

### ALGO 119 - Module 3\_3

Hello everyone! Welcome to week three of the News Algorithms MOOC. We've got a great guest speaker with us this week, Tamar Charney is the managing director for personalization and curation at National Public Radio, better known as NPR perhaps.

This week we're talking about algorithmic curation and Tamar has some experience working with that in a news environment in particular in the context of the NPR One app which ties together some algorithmic recommendation techniques with some very careful editorial thinking. So Tamar, hello and welcome. Thrilled to have you with us today. Thank you for taking the time to share your knowledge with this MOOC.

My pleasure.

So, maybe you could kind of kick things off a little bit for us by giving us an idea of exactly what it is you do in your role as director for personalization and curation. It strikes me as kind of a new role certainly in a new job title, one that I haven't come across before. What does that entail?

Great. Well in some ways we made up the top title quite frankly. I think that from a very high level, I am not the journalist in the room when we're thinking about the algorithms and the technology that's needed to get a localized and personalized stream of content to our listeners.

Mostly on the NPR One app, but this also extends into smart speakers and the NPR app and probably into some of our other digital platforms. So my charges to really think about how could we use algorithms which I think of is really the modern editorial tool, to connect people with the content that they need, and want, and appreciate.

So my job is to really be that editorial person thinking about the editorial implication, thinking about the story arc and the breadth of our coverage, and having the conversations with the data scientists and asking the questions about "but what really would this do if we implemented this algorithm?" or "are there ways we could do this that might enable us to better connect people with the content they're going to value?"

That's great. So, could you go a little bit further and sort of tell us more about how the algorithmic curation and personalization works in the NPR One app. Perhaps the students in the book haven't used it, but maybe you can kind of give us a little bit of background about how it works, what it does, and what kind of information it provides.

Sure. So, NPR One and the real shorthand to understand it as much as I wish we had a better explanation of it is it's almost like Pandora for news. So it's a way of getting individual news stories that are localized and to some degree personalized based on your interests.

So, when you go to the app or you encounter an NPR One experience you know in your car dashboard or on a television set you basically hit play and it starts a continuous stream of content and the algorithm is

basically going through all of the content it has available and fitting various pieces of content into various almost buckets that we've set up based on various criteria that we the curation and personalization editorial team have setup.

So it starts by playing you a national newscast and then it goes and finds your local newscast from your local NPR member station and then it's assembling a mix of news stories from NPR from your local member station podcasts from NPR podcasts from your local member station and even podcasts from third party podcast providers that the algorithm has learned that you like.

So overall what we're trying to do with this is take the best of the decisions human beings can make about content that's really relevant and great storytelling and helps people understand both the news of the day and some of the breadth of the human experience that the serendipity in the driveway moments that NPR is kind of known for, and taking you know, the best of the decisions humans make about those factors and combining it with what algorithms do really well which is you know tracking behavior to some degree and knowing where you are localizing you and really trying to blend the two.

So we think of NPR One as that secret source between human curation and algorithms.

So, can you tell us a little bit more about the sort of the human editorial work that's involved in sort of keeping this algorithm humming along? What does that work look like and how does that work blend into what the algorithm is doing?

Sure. So, a lot of what the humans are doing is really listening to every single audio news story and pretty much a ton of the podcast work that's out there and making some decisions about what role it plays for people, how long is it going to be relevant for, what feels super important and in the moment today or watch us bring a little bit of of breadth and deeper understanding and helping the algorithm understand how a human would interpret these pieces and where it would fit in in the needs of a human being.

I think sometimes you know, I look at other platforms where an algorithm is very heavily used and you'll see something pulled that has this clearly incredibly popular and lots of people have engaged with it, but in a very fast rolling news cycle it sometimes is actually out of date and not the most up to date information, but many algorithms don't really have the tools to know that, that the story has actually advanced a little bit.

And what was incredibly popular and had high engagement times, an hour ago actually isn't the right story to be giving people this hour. So my team is really monitoring that because at this point in time that's something that humans do really well and really efficiently.

Now, in terms of you know, how this algorithm think that people are feeding into it, can you tell us a little bit about how things like the diversity of content or the mix of content maybe get reflected in the curation algorithm that NPR One embodies?

Sure. I mean I think the way to think about it is one of the main jobs of the algorithm is actually the whole localization piece to make sure that you're not just getting a feed of national content that you're getting content that speaks to the community that you are in geographically, but also starts to speak a little bit to the communities you identify with.

And through that you know, the way it's doing that is learning what podcasts you like regardless of whether you actively follow or subscribe to it. So the humans will place great podcasts with a fairly large breadth into the flow of content and see how you respond to it, and if you respond well to it and use you clearly had a good positive experience with it, the algorithm is going to serve it to you again.

If it was something that didn't really work out for you will probably stop feeding it to you. We also use the algorithm to offer other perspectives to use, so the algorithm can learn like "oh you know this person is listening to a whole lot of partisan podcasts, let's give them a chance," in case it was really inadvertent to hear a podcast from a different political point of view.

