

SWDTP Webinar Analysing media discourses-20260505_120015UTC-Meeting Recording

5 May 2026, 12:00pm

1h 5m 3s

JC Jonathan Chow 0:06

Good afternoon, everyone. Good to see some familiar names, good to see some new names as well. Welcome to another session in our data analysis webinar series. My name is Jonathan from the Southwest Doctoral Training Partnership. Two very interesting presentations today on analysing media discourses and also from quite different kind of disciplines of research as well, which is great. If you've never been to one of our sessions before, the way this works is we'll have two 20 minute presentations and there will be 5 minutes for Q&A after each presentation. And if we've got any time left, we will just do a general Q&A. The chat function is turned off, but we have the Q&A feature turned on, so please do use that to drop questions throughout the session. If you'd like to ask a question by turning on your mic, please raise your virtual hand when we come to the Q&A. and I will enable your mic for you. Captions are available. Just be mindful, they are automatically generated. So I think without further ado, I will introduce our first speaker of today, who is Yusuf Olaniyan from the University of Bath Department of Education. Over to you, Yusuf.

YO Yusuf Olaniyan 1:36

Yeah, thank you, Jonathan, and thanks everyone. I was just sharing my screen quickly. So Jonathan, do you want to confirm if you can see my screen?

JC Jonathan Chow 1:57

Yes, yes, I can see your screen.

YO Yusuf Olaniyan 2:00

Yeah. Thank you, Jonathan.

Yeah, so good afternoon everyone. My name is Yusuf and I am a third year Phd student here at the University of Bath and I'm SWDTP funded Phd researcher. So in this presentation, I'll be sharing about

one of my research in IT during my masters of research. So it's basically trying to share, it was to you when I was still learning about media analysis and discourse analysis. And this presentation is more of a learning journey for me as well, because it's a very new thing that I did. Unfortunately, this particular presentation or this paper that I'm presenting, has been published. So you can scan the barcode here and it takes you to the page, like the DOI of the paper.

So without wasting much time, so I'm going to be sharing about...

this research and it's about higher education in Nigeria. So when I started thinking about this particular research, I was thinking about higher education in Nigeria and what the media is saying about higher education in Nigeria. I think specifically there was a very, very disturbing case of a man who completed a university education in Nigeria and after a few years went back to the university to return a certificate. And he kept saying that, get your, like have your certificate back. The certificate you give to me is a scam and it doesn't add any value to my life. So all of those conversations and the things that I have also been observing from the media as well, I kind of influenced the way I think about or why I decided to have, do a research around this issue. And basically what I was trying to do is try to understand how is our education constructed in Nigerian media. But particularly, so I don't just do everything without doing nothing. So I decided to narrow what I'm looking at in the media on post-COVID. Unfortunately, there is a curriculum restructuring that also happens in Nigeria post-COVID as well in 2003. So I was looking at that media discourse from within that year, so it was post-COVID. Yeah, so just to have, just to give you a little bit of some theoretical justifications of why media, in addition to my personal observations in the media, and of course, I'm reviewing literature and looking at the gaps on how I can contribute to those conversations. So there is this Mendez. I love the work of this particular person. So you shared a lot of things about media and how media can shape public understanding, especially our education and how media can be a gatekeeper, especially you are the voice to speak in the media. Anna doesn't have the voice to speak in the media and how media can also be a form of legitimising policy reforms in the media. So and it could also be some site of institutional discourse as well. So these are just some theoretical thinking around a media or if you want to do media analysis, these are some of some of the

justifications behind doing media analysis.

So methodological design. So how did I do my analysis? How do I come about thinking about this analysis? I want to look at what the media is saying about higher education in Nigeria and how do I look at this issue? So I had three phases. So the phase one is about quantitative kind of what happened in the media. So how many articles have been published about higher education in Nigeria? So I was trying to look at that and

count that numbers just to have a frequency, just to have an understanding of how the distribution across month, I want to see how much of the conversation is going on in the media. That was the quantitative scan that I got. And I went there very inductively. I wasn't going there like for specific search criteria.

At this point, I was just going in there. We were just looking widely in the media. Anything about higher education, I was just scanning them. So then I moved to thematic coding. When I have all of those things, then I had to use some search strategies to kind of look, like look closely into that data to do some open coding. Then I do some DHA, which is discussed on historical approaches as well, to further analyse my theory.

I think I'm coding.

Yeah, so sampling strategies. When I was thinking about selection of media to work with in Nigeria, there are a couple of them in Nigeria. So I decided to select three of them. So the Punch, the Daily Trust, and the Daily Post as well. So you can see when I search generally about...

our education. So if you look at this total education record here, so 637 from Punch. When I do a kind of a search strategy, like some keyword search strategies in the 637, I was able to narrow it down to just 200 and 200 and 69.

So, 637 article is about higher education within that time frame that I'm working on, 2021 and 2022, that I like the post-COVID period that I talked about, and I had 637 articles on higher education, but higher education what the what was actually common, the themes that are actually coming up from 637.

And the entire article that I was looking at was about apprenticeship, vocational training, labour market, and so entrepreneurship. So these are the things that were coming up about education. So all of this are like 269 for Punch and Daily Trust like 109. So if you tie it down to 52.

and this daily post like 545 and this is just like 10 of them. So from total of 1289 to just 331 articles that I eventually used for the phase of three. So and here, so these

are just the rationale.

why I selected these free media outlets. So because of their wide readership, because of their national coverage and editorial diversity as well. And I was very particular about excluding some things, especially some campaign materials, because I want to be objective to a very large extent, some advertisement as well.

