# **Introduction to Topic 10**

Welcome to topic 10 in which we discussed natural language

processing in practice. To get us started, let's revisit our course

learning objectives. For this topic, it's slightly different

because unlike all the other topics where we basically discuss

scholarly material, the theory, and the

practice of NLP, here we're focusing

exclusively on the latter. We're focusing on the practice. I'll tell you a little bit

more about that in a moment. But just to revisit our

course learning objectives, there is two that is

appropriate for this topic. We're going to be thinking about appropriate language

analysis techniques for particular

problems because what we're going to be discussing in this topic is very

problem-centric. We're also going to be

thinking about the tooling that you might use because

the whole point of this topic is about NLP and practice so

we're thinking about practical tooling

and techniques. We'll think a little bit about course learning

objective number 3. As I mentioned, this topic

is different to the others, whereas the others was

essentially me imparting my skills and experience

of our many years, decades actually working in NLP and conveying that to you, for this topic, I'm pleased to say that I'm

joined by various experts from the community who have lent their time in the

form of interviews. They're giving their own

unique insight that the presidents a combined total

of many decades in NLP. They have lots of

interesting experiences to convey to you in this topic. That really segways into the

topic learning objectives. You'll see that

there's three here, we'll talk a lot about NLP

skills and competencies. If it's a career that

you want to pursue, and I sincerely hope it is, it's good to go in

with some knowledge of what skills and competence

would be expected of you, where you need to be

hired into a role as an NLP specialist. We'll also talk a

little bit about natural language engineering. One of the things

that's come out across these various discussions

with the interviewees is the importance of basic

engineering skills and to be able to

contextualize what you do, NLP techniques

that you do within the broader contexts in software engineer

and data science. I'm going to also talk

a little bit about NLP trends and developments. I took the opportunity to

question our interviewees on a number of topics including what's hot now and what

might be hot in the future. It'll be interesting to revisit those interviews

perhaps in a couple of years' time and see how many of those predictions have

actually come true. There it is, that topic

10. Let's get started.

# **Phil Gooch interview**

I have with me today a special guest who I have

known for many years, done some amazing stuff

in the NLP community. I'm delighted to welcome, but I'll let him

introduce himself. Could you tell us

a little bit about who you are and what you do? Thanks, Tony. My

name is Phil Gooch, and I'm the co-founder of a

startup called Scholarcy. Is that it? Well, my role is co-founder, but we're a small startup, so I'm also CTO and main

NLP person as well. Yeah. Fantastic. Great. Can you tell us a little bit more about

what Scholarcy does then? I'm guessing that for

our audience to this, who are mainly students, but possibly a few researchers, this is going to be

right up their street. Great. Well, Scholarcy

is an ed-tech startup. Put simply, our technology

distills research materials, or learning materials

in general into bite-size knowledge by turning that into what we call

summary flashcards. It's used by students and

researchers around the world. The idea is just to distill information into

the key findings, really the key takeaways, and the key information

that you need to learn what's in the documents

that you put through it. Yeah. I think you're describing

it very modestly Phil. As someone I've used it myself, I know it's a great tool. Let's talk a little bit about your experiences in

NLP then in practice. You and I have both been

in an NLP quite awhile. I think we all have

our go-to resources. I've got a few on my

bookshelf behind me. What would you say is

your favorite book or blogger resource on NLP? This can be a contemporary thing or it can be a historical thing. It's hard to find one

single favorite thing. Maybe if I mentioned a

few things that I like. I mean, on the NLP side, Seb Ruder who works

for Deep Mind, I think, or Google Brain,

I'm not sure which one now. He has a really

fantastic blog on the latest developments

in state-of-the-art NLP. He's really worth following. He updates his blog quite regularly, he does a newsletter. He was one of the

people who really pioneered this idea

of transfer learning, the natural language processing, which previously had been more of the domain

of image processing where you train a model on an image domain and then

transfer it to another one. He really pioneered the

idea that you could do transfer learning on text. That all lead to

the whole bot and transformers thing that

we see today basically. He's really influential, done

some great work and he has an excellent blog that he keeps up-to-date with the

latest trends in NLP. That's definitely worth reading. In terms of books, I've to say that I find that data science books generally are really useful rather

than NLP specifics. Couple of my favorite

books which I really recommend anybody

who's interested in NLP get familiar

with is Data Science at the Command Line by somebody

called Jeroen Janssens. All that does is

it tells you about the key tools that you can

use to get pipelines and processing and integrations

together to do NLP, but also other data

science tasks as well. It got some great

examples in there. That's a really excellent

book if you want to actually apply some

NLP that you're learning to building

your own tools and building your own solution. To be honest, it

was really helpful when I was starting

building Scholarcy. I had theoretical knowledge and I wanted to learn how to put all this stuff together.

