Transcripts - News Algorithm guest speaker Christina Elmer, Der Spiegel

Hello. Welcome back for the final expert interview of the MOOC. Today we're joined by Christina Elmer who is a member of the editorial board at Der Spiegel in Germany. Der Spiegel has done some really really interesting investigations into algorithms and society as part of their coverage over the last year or so. And so I'm really looking forward to learning more about those types of projects.

Cristina Hello and welcome. It's really wonderful to have you with us today. Speaking to us about these investigations. Thanks very much for sharing with the with the MOOC today.

Of course. You're welcome. Thank you.

So I want to kind of dive in you know some of the some of the students of the MOOC it's a very international class. They may not be familiar with Der Spiegel. I wonder if you could sort of start us out by telling us a little bit about the publication and also about your particular role there.

Yeah. The Spiegel is a German news weekly magazine which has a really long tradition that has been founded in 1947 I think and since 1994 we also have our own news website Spiegel Online. And this news website has its own independent editorial team and the data driven newsroom. We reach around 21 million readers per month. So this is the unique users number. And most of them are in Germany right now. And I started as a science editor there at Spiegel Online and then could boot up the data journalism department which is now consistent or four colleagues with different skill sets. Two of them are journalists who learn to code and two were originally urban planners who started a blog and then published their own things and we found them by this. So the team is quite diverse in its backgrounds and it already combines many different perspectives in itself.

So I had the pleasure to lead this team until last year and now I am working as a member of this editorial board at Spiegel Online so I'm fostering data journalism but also the overall further development of the site or designs towards and workflows for example.

Great. So you know one of the topics for the MOOC this this week is about how to investigate algorithms in society. And I know Spiegel has done several of these investigations. You recently published a chapter in the data journalism handbook. I wonder if you could sort of tell us a little bit more about some of these investigations that you've been undertaking there.

We did this because we believe that algorithms influence many quite important areas in our society already. So for us it was quite natural to go into this direction because we believe that journalists should be able to investigate on them. And there has been a constant discussion on algorithmic accountability at Spiegel Online in this in the years and the recent years. But initially only as an occasion for our reporting not in the form of our own research or analysis project covering and analyzing algorithms.

So we started with these projects basically as a collaboration project because as you know already our data journalism team is quite quite small and we don't have a big data analysis unit in the background for example. So we started doing such projects in 2017 in the run up to our federal elections in Germany. And then we joined forces with a German NGO named Ableism Watch. They are quite early experts in this field and they helped us to gain insights into the presentation of Google search results. So we did the crowd sourcing with them. They did the analysis and then we published the results together and yeah through this project our data journalism department supported the planning and the methodology and the methodology to evaluation of the operation. So it's more or less has been a part where we have been partner of the project but did not do everything on our own was a nice thought into that topic. And as a result it turned out that the personalization was not as intensive as expected. So we could somehow debunk a widespread myth by doing this investigation and just as it was not the result that we had expected as well.

So this is sort of looking at that the filter bubble and the Google search results.

Yeah. Well it was about about the question if this filter bubble really existed and if especially if it existed and in a sense that you get different search results based on your local placement like like your IP address. People always thought that this would be the case and that this would really distinguish the different filter bubbles. But we found out that when you look at the first sides of the news results of the Google search results you don't see that many differences there. So yeah at least according to the to the to the place where you where you are. So yeah.

And at the same time also in the run up to the elections we collaborated with ProPublica and they they had done this ad collect a project in the US and also wanted to do it in several European countries and it was perfect to try this out in Germany. And the goal here was to try to see if Facebook ads targeting has been done by the German parties and to make this transparent. So we thought they they will of course use this option to get to directly address certain target groups for example with their messages. And then again we found out that those parties did not really use ads targeting that much. And of course it was quite interesting to collect all those ads and to get an overview especially on those local ads. This was something that they did. They only published the ads of local political groups in this area which totally makes sense. But when you look at this from Hamburg you don't see them. So. So this was a helpful project and we could make them transparent. And there was this this site where we collected them like a catalog also accessible to our users and readers.

Yeah. It was not the result that we had expected again. And last year this topic of algorithmic accountability reporting really got interesting because we we did and project on an extremely powerful algorithm in Germany the Schufa credit report and record which is used to assess the credit worthiness of private individuals. And this is quite important here in Germany when you apply for anything for a bank account for a mobile phone. For what else. Especially for a flat. You need to have your Schufa report ready to to to prove that you you really can and can pay

the rent. So this is this is really one important algorithm in Germany and it's not transparent at all. So we had to do crowdsourcing again and again we have joined forces with another data team that we knew beforehand quite well in Munich at the Bavarian broadcaster. So we have different target groups. We had different. We are publishing in different areas. They are doing broadcasting. We are doing the magazine and online news so it's more or less it's fitting together. We don't we don't do each other any harm in publishing these stories together. And this worked out quite well. And we also joined forces with two NGOs in Germany because in fact it was their their topic first and they had the idea to do it.