That's something that's just offered. Some people choose to accept us on it, some people don't, but that's something the algorithm has asked to do, or employed to do, so it's really trying to figure out you know how do we use the algorithm in an editorially responsible way to get you the news of the day?

And the curators are often determining what are those important news stories and then how can it be used to support certain interests that honestly if you filter bubble yourself on, that's OK, you know, if you like country music and I like jazz, I don't think there's any harm in me not getting your country music and you're not getting my jazz music.

There's no reason we can't support that interest. Same with books, you know, if I like reading literary fiction and you're really into true crime, we're probably not going to read each other's books. That's a great use of an algorithm, to get really granular in personalization.

Maybe you don't like music interviews at all, maybe you don't need to get them at all, and I think those are places where we use the algorithm to be pretty aggressive. There are other types of content where the algorithm doesn't do anything, it's all up to the curators.

The lead stories of the day, that is solely determined by the curators because we think that's where your skip button is the appropriate tool. If you don't want to hear what's happening today in Syria or with the government shutdown you can hit skip, that's fine, but we're going to offer it to you anyway because we think it's the news of the day, and we are a news app, and we want to make sure you're getting the important news.

So, what are the sorts of topics where you are sort of like directly addressing this idea of filter bubbles and trying to combat that?

I mean, it's mostly in the podcast space, political podcast. I probably can't go into a lot more detail on that.

OK, OK! I wonder if you could maybe talk a little bit about, you know, what is the curation algorithm trying to optimize for? Is it trying to get people to spend more time with the app overall or just listen longer to certain pieces or listen to more a greater diversity of pieces? Like what are the metrics that you're tracking and what are you trying to optimize in this algorithmic curation?

ure. We've really been aggressive about engagement. I mean we're really trying to create a really sticky experience for a number of reasons. One, we want people to really like it and if you really like it you tend to stick around for a long time, and we know that higher levels of engagement translates into people becoming supporters of their local public radio station which is a big part of the funding model. So that's how we use it in service of some of our monetization goals.

But I think from an editorial perspective the more people are staying with us I think as saying we're doing a good job both as curators and as the group that's employing the editorial algorithm in creating experience that is giving you the news you need and the fun and serendipity that makes you want to keep listening.

So we felt that that's a pretty good metric to be driving at, and so far we think it's been pretty successful, so we haven't felt a need to really be using it to try to create a different result.

So, you know, you've been building these systems at NPR for at least a couple of years now. I wonder if you have any advice for other news organizations or other people who might be thinking about creating their own algorithmic curation apps or widgets or newsletters or you know personalized home pages? all these kinds of algorithmic curation possibilities out there. Do you have sort of a broad advice for do's and don'ts, or you know places where you could go wrong things like that?

Yeah I mean, I think it becomes really fundamental to know really clearly who you are and what you do for people and to have data is the starting place for how you use an algorithm. You know what is the job you're doing for people not to get all Clay Christensen on you, but what are you trying to get this algorithm to do?

Because they are the modern editorial tools and you're trying to use the algorithm to support what it is you're trying to do for people call it your mission, call it your purpose, but to really understand what you're trying to get it to accomplish beyond just increasing page views or increasing engagement.

I think that's really easy to hit you know a Key Performance indicators metric. I think it's a lot harder to really deep in your bones know why you're doing it and really understand what your editorial guardrails are.

And I think organizations not being clear on some of those fundamentals about who you are and what you're trying to accomplish is really how we end up in these situations that filter bubbles where we're just driving at the bottom line or a certain number of people instead of thinking really deeply and hard about how we want the algorithm to improve people's experiences or further our mission as journalists and journalistic organizations.

In some ways this is my job in the room as the journalist in these conversations is to really ask these tough questions about: what is it that we're trying to accomplish? You know journalism is there in service of our democracy, and I think the algorithms need to work in service of that as well.

And I think we sometimes forget that piece of it in the midst of: oh my gosh you know we can optimize for this, that, and the other thing. Well, why? And how is that helping people? And how is that helping people understand our political system, and our communities, and what's happening you know down the street and in our schools?

And I think if we're really clear about that we may be able to use them for good and people aren't tearing ... I don't want personalization, I want the news. Well if the personalization is helping you get the news in a more efficient and effective manner, and helping you better understand the things that are important to you, you might feel better about that or differently about them.

So one of the sort of I guess more pragmatic elements to implementing these kinds of ideas (coming from an editorial background) is how do you work with a data scientist or someone who's doing the machine learning and actually building that technology? Do you have any thoughts for how to make that kind of collaboration fruitful?

Yeah that's a great question. I mean I think it's tricky because there's almost two opposing forces. On one hand we're down this road of data scientists and machine learning because it's really complicated, but from the editorial perspective we want this level of transparency, and we want to understand what it's doing, and why people are getting what they're getting, and how this may be inadvertently skewing what we're trying to accomplish editorially.