I was very particular about them, so I removed them. I didn't include them in these numbers that you're seeing here. So these are just my exclusion criteria. So my inclusion criteria are this and these are my exclusion criteria.

And quantitative mapping, you know, I said I did the analysis in three phases. One is the quantitative mapping, just to see how much of the conversation around our education is going on in Nigerian media, especially the post-policy reform and the COVID. So here, this is a, this is like a graph.

And if you look, after this presentation, you can even look at the publication. There is a couple of them, but I just added this one. If you look at this graph here, higher education related news in 2021, that's great. In January, you can see what's going on here. So higher education narrative construction 2021. So here, the blue one, you can see the blue one here. So this green one is more of anything about higher education related.

news is just here. It doesn't have to be, I don't care at that moment what they're saying about higher education. As long as you're talking about higher education, yes, I shortlisted them here. So then higher education narrative construction, because some articles just talk about higher education without talking about what I'm really looking out for. So

the ones that are actually constructing our education, talking about that, doing some ideological reshaping or kind of pushing some narrative about our educations. So they're like the blue ones there. And this is for 2022 and narrative construction for 2022. This is like the frequency count. And I did some analysis of this. I look at the trends in this publication as well. And I observed that it was when that policy reform happened, and that was like the research in the number of the publications that were released by higher education because of the policy. And that kind of helped me to trace the ideology. What are people saying when that policy was released? Why do we need that policy change?

Who does that policy change is going to serve, or what intentions behind those policies? So, these are these are one of the advantages of doing this quantitative mapping. So, from quantitative mapping that I moved to thematic coding and

discuss historical approach. So, looking at...

this number of articles talk about this number of articles talk about this, and I do a close thematic coding, then, okay, what are you actually saying using this guy framework, which is, which really, really resonates with what I'm looking at. I know a lot of people use for cloud three-dimensional, but this Real and Wodak 2016 model actually

resonate with what I want to look at because Nigeria has a colonial history and I needed a discourse analysis approach that could help me uncover some historical context about what I'm researching because I also want to really have an understanding of how colonisation history kind of influenced the way Nigeria think about higher education, because Nigerian higher education is built on colonial structure. So that's the phase two now. So the phase two and phase three. So this is just like what DHA, this is the equal approach, this is their own thinking about how to analyse

discourse or how to do discourse, basically. So one is they try to, this is like, if you look at this diagram, it's from their book that I have. And if you look at the inductive and deductive and different, the discourse approach that is out there, and how this different discourse approaches influenced this particular discourse historical approach. So, and they talk about different things around critique, around ideology, power, discourses. This is what their own discourse historical approach are actually tracing in media or in a particular text. These are things that you actually are particular about. So, which is

very similar to what I'm looking out for in the media as well. So the critique, the ideology, the power, and the discourse as well.

So on DHA in practice, so thinking about how they think about using that particular model to analyse data. So very similar to Fair Cloud three-dimensional model, they have their own three-dimensional model as well. And the first one is about content and topics. You look at the content, what specific topics and theme are you being addressed?

In my own case, it's about education. So what strategies are used to construct the meaning and legitimacy, which is like that, that discussive strategy is very, very important to me, like to my study, because I use that discussive strategy a lot to kind of analyse my themes for that and the linguistic realizations. I don't think I use them. step three of their model, the grammatical, lexicon, and theoretical devices. I wasn't really interested in that step three, but I think the step one and step two were the

method or the stages are very applicable to my own study. So the five discussive structures, you can see them here in action. You can see the nomination is about how actors and event are named and referred to linguistically.

In my own research, I was looking at higher education was being named as like a workforce pipeline or something that should produce a graduate. That was what was going on there. The actors in the media and the government, the ministries and the local authorities, that's how they are talking about higher education. So and the predication as well.

So, the kind of statistics and the qualities that attributed to this to these things, they're talking about the argumentation, the perspectivization, and intensification as well. So, these are like their discussive strategies, and you will see how the very how they were applicable to my studies varies in few slides to go.

So this is just my coding process, just for transparency and audit trail, because this course as well, oftentimes when we do discourse, most people forget to talk about transparency and audit trail because I'm a Nigerian myself. I'm very, very much in the Nigerian media and there is a way that shaped the way I think about higher education.

And I think about the media as well. So these are just the things that I do to just make sure that my own positionality doesn't really influence what I'm doing. So open coding was what I did. The 331 articles that I should listed after the frequency count. Then I framed the identification of these.

Of this articles, then I apply DHA, then I do some contextual reading of what I come up with.

And yes, this is DHA in action. So this like more specific, I added more specific here. So nomination. So when I shortlisted the 331 articles and I was reading those articles, I was actually using each of this discussive strategies by Woodak and Ray. I was using the discussive strategies to look at those articles closely.

One, I was looking at how social actors are being represented. So higher education there, the social actors they are saying higher education is an engine for productivity, is the solution to unemployment, it needs to be a driver for national development, and it's a workforce pipeline. That's what they're talking about. And the quality and characteristics attributed to social actors and how actors are described as well.

