Intro: When you take a magnifying glass and look at a picture, you see that it's made up of thousands of dots and when you pull back, those dots become the details that create the picture. There's real power in the details, because when you have more detail, the bigger picture becomes sharper and wider, and a story emerges.

Hi, this is Amie Moreno and you're listening to "Seeing the Big Picture: Conversations on how Data and Artificial Intelligence can add the Details that Fuel Deeper Insights in the Life Sciences Industry."

Amie Moreno: Hi everyone and welcome to the podcast. My name is Amie Moreno. I am a Director in the Data, Advanced Analytics and Tools team in our Life Sciences group at Optum. Today we're going to be talking about artificial intelligence in health care where it's come, where it is now, and what we see the future looking like. And you're lucky enough to have our guest, Bertrand Lefebvre with us today. So, Bertrand, would you introduce yourself and tell us a little bit about your background, please?

Bertrand Lefebvre:
Hello, Amie and hello, everyone. I am Bertrand Lefebvre, Data Scientist here at Optum Analytics where I'm working in the natural language processing group, working on electronic medical records and specifically on the free-text part of the notes. My background is originally in physics where I spent a few years doing academic research, doing computational physics and working with a lot of data, which got me interested in statistics and machine learning and wanting to transition to more real-world applications and that's how I got into natural language processing years ago. When I first started working on virtual assistance, also conversational agents, and now on natural language processing on medical records.

Amie: Great, thank you. So, we've all heard about AI, artificial intelligence, machine learning, natural language processing. Can you explain to us a little bit about how those are not necessarily synonymous, but how they differ and how they all kind of fit into the grander scheme of things?

Bertrand: So, artificial intelligence has the grand goal of creating an artificial human like intelligence, which is not a very well-defined objective as it's not entirely clear how human intelligence works and how to define it, but it has multiple components to it, one of which is natural language processing. Natural language is how we humans communicate and how we not only communicate, but how we encode our knowledge, and this is necessary to create an artificial intelligence that will be able to understand and communicate with humans and communicate in a human-like way.

Machine-learning is another subset of artificial intelligence, which consists in teaching machines how to do tasks without explicitly programming

them, and how they do that is by learning from fairly large amounts of data while they try to identify some patterns and learn how to internalize what they've learned from the data to other types of data they've never seen before. And this has some pretty big intersections with modern natural language processing as well, which uses a lot of machine learning. So, all these are, both machine learning and natural language processing are sub-fields of artificial intelligence. There are some other sub-fields like computer vision, robotics, it's a complex, big domain.

Amie: What are some of the challenges that you find with the interpretation of natural language processing or NLP?

Bertrand: Well, there are multiple. In our specific domain, the medical domain, there are such a variety of EMR vendors, so many different configurations, there's no standard way to configure and use EMR's, there's absolutely no standards and because of that, that's a huge variability in what we get as data. At the same time, that means this variability means it's extremely rich where we get multiple perspectives from multiple caregivers, multiple institutions and that's the richness and at the same time, an obstacle in some way. The other is in terms of natural language itself, it's a fairly, very specialized domain, the medical domain, at the same time, it has a huge vocabulary. We're working with medical requests from all sorts of medical specialties, also other institutions and again, so that implies a huge variability. Added to that, human language is ambiguous. Human language is not always, there’s some uncertainty that's expressed in these notes. There are some inaccuracies sometimes. People don't always pay as much attention as they should because they don't have time. For example, when they write down these medical notes and these are changes from natural language processing, point of view, which makes natural language processing challenging and interesting at the same time from our perspective.

Amie: Yeah, with that in mind, it’s not only complex, but it's relatively novel, right? But we've seen in some of our market research is that there's still some concern about the accuracy of this type of data and so, can you tell me what you would consider is necessary to be in place, what kind of controls and quality checks need to exist in order to account for any of that variability?

Bertrand: There are, indeed, many challenges to natural language processing and natural language processing in the medical domain, we can talk to that a little bit more later on. In terms of quality check, there are multiple levels, which can ensure quality of the results. First and foremost, when we’re training machine learning models, it gives us data that has both the raw input, which in our case is the raw medical notes, the text, and expected outputs. So, for example, what kind of information should be extracted

from these medical records? This expected output has been labeled by humans, like domain experts, clinicians, very often, who defined what the expected output should be. But, some of this data is used to train the machine algorithms so it knows how to perform the information extraction task. And some of this data is used to test the algorithms, so we can score them quantitatively, so when we present them with some inputs that they have never seen before and make predictions that extract information and we can compare that to what the expected extracted information should be. I mean, that way we can quantitatively score the performance of the systems. So, that's one way to do it.

Another way, which is another safeguard perhaps, less quantitative, but would be extrinsic evaluation. So, once we have a system that is able to extract, on a large-scale, information from medical records, then we can build some models from this extracted information and measure the performance of these models or compare with known results. So for example, if we extract information from a large population of patients with prostate cancer and we extract the information about the histological type of the tumor, then we know that about 95% of the patients, the tumor should be of the adenocarcinoma type and so it's easy to compare if the information has been extracted correctly, if that matches the statistics, the well known medical statistics. That's another way to measure, to quantify the performance of a system.