So we did the crowd sourcing together and then the two data teams and the media data teams did the investigation with, based on those data sets together.

And what did you find in those data sets about the Schufa credit score?

We we could not really find out how it works in detail in every aspect because we had I think around 2012, 2500 different reports crowdsourced. So it was quite a lot but not that many to really investigate on every different or every single aspect. But we found that the source of the score privileges older and female individuals as well as individuals who changed their addresses less frequently which was quite interesting. And we saw for example that there have been different versions of scoring algorithms used within this huge score. So with different outcomes for people with the same background information.

So depending on which bank you you are having your money on, you can get different results. So some banks used older versions while others have already updated their systems and that has not been made transparent at all and reports and discriminated against certain groups of people. So there have been some some yeah. Some results.

So you know it seems like across all these products you mentioned a lot about collaboration and working with other teams. I wonder if you could sort of elaborate on that and sort of explain more about how you organize these types of projects both within your own team and your own organization and also working with these with these other organizations and groups like what kind of skills do you need on these teams in order to pull off these really big projects.

I think you definitely need your own data journalism department besides all those collaboration possibilities because you need the colleagues who can handle huge sets of documents and huge data sets in-house but they should also be able to write the necessary code and assess the journalistic relevance of the results because this is a back and forth that we always have in such projects that we see something in the data sets and we have to ask ourselves Is this really worth it? And worth it and is this really a result that we would publish and yet concerning or relating to our journalistic relevance criteria. So we have to have both on board and I think you often need to have a methodology for understanding the team how such a crowd sourcing process can work because you have to set up this like design framework for us with all your

partners. I think many other things can be done via collaborations with other data sets or NGOs or scientific partners for example.

But I think it's crucial to have your own team working on the data set with a journalistic view. And we saw that. I mean we brought together different factors from different systems so we first had to discuss our relevance criteria. We had to discuss the requirements and the capabilities beforehand. Also how would you report on the results. What would you say? What what is the one sentence that we all should be using not in terms of PR but in terms not to get anything wrong. And we of course had working towards that we use together for example with the Varian data team. We had Slack channel that was on our workspace and select only set up for this project because it's so big that you really need some. Yeah I won't call Slack elaborate but you need to be to make more yet. You need such tools yeah.

And then what about sort of the external collaborators. How did you think about what they were bringing to the table in terms of these projects?

I think they could approach the topic from a different angle. That was quite helpful. For example the NGOs could really fight for the crowdsourcing and really come up with with their really distinguished opinion about making this transparent and also be a bit more aggressive on that. I would say. [Video skips] it was more like we we what we are not neutral because we think transparency is really important but concerning the algorithm itself. We want to be be more neutral compared to the NGOs and we should also be independent from them. So because in the end we publish results and we want to be taken seriously by the public but also by the Schufar for example which is a company making a lot of money out of this algorithm. So we had to stay independent. So this is what NGOs can bring in and concerning the scientific partners. Some of them have also been involved in the Algorithm Watch Project in the run up to the elections. They can really bring in methodology and capabilities in terms of analysis skills and Data Handling and the like like they could program the the Firefox add on that we use for example to collect all those or those search results whereas ProPublica for example could code this add on that we used to collect those Facebook ads we used a Facebook add on. Yeah. No it was Firefox add on to collect the Facebook ads but I think we could not have done this internally. We don't have the people here. So yeah they're different different skills but also different position of of those external collaboration partners that can really help.

Right. Right. So now that you've worked through several of these projects I mean is there anything that you've noticed is particularly difficult or challenging and kind of pulling together these investigations.

I think one one challenging aspect is the project management because you need project management to skills at least to to really foster these investigations with many partners and with many partners in house and externally. So so this is something that is not widespread in journalistic organizations right now. So this is really important. And in Germany the most challenging aspect besides that is that so many of the algorithms that private actors on the

stage is working with are not transparent right now so we have to set up experiments that can reveal at least some discriminating factors or effects based on certain sets of outcomes but not the algorithms themselves. So we would never be able to get a holistic picture of what they do.

If we can investigate the entire system so this is not that not that great about Germany right now and in most cases we don't even know of algorithms that are used or not used for example in predictive policing. We know that some units are using them, some are not using them and there is no information available at all. So it is a serious use problem if you do those projects in Germany right now.