So they're saying universities in Nigeria are struggling with the quality of the graduates, they're lacking skills, not industry ready, whatever they are doing in the university and the labour market are not connect, are not connecting. So these are

like, but are they predicating what they're saying? So our argumentation, this is more of a how they are legitimising that thing to be true, like yes, it needs a serious attention and it needs to change and reform needs to happen. So that was those ideologies I was trying to look at and it was because of their discursive strategies. I was able to look at those ideologies in the media and in historical embedding as well, the colonial manpower logic in Nigeria. and the marketization thing that was coming in in the conversation that I observed and in the liberal policy of borrowing as well. So after that policy, the Nigerian government eventually launched a student loan, which is very similar to the UK as well. So all of those conversations were what I was tracing with this discussive strategy.

And after tracing the discussive strategies and doing the theming analysis, I was very interested in whose voices are dominant and whose voices are silenced in the media. How many, like all these articles that I'm talking about, if there are 31 of them, how many of who are the people speaking in these articles or in these media articles? I saw that just 64, 63, approximately 64% are the government officials. I mean, the ministry, the local government, the minister for education, these are the people that have the most voices in the media. And just 5.4% is the students, which are like the consumer. These are the major consumers of my education and they are like the least, like just 5.4% of their voices are represented in the way our education is being shaped in Nigeria. And this few 7% are NGOs and 24% are the academics. So it was very concerning as well. So this just brought that knowledge of whose knowledge count and silence in policy. So yeah, thank you very much. If you have any question, I'm happy to take your question up.

JC **Jonathan Chow** 18:22

I love, love just the how to kind of nature of your presentation. Really a great resource for a lot of our audience today who obviously are PGRs, ECRs themselves. I already have a hand up from Sharon Walker, so I'm just going to enable you to unmute yourself. You should be able to do that now, Sharon and us.

Yo **Yusuf Olaniyan** 18:45
Yeah.

SW Sharon Walker 18:49

There we go. That's lovely. Thank you. Thank you very much for your presentation, Yusef. That was really interesting to learn about your research and to see some of the things that are coming out of it. My question relates to software. So, for example, when you spoke about, you spoke about

YO Yusuf Olaniyan 18:52

Thank you, Sharon.

SW Sharon Walker 19:08

doing a frequency count, keyword count. You talked about uploading, preparing your corpus, and you also spoke about doing a quantitative mapping. I want to know the software you use to do that. What did you upload those articles to? What did you use to do that?

YO Yusuf Olaniyan 19:12

Yeah.

SW Sharon Walker 19:27

quantitative mapping? Are you using something like search engine? What are you using to do that? That's what I'd be really interested to know.

YO Yusuf Olaniyan 19:35

Yeah, thank you. Jonathan, can I go?

JC Jonathan Chow 19:39

Yes, of course. Go ahead, Yusuf.

YO Yusuf Olaniyan 19:40

OK, OK, so thank you, Sharon, for that question. So, unfortunately, when I did this research, it was in 2021 and there wasn't even AI or something to help speed up the whole process. So, and it was also very, very painful that the media outlet that I choose, do you not have that capacity of kind of that search power of filtering. They don't have the filtering functions in their

website or on their website. So I did it manually. It was very tough. It was very tough. I had a very big Excel sheet where I was counting. I was counting each and every article. So it was a very challenging task that I did.

Yeah, that was why I had to leave it to that, 2020 and 2021, because it was 2021, 2020 was very challenging. So I did it manually without the use of any app. Yeah, and I use XLS well. I use Excel to kind of count and do those chats.

JC **Jonathan Chow** 20:41
Wow.

YO **Yusuf Olaniyan** 20:41
Does that answer your question, Sharon? Yeah.

SW **Sharon Walker** 20:44
Sorry, I'm just trying to figure out how to use the buttons. Yeah, that answers my question. That must have taken you a lifetime, Yusef, to do that.

YO **Yusuf Olaniyan** 20:46
Yes.
Yeah, yeah.

SW **Sharon Walker** 20:53
Okay, alright.

YO **Yusuf Olaniyan** 20:53
It was when I was doing my masters of research, so I had I had loads of time in my hand, so I spent close to a month doing that.

SW **Sharon Walker** 21:02
Okay, all right. Thank you. Thanks for that. Thank you.

YO **Yusuf Olaniyan** 21:05
Yeah, you're welcome.

JC **Jonathan Chow** 21:07

Wow, lots of blood, sweat and tears, Yusuf. Got another few more hands, but obviously if you would prefer to just type your question, please do make use of the Q&A feature where you can just drop a question there. I'm going to now allow Diana to unmute.

Here you go, you should be able to.

DD **Daiana Damasceno** 21:31

Hi everyone. I hope you all can listen to me. First of all, thank you for the presentation. Just a quick presentation. I'm Diana. I'm currently researching global partnerships and equity international education, mainly higher education. Your analysis of Nigeria media narrative is very insightful.

Thank you for that. Based on your findings from 2021 and 2022, what interferences can you make regarding how the media addressed the tension between internationalisation and the local demands for DEI, diversity, equality and inclusion? in Nigeria High Education.

Thank you.