That's an excellent book. Another one also is really good, it's along similar

lines is Data Science from Scratch by Joel Grus. They're really good books

that are worth learning. On the NLP side, obviously, there's the classic speech

and language therapy, which I think is in its third, fourth edition

now, I'm not sure. There's a couple

of books on spaCy. SpaCy is an excellent NLP

library from [inaudible]. They've got a book I think

called Mastering spaCy, which is worth reading and

there's huge amount of online resources

on their website including a free online course. I recommend checking all

those out if you can. Great. Excellent.

Thank you, Phil. Yeah, it just so

happens actually, Jurafsky and Martin

book is one of the main ones that we

use on this course. Yeah, it's a huge text, but it's the go-to

book, isn't it really? Let's talk a little

bit more about your experiences in NLP then. Again, you can think

about this from academic point of

view or for well, actually a practitioner

point of view I think would probably be of greater interest for

this particular session. But, you tell us

about an NLP project that you felt was a real

learning experience where it turned a corner or a key milestone in

your understanding of the field of NLP? I think one of the

best example was when I moved from academia to the

commercial world really. I mean, I had worked previously

before I did my PhD, I was working for

companies but not in an NLP role and in more of

an analyst role really. After I did my PhD, after a couple of years as a research developer

at King's College, I got a job with startup, as it was back then

called Mendeley. I was their third

data scientist. They only had two data

scientists at the time, and I was the third

one to join a team, and I had not done any

data science before. I had built some NLP

pipelines before in GATE, which is not a great

NLP tool that, hopefully, your students

will be familiar with, but I hadn't really done

any production work in terms of actually building stuff that was actually

going to run on its own as a part of a

service and at scale. That was a real learning

experience because my task there initially

at Mendeley was to basically figure out how to extract metadata

from PDF documents. You automate the process of importing documents into

your Mendeley library, and extract all the information like the author, the

title, the abstract. That was interesting

task, and I was using some open source tools

to help with that, but it was all the

engineering around it I had to learn about, and I was working with

a really great team of engineers who really helped

us productionize that, so I was still working in

a slightly R&D bubble, but with really great engineers

to help productionize, and so through that, I learned how all this stuff works, the whole process of you do your initial

experimentation, you build your models,

you evaluate them, you iterate on that, and then you've actually got

to productionize it. I was working with the

bigger engineering team, which was new to me, and that was a really

great experience. I did learn a lot

from that in terms of that whole

end-to-end process. Great. Thank you. You

mentioned earlier, just before we actually

started recording, about Scholarcy expanding and being able to support a greater number of stuff

than, say, a year ago. If you were hiring

NLP specialists, perhaps a graduate MSC student, what kind of skills

or competencies do you think you would look for? I think what's

important, and again, coming back to my

experience at Mendeley, this is something that was quite important

there as well, and particularly for startup, where people have to wear

multiple hats really, particularly in the early days, somebody who's a self-starter and can prototype a

solution independently. We would look for somebody who can go out and find datasets, clean them up, prepare them, and process them and build

some models and train them. They don't actually

have to build a production-level system, but somebody who can actually do the data engineering

part, really. I think that's really important, and maybe an overlook part of NLP in terms of the data

preparation and cleaning. You're not always going

to have clean data, so if you've got some

experience of doing that, or just general programming experience is really helpful, but really, I think most importantly, someone

who's curious. If you are curious about data science or

you're curious about engineering or machine learning and you want to learn that,

that's really important. I think, also, look

for someone who's maybe done a little bit

of open source work. Maybe they've contributed

to some repositories on GitHub or maybe got some code in the public

domain that we can try out. I think it's really good to work on some projects, and also, a really great way of improving your general skills

as a practitioner is if you can help out work with other people who are building

open-source software. Any of that experience is

really helpful as well, but I think the

engineering side or just knowing enough to clean up data and process it and build a model from that is really useful. Great. Let's turn to the NLP as a field or

profession in its own right. I know that it's undergone quite a significant

transformation particularly over the last five, perhaps even 10 years, and the leaderboards are

constantly changing with new models and new variants

of models and so on. In a moment, I'll ask

you about the future, but what's your take on

what's hot right now in NLP? What do you think is the flavor of the month or the zeitgeist? Transformers are still hot, and they have been for

a couple of years, and they're still the

dominant paradigm, but what's interesting about

transformers is they use this thing called

subword tokenization. In traditional NLP, you'll

tokenize on whitespace, if it's a romance

language typically, or you'll find some

other method of splitting sentences into words. With transformers, you use something called

subword tokenization, which is basically statistically based to slice up words not necessarily on

word boundaries, so it's parts of words, and it's not really exactly morphemes or anything like that from a linguistic

point of view, but just commonly occurring sequences of characters

that tend to be split up. Now, one of the most recent

advances on that, I think, is something called car former from a team at Google, and they were