Wow. Seems like a really really challenging beat to work on but maybe also that that challenge can can motivate some more people to get interested in pursuing these things. On that note I mean what kind of advice would you offer to other journalists or other news organizations that are thinking about getting into the algorithms beat so to say. Yeah. What would be your advice to them?

I think the most important thing is to do it anyway. Just um. I mean this is such an important topic and we have to investigate on that. We have to look at algorithms and in our case I mean we we are not that strong as a data team. And are these algorithms are not that transparent. But at least we could do something. So any, every bit of transparency is really meaningful now or useful. So I would first say you should do it. And besides that yeah collaboration can be extremely helpful as we discussed with other teams with scientists, actors from different systems who can approach algorithms in different ways. And it's also internally and newsrooms for example some of those collaborations might not be that yet established in the past but it was really really easy to to tell everyone why we do it this way and why we have to do it this way. Then I think you should really take some time to set up a meaningful research design in the beginning and also it would be use useful to speak with computer scientists or social and cultural researchers about this topic because in the end the reliability of your results really is based on this research design.

So it's the scientific process in the beginning and I think the most important thing in the end is to to define your goals in a sound and comprehensive way. Because I mean first strong goal might also be to raise awareness because many many people don't know about algorithms that are working in certain fields of society so this is a goal in itself. It's really important.

And don't try to deconstruct the algorithm or to investigate every discriminatory effect it might have because I think you wouldn't really need extensive datasets to do it in most cases they are not available. And from the reader's perspective, I would say that the algorithm itself is not the most interesting thing anyway. So they are really much more interested interested in the results and what their personal impact might be that they can feel in their everyday lives. These small bits might be also really important for the readers. So the reader's perspective is really important. I think.

So something is better than nothing I guess. And just dive in. I think that's I think that's. I like that approach.

Well so another topic of the MOOC this week is sort of more about transparency and how journalists themselves can be more transparent in terms of the data and the algorithms that they might be using in their own work.

I wonder if you could maybe tell us a little bit about how how you and how Der Spiegel sort of approaches issues like transparency for big data journalism or big computational journalism projects.

I think also it's a concerning this topic. You would need to distinguish distinguish between different audiences here because when we think of our readers I think they really need explanatory article about articles about our workflows and that I really see that we are really transparent about methodology methodological problems and gaps in our datasets for example.

We of course link to our primary sources whenever this is possible. If it's not a link we can do this of course and also publish raw data sets as a signal to them that we of course are willing to to also share our sources.

But in my view it's it's really crucial not only to make everything accessible but also to explain the limitations and implications. Oh sorry for the sounds. And in another way. So I think besides that you could also think of maps or tables like just to it's to to make more raw data points accessible that are interesting for reader groups but I don't think that the yeah that our code for example is useful in this respect.

So on the other hand side when when we don't think about readers only but also about our colleagues from different data from other data journalism teams or scientific researchers. We of course share much more. We we share our methodology and of course explain our workflows. We give them insights to our work for example be it on conferences or indirect prisoner exchange. And there are also some students that work with our datasets.

So we do this but not publish everything on the Web site because this was would be really really hard to do given that the team is quite small and we would have to to really clean up everything afterwards in a much more extensive way than we do it either way. So yeah but.

So it creates additional work basically having to clean up and make sure it's publicly publishable.

Yeah of course. Yeah. And we don't think that it's really some some some kind of signal that that our readers really need. I think for them we should really be more explicit in explaining our methodology and being transparent about those gaps and those problems that we have.

And what you can you can understand and what would be a misunderstanding like those examples for example. Yeah but that. Yeah. When it comes to other colleagues or researchers for example we we and the direct context you can also of course talk about those those problems that that are occurring in every code snippet for example that you do.

This is really really helpful Christina. Any other parting words of words of wisdom from leading these teams and sort of either investigating algorithms or trying to be transparent with algorithms.

I think when it comes to algorithms and their role in society for example it's really really important to have a diverse team. Oue team could be more diverse because you really have to see those effects and those topics and it's so much easier if if if there are people on the team that are really bringing this in from their personal view and from their reality in life. So yeah it would be really really helpful to have a different, different fields in society or different regions or different age groups in your team.

Of course a woman. And this is this is really really important because this is where the really good questions come from from your everyday life. Also as a data journalist. Of course you should also listen to your readers and should really ask them what what they want to know or what they experience in their everyday life.

This was also quite helpful for these projects but to get in many different perspectives I think is a crucial point and because this really gives your feelings for what an algorithm does and how it discriminates maybe in certain groups in society.

I think that's I think that's perfect perfect advice and in a perfect way to close the interview. So thank you again for spending some time with me and with the MOOC today. It was really great to have you here. Thanks.

Thank you. You're welcome. Thanks.