YO **Yusuf Olaniyan** 22:14

Yeah, thank you, Diana, for asking the question. So, you know, as much as that was not really what I was actually looking for when I was searching the media, I was more of what is Nigerian media saying about education, but what you just mentioned about internationalisation was actually very

It was very imminent as well because there are traces of Nigerian government saying things like Nigerian higher education curriculum doesn't fit the international standard of ideal. And there is a need for Nigerian higher education to change their curriculum so Nigerian graduates can be international ready kind of conversation happening. And there is

and instances where Nigerian universities are partner, they are in partnership with different universities here in the UK as well, just coming to share their curriculum and things. So there wasn't really much thing from what I was able to see from Nigerian media saying how they can

kind of contextualise the higher education curriculum to suit the local reality. What I saw in the media, there wasn't much conversation around that. Yeah, I don't even think I was able to see anything around that. It was more about, let's make our student graduate ready. Let's make our education a place where

and we can produce more graduates, we can make sure we're giving them something that make them function effectively on the job. That was what was going on around that policy reform. So that policy reform wasn't really about making the policy responsive to the local reality or local needs. So if that answer your question. Yes, there was an internationalisation conversation and less of about local reality.

DD **Daiana Damasceno** 24:04

Thanks, Hodgkin. Thank you.

YO **Yusuf Olaniyan** 24:06

Yeah, thank you that.

JC **Jonathan Chow** 24:08

Thanks, Youssef. I'll ask a question as well, if I may. I think it was a second last slide where you kind of identified that obviously a lot of the voices that you kind of came across in media were not those of

YO **Yusuf Olaniyan** 24:17

Yeah.

Yeah.

Yeah.

JC **Jonathan Chow** 24:27

the students, perhaps even of a staff. And you know, I assume that kind of points to something more systemic in terms of media and the press. Did you kind of try and did you just kind of perhaps, how did you use that in your research? Did you kind of try and just use that as a critique? Did you try and find these alternative voices from other sources? How did you kind of approach them?

YO **Yusuf Olaniyan** 24:51

Yeah, so because it was just a master's research, so I couldn't do much. So this was just what I had and I did my analysis on these just to share this descriptively, but this is what's going on in Nigerian media. These are the people who have the most voices when it comes to conversations around my education and these are the people who doesn't

my voices. So it was more of me describing what was going on, not really looking for different alternatives to kind of triangulate what I found. Because it was a master's thesis, so I didn't have that luxury of time to dig more.

JC Jonathan Chow 25:27

Gotcha. Thank you. Thank you. All right. I'll probably take one more question that's just come in from Saroop in the chat. So Saroop's question is about use of your purposive sampling strategy. He was wondering how you selected the articles to go into your final corpus. Did you use

YO Yusuf Olaniyan 25:29

Yeah, thank you, Jana.

JC Jonathan Chow 25:48

the quantitative phase to inform this? Did you read all the articles or just headlines? Did you purposefully select a certain kind of number of articles per year or randomly select? How did you do it?

YO Yusuf Olaniyan 26:02

Yeah, so just like I shared, when I started the analysis, I just went on their website. So here, this is the website. Let's say for instance, this is the punch of news media. I just went there, I go to their search button, I just search for higher education. There isn't anything about, they don't have the functions of years or something. So when I search, I just had to be clicking next, next, next, next, next, just to get to 20, 21 that I needed. Then I started clicking on each and every one of them. Then I started reading. So I started reading, okay, if this is about higher education, then I put it that one. Higher education 2, So that was what I was doing in the first phase. So when I had that 1000 article in the first phase, quantitatively counting all of the articles that were talking about higher education, I think I have another column in Excel where I was putting the links to those articles as well. So the links to this article, I was shortlisting them. Then I went in again.

So okay, I have heard all of these articles that were talking about higher education. Then how many of these articles about higher education? What are they actually saying about higher education? It is something to talk about higher education, but

different than to talk about something about higher education. So anything I talk about higher education, I should at least. Then what are you saying about higher education? So what are you saying about higher education is the second?

layout of analysis there. If you're talking about these, so I realised the majority of them are talking about the policy reform and it was about labour market, quality of graduate, labour market. That was very, very, that was very, very recurring in the articles that I shortlisted. So

I made that like, okay, this is, this is some, this is the patterns in this article. So that pattern was what I actually focused on. So that pattern gave me my sampling strategy. So it was the data that spoke to me because when you started checking articles and majority of the articles are talking about the same thing, then that became my sampling. So every article that talks about that same thing.

and I kept shortlisting them. And 31 of them were kind of saying the same thing. Other things, I wasn't really, I wasn't much interested because they weren't saying something about constructing what our education is. Some of them are talking about education buildings. Some of them are talking about a number of buildings Nigerian government need to build.

So about, so it wasn't, it wasn't really much about ideological thinking about higher education. It was more of an infrastructural, it was more about food and, you know, so it wasn't much about what I was looking for. So that was how I did that counting. I hope that helps. I'm Sarah.

But, but the article is actually published, so because of the because of the timing, so it's published, you can read more in the publication, yeah.

 **Jonathan Chow** 28:53

Yeah, that's wonderful. Thank you, Yusuf. I'd be happy to share that after the session as well. Right, okay, let's switch gears a little bit now and look at another piece of research on social media influences and veganism. Up next is our next speaker, Adrianna Jezierska from Business Studies at the University of Bristol.

 **Yusuf Olaniyan** 28:55

Yeah.

 **Adrianna Jezierska** 29:10

Yeah.

Okay.

 **Jonathan Chow** 29:14

Apologise in advance if Adrianna's mic kind of does the static-y sound. Hopefully, we'll be able to get through this without too much problems.

 **Adrianna Jezierska** 29:15

Play.

Call.

No.

 **Jonathan Chow** 29:24

Adrianna, over to you.