So we need to understand what the algorithm is doing, but it's incredibly complicated math that almost human beings can't understand, and I think this is where we need some some real thinking in how do we breakdown and really understand what is being accomplished and how those algorithms actually end up distributing content to people and what that looks like.

And you know, maybe this is where we need the people like you to lead us a hand in this layer in between the journalists and the data scientists to help create a way so that it's understandable, it's translatable, and that the two parties that need to be involved in making it successful can both talk to each other, understand each other's priorities, and understand what's being accomplished with it.

Have you found yourself having to learn some of the data science jargon to communicate? Or is this kind of more of a back and forth?

I think it's a little of both. I mean I won't lie, I've been in conversations where I've been you know the person like: wait, what? I totally don't overstep that, but I've also seen you know the data scientists and some conversations look at me and go.

Well that's just you know editorial fancy fancy language. No no, it's really important. And figuring out how we both understand the importance of what each other does because this is not going away. You know, we need these algorithms. The way people are getting content on modern platforms require the use of them.

And we have to figure out how journalists and data scientists can really work together and accomplish what needs to be accomplished. Both you know, increasing performance and also increasing the value of our journalism content to people, and that's that's going to be tricky to figure out. I think that's some important work we need to do.

For sure! So, I want to actually pick up on something you mentioned a few minutes ago because next week we're actually going to be talking about how how journalists can be more responsible with the algorithms that they're building in terms of being more transparent with the systems that they're creating.

So I wonder if you could sort of go into a little bit more detail on how do you and how does NPR think about transparency with respect to the curation algorithms that you're developing and using?

Quite honestly it was probably two years ago that I wrote a post for the NPR.org site explaining what we're trying to do with the algorithm and how we use it so that people understood you know, if you're using NPR One, you are going to get the day's news and you are going to get serendipity and you're not going to just get a very narrow slice of the news.

So I wanted to be really open about that and really pull back the curtain a bit so that even if we didn't go into the nitty gritty about exactly how people understood what our philosophy was toward using it and to reassure people that this hasn't just been left to the technologists and the data scientists no offense, but that there were journalists in the conversation actively looking at what is being placed into the system and probably also sort of mentioned this evaluating what's coming back out.

All of the curators have multiple accounts that we use to listen to NPR One to see what various people might be getting back based on what other things they've heard and what's in the news that morning just to make sure that we feel good about what's coming back out from that algorithm for people.

That's fascinating. I mean could you could you talk a little bit more about sort of I guess the monitoring work that your team needs to do? Like how does supervision of this system work?

Yeah, I mean we have various reports that we can run that basically show us what pieces the algorithm would be pulling and then what some of the fallback or backup options would be based on a different ways a listener might be listening.

I think I have maybe 12 e-mail accounts I used to listen and the members of my team also have multiple accounts. It's not to say we listen to all twelve of them or you know multiples in any given day, but we're often rotating between accounts just to see what else might be popping up because everybody listens to

different things and the system knows what you heard yesterday or two hours ago, and you might get a different ordering of pieces and sometimes ordering of pieces can feel differently.

When you're producing a radio show you think very hard about the sequencing. What you can't control when you're programming your radio show is when people tune in and out. It's almost the absolute inverse with NPR One.

We know when people come in, we can't always control exactly what sequence they hear things in, or what they did in their last listening session, this current listening session could be a little bit different. So in either instance you have some loss of control of the experience, it's just a different type of loss of control.

That's really fascinating. So, what's next for NPR in terms of this whole area of personalization and algorithmic curation?

Yeah, I think it's thinking about what other platforms it makes sense on. I think it's also how do we as journalists get more comfortable with using it as an editorial tool. I think it initially can feel a little scary to journalists because there is this loss of control and we're used to.. you know, stories are written in a very rigid format, radio shows are produced at a very rigid clock, interviews we do on the air have a story arc.

There's been a lot of control and the use of an algorithm does require some relinquishing of control, and some trust that you've set up good rules, and you set up good templates, and you've put the right things into the hopper, and I think it's us as an industry just getting comfortable with that and figuring out what outcome we want from it, and making sure that what we're getting out [Video skips]

Is there anything else that you think the MOOC students should know as they're kind of heading into building their own algorithmic curation, applications, and systems? Any final advice to them.

I mean again, I think they've to be really clear about what you want to accomplish, and not just you know, the number of people you want doing X, Y or Z, but really why? Why are you using it? What are your journalistic goals with it?

And always come back to that because I think that's an important conversation to be having with the data scientists.

So really making the algorithm you know, do journalism.

Yeah, exactly. That's a great way to put it. Yep, because you think really hard about what you want your front page to look like, or what you want your radio show to sound like, or your TV broadcast.

What do you want the output of your algorithm to be on the platform that's driven by it, and to make sure that it is as an editorial tool doing and accomplishing what you want it to accomplish.

Yeah. Well this has been fantastic. Thank you again for joining us today. It's been great to have you here and I look forward to listening to some NPR One later today.

Great! Thank you Nick, my pleasure.

All right. Thanks.