Amie: And so just a point of clarification, Bertrand, when you're speaking about EMRs, which is electronic medical records, we're thinking of a specific either record or a specific platform, because that's the data that you're pulling in and I just want to clarify that as it compares to EHR, which you'll hear us say a lot, which is an electronic health record, and we say that because there are multiple different EMRs. Multiple different EMR platforms or multiple different interactions of care, whether it be at your PCP or a hospitalization, but the point in discussing EHR is that an EHR implies the holistic view of the patient journey across multiple EMR systems across multiple sites of care, etc. So, just to make that point clear.

So Bertrand, from your perspective, what is the value of natural language processing, what are the types of information and types of data that you can extract from this that makes it so valuable?

Bertrand: Well first of all, there's a lot of information and knowledge that's encoded as natural language, and particularly in the electronic medical records. Because these medical records are completely patient-centric, they are a narrative that usually covers a lot of information from the patient, family history, and medical history through the treatment pathways and it includes, describes some of the thinking process that the practitioners are

going through when treating the patients. So, this is an invaluable source of information for anybody interested in health care in general and how patients are treated and outcomes, for example.

So some of the information that can be extracted includes the disease and the conditions that the patient has. I mentioned previously the medical history, and that includes, as well, the family history that sometimes matter. There's also other factors such as social determinants, for example, that might affect how a patient is being treated because they cannot afford a specific treatment or how their lifestyle as well might influence certain outcomes, information about treatment pathways, how are they being treated and why treatments might be changed to a different one, what are the sequence of treatments doctors might decide to do for their specific patients. How patients are responding to treatments, that's a type of information as well about things like side effects that might occur. There's some mention of their adherence to drugs, which is another interesting part of information. What diagnostic tools have been used?

Amie: So can you talk to us a little bit about any particular project that you're working on or something you're excited about where you can see, even in the future, potential enhancements or just something that's a capability that you see in the future.

Bertrand: So, currently we have pretty big projects that focusing on oncology for a whole variety of studied tumors. It all started working on prostate cancer and we're working extending that to a whole other range of common and more rare cancers. There is so much information buried in these notes that's not available in structured data, information about the history of the tumor or the family history, the histological type as biopsy results. There are information about the metastasis and how and when they've been diagnosed, about the stages, information about biomarkers, for example. The biomarker is another important piece of information that is very interesting for people studying oncology. Oncology is a very big and a very interesting area with lots of potential applications.

Amie: So, Bertrand, what's next? What's next in artificial intelligence? What do you see are some of the big trends coming up?

Bertrand: So, first, I would start by answering what I don't see happening any time soon, which is the robot of artificial intelligence -- doctor, for example, or nurse -- or the artificial intelligence that is going to cure cancer. I think this is a field where one should stay away from overpromising, but what is really important is to have a dialogue to understand what are the right questions that we can ask, and we can answer right now. And where I see the future going is artificial intelligence including natural language processing being used to assist the caregivers, the physicians, assist the

patients, and assist the researchers to, in particular, unlock access to all the relevant information that they use to make health care more efficient and improve the outcomes of the decision-making, and integrate most seamlessly with the physicians' workflows, break the data silos, and make use of all the various types of data that is available, whether it's electronic medical records, whether it's claims data, whether it's genomics data, or even data coming from wearable devices or smart phones or also medical devices that measure, collect data of the patients. This is one of the, I think, the big challenges and exciting opportunities for artificial intelligence in health care in the future.

Amie: So, I heard you mention specifically wearable devices, which is a big trend. Do you think that natural language processing could influence or help interpret drivers of social behavior, for example?

Bertrand: Oh absolutely. There's some work being done on that. A lot of researchers are following Twitter, for example, to try to follow, for example, epidemic outbreaks or our social behaviors that is, indeed, a very rich source of information that could be as well integrated and used in addition to all other sources of data that we have and give them a holistic view of patient population. That is another area with rich potential to gather all this various information sources.

Amie: So, Bertrand, as we're talking about the different kind of social and behavioral aspects of data that we can tap into using natural language processing, what do you think some of the types of questions are that we can answer or some of the types of information that would be important using that type of data?

Bertrand: Well that can give a lot of insight into how patients are perceiving treatments, for example, or they see various drugs and treatments. That can also give lots of insight about their social behavior that might influence their health outcomes. For example, what they eat, how much social life they have, interaction with other people, what their interests are, and it's important to follow them, to follow the patients, not just when they go to visit the doctor and they're sick or on the information collected is going to be focused on what the current symptoms and conditions are, but to have a more thorough follow of their social life and social behaviors, over the year, and including are they healthy, see what kind of habits they have that might be changed for better outcomes. For example, what their diet is. As a lot of this information that is actually available on the social media that could be collated with more specialized medical information from the electronic health records.

Amie: So that can really help a physician, for example, understand different ways to speak to or intervene with a different patient based on some of that

information outside of just clinical, and I think once we start to take that data and that information and integrate it with clinical information and integrate it with claims databases, for example, that's really where you're going to get the holistic picture of a patient and really be able to improve health outcomes.

Bertrand: Absolutely.

Amie: Well Bertrand, I just want to thank you for speaking with us today. It's been extremely informative, and I appreciate you taking the time.

Bertrand: Well thank you very much, Amie, it's been a pleasure talking to you.

Amie: Great.