 **Adrianna Jezierska** 29:25

Hey, hi everyone. OK, just...

PowerPoint has taken a few seconds to load.

Okay, I think this is working. Oops.

Okay, so hi everyone and thanks Yusuf for the presentation. As I feel like your presentation like really nicely introduced the discourse because I decided to take a slightly different approach and focus really on the technical part aspect of the of the analysis and the methods I use in my Phd.

So I'm currently in my final year of the Phd at the University of Bristol Business School. And broadly speaking, I analyse how social media influencers legitimise veganism over time. And as I said before, I would really focus the methodological approach I developed

to analyse how veganess is legitimized. And the broader aim is to really show you how the computational methods can be combined with this qualitative interpretative discourse analysis that Yusuf introduced very nicely to study the legitimacy construction.

But I promise it won't be very technical, because I'm a social scientist myself, and a lot of these concepts are very scary at first. But I want to show you that even social scientists, they can use these methods with a small support of maybe computational scientists.

because across universities, this support is available. And if you're really interested, we'll be very happy to share further resources of where you can find like summer schools or like more local resources to help you start it. But before diving deep in. to the technologies, a bit of the research background. So we know that the social media platforms have fundamentally reshaped this whole information environment by amplifying the voices of diverse actors, including those social media influences where these really charismatic leaders

They leverage the storytelling and narratives, but trying to be a little bit authentic, but also very often managing the commercial interest. They very often sell products, I think.

I think initially influencers really were associated with purely just marketing, where there were these extensions, extension of specific brands promoting their products. But this has been really shifting recently and influencers themselves create products. And surprisingly, they are more and more increasingly engaged in this, in these like different social issues, discussions on climate change, for example, or the politics. So there have been these recent calls from the management to theorise those influencers.

as those independent actors rather than extensions of these organisations they endorse. And more specifically, so I'm very happy to take this small research gap in there. And in addition, I also use the discursive legitimation.

this discursive legitimacy theory, which emphasises how this legitimacy is constructed for language communication. So really, I hope that this can allow us to move from beyond this influencer as marketing device towards understanding these influences as actors who actively construct

influences in contested domains. And as I mentioned, the context of this study is... slightly alluded to that is the contested terrain, which in my case is veganism. So, you know, some of you might have come across Veganuary. This is the annual social campaign which invites people to go vegan for a month. You know, we have seen a plethora of

vegan recipes that keep popping up on social media. Influencers just chatting about their daily life, whereas having this massive bowls of vegan food, showing their, you know, shopping, grocery shopping trips, recording live from the shop aisles and going through the content

of the ingredients of different products. But at the same time, we have this other side of the spectrum, like the keto carnivore diets, which have, I think, very recently

became even more popular. So there is this constant negotiation of meanings that I think makes vegan is a really interesting case to study how legitimacy is constructed and challenge on life.

However, these examples already illustrate the diversity of the discourse within this vegan influencer content. So there is of course the challenge in identifying recurring patterns across these variations. Here I only showed you this more visual aspect of social media. In

The video formats, we could of course look at the transcripts themselves or as Yusuf also looked at into the articles themselves. However, individual transcripts are very long, they are unstructured, they often contain multiple topics because the videos can last from

5 minutes to over an hour. So if we want to effectively utilise social media data at scale, including hundreds of social media posts, manual analysis might probably be too large for, manual analysis might be challenging, especially if we.

There, if we want to consider the content, the amount of social media posts, but also, like, if we want to, what I try to do is to really focus on this, this temporal analysis, like seeing the seeing the changes in the the discursive trends over time.

So, how can we systematically identify those legitimising discourses and track their evolution over time?

Well, for this reason, I came to learn a little bit about computational approaches, including topic modelling with qualitative interpretation. So in my Phd, I focused on YouTube. I chose it on purpose because it's the second most used platform globally. And it's very much underexplored in research compared to Twitter or Facebook. And in addition, its API is also relatively open compared to Twitter or X, which are very expensive to access and usually you can access for free.

I collected data from vegan-related channels between 2014 and 2024, and this is the temporal aspect, the 10-year period. And initially, I ended up with a data set of 25,000 videos and almost 9,000 unique channels.

So I needed to ensure some filtering steps, such as, you know, I look at some of the engagement metrics, such as comments, the likes. I also manually reviewed a large number, pretty much all of the accounts, because I wanted to ensure that there are no...

businesses, no media, and my data set just to really ensure on this like influencer content. And later to better understand how legitimacy is constructed, I use something that's called bad topic modeling, which helps to identify the clusters of

recurring themes in text.

So this computational stage really identifies the semantic patterns across the data. And later, finally, the qualitative stage interprets those patterns through legitimacy, through the lens of legitimacy theory.

So, but what actually is bad topic modeling? So, topic models have been around for a very long time. I think maybe some of you may have come across a term called ALDA, the Latent Derefit Allocation. I'm then going to do...

I'm not going to go too much in detail here, but it's a like probabilistic model and it uses statistics out to find like Bayesian interference more precisely to find some underlying topics in corpus text. That topic, on the other hand, hand relies on transformers, which are

like this pre-trained models in natural language processing. And they've been trained on large data sets, such as, for example, the Google News data sets. Again, not much technical details here, but I will try to explain the simplified steps that you can see on the slide here.

So just to give you a little bit of understanding what's going on behind this whole process and what defines the actual raw topic output that we will see later on. So as I said, that topic identifies these recurring themes in the semantic way.

