into the digital twin, soil permeability. We understand transpiration rates on various vegetation. We've got detailed vegetation understandings right across Canada. It wouldn't be a stretch for us to now turn on flood modeling from there.

Or you could do what we did, which is look at something completely orthogonal to the problem that we set the digital twin up for and build that into the digital twin, and then see if there's some correlation between the two. One of the things we wanted to know is who is likely to inherit your farm? It's absolutely orthogonal to anything to do with climate change, but that piece of information might actually figure in projecting climate change down the road, because the new owner of the farm might be better educated, might understand things that they can do more effectively.

Aisha Greene

We're coming closer to time, but what's your message for those who are thinking about exploring this sector, don't know where to get started, and they may have data in various formats that they're thinking of maybe starting to model, what would your advice be to them?

Arthur Berrill

Definitely dive in. That'd be point number one. Point number two, go have a look at a project called The World Avatar. The World Avatar in the UK. It's a brilliant piece of work. It's a beautiful exemplar of a digital twin. It's fully functional and it'll give you some ideas. You can walk away from it and going, "Well, that's a little bigger than I can conceivably handle," but they've published all the source. It's all open source. We used it. In some cases, we had to go back to them and say, "You've got something wrong here, and we helped them fix it. They told us where they were developing new ideas. I think those are two good places to start. Incidentally, before we leave this, there's one challenge left, one more than a speed bump. We think we have an answer for it, but the problem with simulation is that when you run a simulation in the digital twin, by definition, it's going to change the digital twin.

In the bad old days, the way that was solved was that you made an exact copy of the original digital twin. Well, if you've got a 10 billion node knowledge graph, that's not a trivial exercise to do that. The World Avatar team put together a solution for that, which is a partial solution, and we think we've got the rest of the solution. There are some technical speed bumps. I think my contact details will probably be attached to this. If you get involved in it, by all means reach out and I'll see if I can point you in the right direction.

Aisha Greene

I want to expand on that challenges statement, because those risk and speed funds can also be ways that people might want to think about how do we jump over them. One, if you create a simulation of a simulation of a simulation, what's the impact of anybody understanding what is the true impact of the situation that you're trying to model? Simply because sometimes those simulations can have variables that you did not account for, or they might also be issues with the replication of the data. How do you account for that?

Arthur Berrill

I don't think you would run a simulation on top of a simulation. What you would typically do is you would do parallel simulations. The magic then is in the comparison of the results. I've run simulation number one with these controls. I've run simulation number two with these controls. If you use the parallel world approach that we're using, then you can say, "This result set and this result set, let's compare them and then discover the explainability of the difference." At that point, you, you're in pretty good shape, because you can say, "Okay, I can see where this simulation goes. I've simulated for two years, let's crank that up to four years because I have every expectation that the projections are going to continue to be reasonably sensible."

Aisha Greene

The effect of having the parallel simulations is almost like your validation check to make sure that what's happening in one scenario, you can remodel and replicate.

Arthur Berrill

Actually, that's a possibility. Where I was going with this was that the parameters that I use for this simulation, that I tweaked for this simulation, the parameters I tweaked for this simulation are slightly different. What I want to know is what's the difference in the results? How does that relate to the changes that I made in the first parameters? It's going to rain 12 inches tonight in Toronto or it's going to rain six inches tonight in Toronto. I got two simulation paths. Now I compare the two simulation paths and go, "Does the explainability factor work backwards to make sense here?" It's also a way of testing the digital twin.

Aisha Greene

I guess maybe as a final thought that you might have for people who are listening to this podcast and maybe even share an insight of where RBC is going, can you give us some reflections on that?

Arthur Berrill

Yeah. I did mention earlier on that we're examining the path forward. The POC is working and we're very pleased with it. There are some tweaks that we want to work on. For example, the parallel world scenario. The next job really is let's decide which extension to the digital twin we'll do next. I would seriously love to have outside influences on us. People come along and say, "Look, we'll give you some anonymized data. If you can build that into the digital twin and show us how the results are, that would be very cool." The neat thing is that we have one under our belt now, so we have some confidence that, moving forward, we can do some exceptionally clever stuff. It's more now a matter of where's the first value, where's the second value, where's the... Prioritize them by the value. Not necessarily just the value to RBC, but the value to the rest of Canada.

Aisha Greene

I like that call to action. We've also given everybody a little bit of a call to action to reach out to you if they've got some data that we can use for this. Arthur, thank you so much for joining us for this podcast. Really appreciate the insights that you've shared. We're looking forward for future conversations with you as well.

Arthur Berrill

Beautiful. It's my pleasure.

Aisha Greene

To you, the listener, thank you for tuning in to the Deloitte AI Institute Canada podcast, Demystifying AI. If you'd like to learn more about this podcast or other podcast within our series, please visit us on the Deloitte AI Institute webpage. We look forward to seeing you next time.