And the first step is this is where the transcripts are converted into embeddings. And I will tell you a little bit more about embeddings in the next slide, because for me as a social scientist, this was really helpful to understand what embedding is.

and it made much more sense to understand the output and how it actually works.

Then in the second step, these embeddings are later simplified to preserve these semantic relationships. Texts are then grouped into semantically similar clusters. And finally, we can see that

these clusters are represented as topics, which we can then take into this iterative, qualitative analysis.

So, what are word embeddings? So, by its topic, as I said, analyses the text or in so something that is called string. So, in other words, we can analyse a sentence, for example, in a document. We can also analyse just a chunk of text.

And since that is really like a machine, and machines generally don't like text, they enjoy numbers. So we need to turn this text into the numbers.

Um...

So, so this is what really text embeddings.

the embeddings are. Where the birds, when we feed this document, like let's say the

sentence, like you can see in the example here, I think it's the...

No, these are just words, so we can we can have the cats, kitten, dog, and houses. So, these are the separate words, and as we can see, we they are then turned into into this different different numbers here, list of numbers. So then...

What that does really, it reads these numbers, and the closer the numbers are together, it means that they are semantically similar. So we can see that clearly that cats and kittens, well, they have something in common, whereas the houses and dog, well, not so much. So this is

what embeddings really, really do, and this is what the machine does, like behind the scenes.

overview.

Now, having this idea of what does, how it actually I applied it in this in my research.

So, instead of words or sentences, since these transcripts were very long, I actually decided to work with the chunks of text. I split

each transcript into approximately like 500 characters.

characters, text.

Which, as I said, relates to the fact that...

You know, each of the video might have various representation of the topics, but also it's one of the technical technical flaw of that, which doesn't allow you to have long enough texts, because simply...

after these 500, approximately 500 characters, it will basically chunk crop the text and will only analyze, so do the embedding, so the semantic analysis of how closely the words or sentences are related. We'll only do this for the first part of the text. So I first needed to chunk the documents.

Then second, I created the classes and the classes determines the smallest number of text chunks that required the topic model to recognise a thematic pattern. So I set this to two, meaning that at least two semantically similar chunks were required for a topic to be formed.

As a result, I received 220 topics, from which I serve as the vegan discourses over time. And then I followed with qualitative analysis of these topics and the

representative videos. And the final two steps is really the iterative process. I kept moving between the raw output, which I will show you in just a second.

and the actual interpretation of the topic after watching several videos, reading some of the comments, reading some of the influences proper, just trying to better contextualise what the topic actually represents.

So this slide really illustrates this raw output of a topic. So here you can see like a sample topic identified by Beth.

Um...

So, for example, here in the topic 70, we can see keywords such as the methane, the carbon emissions, carbon dioxide, atmosphere, cycle. So, which we can interpret loosely as debating

some kind of environmental impact of the veganism. And this is what I did for all of the 120 topics. Initially, I just had some rough ideas about what they discussed. In the second, once this has been completed,

I tried to name all of the topics together. Like, so as I said, in the topic 70 example, this could relate to some debates about environmental impact of vegan, vegan diet and of farming more generally. Before,

I moved to the 2nd order code. I went back to this raw output, but also the videos themselves. And I started watching the most representative videos that appeared, or the video chunks that appeared within this topic. I started reading about

as I said, about the influences, who they are. And this really started forming like the 2nd order coding, which I came to understand as positioning veganism as a solution to climate change. So rather than working with these

topics, this raw topics, they both started like, you know, grouping, I started grouping them together. And then the aggregated dimensions, like I mentioned, I used the legitimacy theory. Well, these are very much like moralistic claims. So this is where these aggregated dimension and the final output.

Of the of the topic analysis.

But I also mentioned that, oh, didn't you just jank your text? Like, did you analyse like the full video or like what happened to that? Well,

A little bit, yes. So...

So what really happens with the chunking is, as I said, it's a technical flaw of BERT. It doesn't let you input very long documents. It's likes when it's when it's short. So I needed to cut these videos down. But when I was reading about it, I actually realised that, oh, actually, you know, in the case of the videos,

They're usually very long. They are probably looking at different topics at the same time. So like in this example here, this was a video posted by one of the influencers that appears very often in the data, and there were two chunks that I identified with two different topics.

And each of these topics, each of this chunk receives a separate topic. So we can see

that the chunk one was topic 5, which was the integrating veganness into family life. And the probability probability of this topic appearing of this of this chunk was just 25%, whereas the the the chunk to the topic then protecting sorry, the screen is very small. I only have one screen.

promoting sustainable cooking and shopping and shopping practices, which was representing 75% of that video. So having this understanding of how these different topics come together is very nice approach to then fit further back

Into this raw output, because I can counting, you can focus on the occurrence of these topics together. Are these, like, from this example, we can understand that this influencer really focuses on this family cooking and discussing these ideas of...

you know, what happens to the family settings when one of your partners or you yourself become vegan, like what happens to other family members.

And in fact, this video was titled something alongside the lines of turning your family vegan. So this can tell you a little bit more as well, like this junking, although it's, I take it as a technical flow, but in the case of these videos, a long text.

It can actually help you further understand what topics and how they occur together next to each other.

Okay, so that was the basic modeling, but what happens to the temporal, this approach? Well, bad topic modelling allows you to also track changes in the top words represented by the topic. So remember a few slides, we look at this raw outputs.

Of a topic which was at any given point. However, since when I was collecting the data, I also had the metadata including the date when the video was posted, and this allows me to track the evolution of the keywords.

In this case here, we can see that this discourse transitions from this scientific and biological rationalization. We can see like the references to Catherine Milton, who in 2014 or, and she's an anthropologist of brain evolution in 2015.

So there's really like this, like, as I said, like this bio, more biological, like scientific, rational compared to, for example, hunter-gatherers, I think started appearing later on.

Um...

Country gatherers, like, for example, in 2021, the tribes, plants, and animal food, plants, animal foods.

And when you get to see the full data, it actually started reflecting this ancestral eating trend that has been covered by the influences. However, looking at the

broader media and what has been happening on socials, it can be also the fact that, like, as I said, keto diets have been recently on the rise, so this is really reflected in this like temporal analysis. You can track these changes, but this is again very, very niche and very...

detailed analysis. So what I'm actually doing and looking at as well in my work is really looking at this like broad mechanism that I identified. There is 8 or 9 in total and really tracking what happens to them over time. So just to briefly give you an here, like we can see this dominant legitimising discourse, which is making the everyday vegan practise accessible. And we can see that it peaked around 2017, 2017, and then we saw a big, big dip in the amount of videos representing this mechanism around 2022. Yet it remains the most frequent topic. So here we can draw some, you know, insights on whether, why is this happening? So we might assume that, oh, maybe public is most interested in how to be vegan.

rather than why, because we see that on the other hand, the topics related to the moral legitimacy, which is the light pink, it's the occurrence, the amount of videos appearing over time in these discourses, well, it's relatively low compared to them. Pet pink one.

Okay, so just to wrap up, so I think that applying this bad topic to long form YouTube transcript post or long text requires several methodological design decisions.

However, there's much more to explore and just wanted to wrap up with like a two order that I tested myself and were also quite nice for data exploration.

but also considered as a good practise in topic modelling generally. So we already touched on chunking, so like the splitting of the text, because bad like this machine likes when it's rather short. The second is something, there's the text preprocessing, and this can balance the output that you might have.

So pre-processing usually involves deleting the dots or commas. And it's a good practice, but the documentation does not mention it as a required step. And in my own research, I did not see substantial changes. However, I have friends who also apply these methods.

And they, they, they say that they can they can see several differences, and finally, topic outputs require this interactive interpretation, since, and since that topic is, I think that a lot of you might ask, like, oh, but how do you know how many topics? actually a good representation of the corpus of the data set that you have. Well, there is no right answer. And that topic, unfortunately, doesn't let you predefine the

number of topic. And then they let you decide the cluster. So then you can say that I want at least two documents

Similar documents to be put together to get this thematic output.

So given that you cannot really define this number of topics, and there is also like no appropriate measure for that topic, there is something that is called, oh no, it disappeared. I'm sorry, I didn't cheque this one, but something that's called like really like a hierarchical output.

And since you cannot see the, since you cannot see the visual, I won't be describing it in detail because it makes much more sense when it's visually, but it really allows you then to combine some of the very, very semantically similar topics. I will make sure I will add it in the final slides after the session.

So, well, I thank you very much. I hope it was useful. Not too technical, but yes, if you have any questions, let me know. Thank you.

 **Jonathan Chow** 56:04

Wow, thank you, Adrianna. I mean, I'm not even going to try and hide it. Like, I'm not, I'm not trained in this way at all. So, you know, but it was fascinating to see just how much you can get out of the model, like just the incredible ability to work through so much data. Thankfully, there is a question.

 **Adrianna Jezierska** 56:19

Thank you.

 **Jonathan Chow** 56:25

from Jezierska. Hi again, Jezierska. Thanks for joining us. I now unmuted you so that you can ask your question.

 **Adrianna Jezierska** 56:25

No.

 **Jonathan Chow** 56:37

Oh, sorry, sorry, Jezierska. I've enabled you to unmute yourself, rather.

Was that a false hand? Jezierska, we cannot hear you yet.

Okay, we might try and come back to Jezierska. We still, on my end, you still look like

you've got your mic muted. If there are obviously any other questions, please do feel free to raise your hand.

 **Adrianna Jezierska** 57:06

Oh
Yeah.

 **Angeline Mbogo Barrett** 57:07

Yes.

 **Adrianna Jezierska** 57:09

Okay.

 **Angeline Mbogo Barrett** 57:09

Question in the Q&A from Lee Poetzsch talk, because we're running out of time, Lee. I hope you're happy for me to ask your question. And first of all, he says thank you to both Youssef and Adrianna. I'll echo that. Thank you both very much for your presentations. The questions about ethics and data handling.

 **Adrianna Jezierska** 57:11

Yeah.

 **Jonathan Chow** 57:13

Ohh, okay, perfect.

 **Adrianna Jezierska** 57:16

The.
YouTube.
Call Cortana.

 **Angeline Mbogo Barrett** 57:31

Could you share how you stored your data and how long you stored it for? Was that an issue for this data, Adrianna? Because I hope it was already, for both of you and you said, it was already in the public domain, right?



Adrianna Jezierska 57:43

Yes, yes. So generally speaking, the social media data collection, although it may seem that we might not require ethical ethics approval, we do because there is still humans involved. It's A slightly, there is a very useful

There are a lot of guides online, especially from the Association of Internet Researchers, and they keep updating this, which is a very useful tool for approaching the task. However, what we need to keep in mind for the ethical approval is the fact that, although it's a public data, these people put it out.

There, so, but still, these are these are humans, so we need to need to remain considerate. What I, for example, see very often, even if we if we keep screenshots of the data, we, for example, blur the faces of the people.

I hope that helps. And in terms of data storage, so I use the computational resources from Bristol Uni because it's a large amount of data. So my laptop just cannot handle it. I thought it would, but hence the noise that you might probably hear in the background.

So, I think it broke it slightly, so if you have most of the universities will have some computational resources to help you store this large amount of data.



Angeline Mbogo Barrett 59:11

That's a very good point. Thank you, Adrianna. And the resource you mentioned, just to say, I've popped that in the chat. Association of Internet Researchers. Thank you.



Adrianna Jezierska 59:11

How do you?



Jonathan Chow 59:25

Can I ask Adrianna about kind of how you, what the learning process was like for you? Did you already come into this with a little bit of prior knowledge or is it brand new to you? And if so, how did you kind of, how did you figure, work your way through?



Adrianna Jezierska 59:30

Okay.

No.

Okay.

No.

Hello. So I really started doing some computational social science, how it's nicely called. In the first year of my Phd, I started broadly with a summer school, which was a very, very nice way to get into

Into the field, and you also get to meet people who are a bit more experienced with coding, so you get to learn a lot from them, but then it's really finding the materials and the resources online. There is a lot of them, and there is also a lot of like step-by-step guides, and...

Also, I also decided specifically to work with Birds because the documentation online is very detailed, and for someone who doesn't necessarily know how to code, doesn't have a wealth of experience, I managed to really follow the documents with just a little help of computational of.

engineers or computer scientists. Like in Bris, if you're based in Bristol, there's a great resource, the Jin Golding Institute. The guys are great in great and they work with these type of methods. So so so it was really a lot of.

Yeah, a lot of exploration, I would say, I would say, and finding the right people alongside. But for best specifically this method, I would say that the documentation also made it rather straightforward and understandable for someone who doesn't have a lot of exposure to this type of methodology.

 **Jonathan Chow** 1:01:21

Okay, well, that's encouraging. I hope that if you in the audience were inspired by that and didn't know where to start, hopefully that will get you started a little bit. We are out of time now, unfortunately. I just want to quickly point to a link that I've put in the Q&A to a survey.

 **Adrianna Jezierska** 1:01:31

Okay.

A lot.

The.

 **Jonathan Chow** 1:01:41

that we like to do. It just really is very helpful in helping us make sure that any future training we put on is beneficial to you. So we'd love to hear from you. And also

there's a link there to sign up to any of the remaining sessions coming up in the series, including one on critical discourse analysis.

 **Adrianna Jezierska** 1:01:44

Okay.

Laura.

Bing, Bing.

 **Jonathan Chow** 1:02:01

on the, I believe, the 14th of May.

Sorry, it's getting quite loud now, I think, Adrianna, where you are. Yeah, so, sorry, I do see a hand. What I will do is, if you do have to go, because we are out of time, please do feel free to leave. Thank you again for joining us. But if Adrianna and Yusuf, if you could stay behind for a minute, and see if the question pertains to either of you. We can take that one as well if that's okay. So thank you to the rest of you for joining us and then I will now allow Santiago to unmute.

 **Angeline Mbogo Barrett** 1:02:50

Santi, you'll have to unmute your mic.

 **Jonathan Chow** 1:02:51

Student.

Could be a force, perhaps.

Oh, there we go. Hi, yes, we can hear you.

 **Santiago Ayuso Arcas** 1:02:59

Hello, can you hear me?

 **Angeline Mbogo Barrett** 1:03:00

Yes, I can hear you.

 **Santiago Ayuso Arcas** 1:03:01

Cool. Yeah. Well, first of all, thanks very much for organising the talk. It's been great. I'm using critical discourse analysis and I could see many resemblances with the

different strategies and
and concepts coming out during the session. So thanks very much for doing that.
And Adrianna and the rest of the presenters, amazing job. And it was very quick. My
question, is it going to be available this, because I can see is recorded, so is it going
to be available this session?
to, because I found so many interesting bits that I'd like to go through again. So I
was wondering if I could access again the presentation, the different presentations,
because yeah, I'm
As I said, I'm working with CDA and I didn't come across many of the things that I'd
like to explore and see if I can draw on as well. So yeah, that was all. Thank you.

 **Jonathan Chow** 1:04:18

Okay, yeah, yeah, no, so it will be shared. It won't be available right away. You won't
be able to access it through the Teams. Sometimes it pops up for you automatically.
You won't be able to access it that way, but we will, once it's ready, we will send a
link to you and all the people who have registered.

 **Santiago Ayuso Arcas** 1:04:36

Okay, thanks very much.

 **Jonathan Chow** 1:04:38

Yeah, no worries. Right, okay. In that case, yeah, thanks for thanks to the rest of you
and thank you again, Adrianna and Yusuf. Fantastic presentations.

 **Adrianna Jezierska** 1:04:51

Thanks for hosting us, and...

 **Yusuf Olaniyan** 1:04:51

Thank you, Jonathan.

 **Jonathan Chow** 1:04:53

Yeah, cool. Stay in touch then.

 **Yusuf Olaniyan** 1:04:53

Yeah.

stopped transcription