looking at, actually, can you learn this

subword tokenization to input the stuff into the transformer and do it

across multiple languages? They were able to

successfully build some models that worked across, I can't remember how

many languages now, but it could learn to do

this subword tokenization, and they evaluated it against some other

methods than this- Effectively rules-based

subword tokenization that was currently being used and character-based

tokenization and it was superior to that. That's quite an

interesting advance. My only concern about

that is it very much tie into the whole

transformers paradigm. I wonder if, in future, whether character-based models

will ultimately catch up. Because that would

be the ideal thing, it's not really

tough to worry about words and just

really think about streams of characters or

streams of bytes, really. That brings me on to other stuff that I think is

quite hot at the moment, which is the joint learning

of image in language data. Again, if you're learning

from streams of bytes, does it really matter

whether these bytes are representing texts are

representing images? And can you combine

those models so you jointly learn both the

image and its caption, or the image and

its description, or summary of the image, or scene, or whatever. So that's what's interesting

areas at the moment. Okay. Great. This is

always the tough one, the predictions for the future. I often wonder what, if I'd asked this

question five-years ago, who would have

predicted transformers? No, I wouldn't. Not many. But as

to 10 years ago, who would have predicted

the embedding, and dense factors, and all the stuff that

we take for granted now? Probably not that many again. If you were a betting

man, I'm sure you're not. But if you were,

what would you put your money on or

what do you think the future might go for

the next five, 10 years? Well, I'm going to go

out on a limb here and talk about quantum computing because I think that's

a really exciting area. I think quantum machine

learning and quantum NLP could be a game changer. If the current developments in quantum computing and the

number of available cubits, which are necessary to basically run constant based

algorithms on this hardware, if this continues, then I think we'll see some really interesting

developments in getting the

performance that we can get with these really

large language models in a much smaller

language models. Also, actually having

some progress on real natural language

understanding with language models

that are grammar aware. Because the current

language models aren't really grammar aware, they're character aware, they're word aware,

they're context-aware, but they don't really

encode grammar rules. I think with quantum

deep learning, I think there's a

possibility that the prospect of

actually encoding some of this information. I'm not by any means an

expert on quantum computing, I'm a complete beginner on it, but from what I've read,

it's really exciting. There's some initial

work being done. Well, it seems the MIT and on others and on

actually implementing some NLP algorithms on quantum hardware as

is at the moment. So I think that's a really exciting area

and I think that's going to be quite big

in the next 5-10 years. Right. All right. Well,

you heard it here first. Good. Thank you, Phil. I'll play this back to you

probably this time we'll see what's happening. All right. Just one final question

to wrap us up, and this is a sort of a bit of a slightly tongue-in-cheek

question, I guess. So if you could change one

thing about the profession or even your role within

it, what would it be? What would I change

about profession? I mean, we all very much

in machine learning, probabilistic deep

learning paradigm for NLP, and that's great. I mean, there's been

some huge advances, but I think people

maybe have forgotten some early work that's

been done on pausing, and grammars, and so on. That's really important

to learn because you can actually achieve a lot

without doing deep learning. You can actually achieve

a lot with rules. I'm not saying that you'd

want to productionize that, but it's really important to know the basics of

NLP and how to put stuff together and build a simple rules-based system

from which you could use to generate

training data, at least. So then you would

then correct and then maybe eventually

feed into a model. I'm a big fan of all

the new developments, but that's not the whole story. I think it's really important to go back to basic as well, and I think some people

might have forgotten that. I think it's

important to remember there's some great

work that was done 10, 20, 30 years ago. They're still

relevant, I mean stuff like [inaudible] for example, are still very useful

for doing things like hypernym

detection and so on. There's great stuff out there that will get

you good results. So don't forget about all the stuff that

was done years ago. Okay. Yeah. Excellent. Great. Well,

thanks very much, Phil. We covered a lot

of ground there. Yeah, some very interesting insights so I'm really grateful. Thank you very much for that and all the best for Scholarcy. Thanks very much, Tony. Cheers.

# **Diana Maynard interview**

Great, so I'm very pleased to say we've

got a special guest with us today. So to kick us off, can you tell

us who you are and what you do? >> My name is Dr. Diana Maynard. I am a senior research fellow

at the University of Sheffield, where I work in the natural

language processing group, which is part of the department

of computer science, where I've been for the last 21 years. >> You must really like it there. >> [LAUGH] Yes. >> You've got some lovely things, lovely places on your doorstep as

part of the world I'm very fond of. And a department that I have a certain

affinity for, so I don't blame you. 21 is good innings,

you get a gold watch or anything? >> [LAUGH] I wish,

I might get something at 50 years. >> [LAUGH]

>> If [INAUDIBLE] have that long. >> Well, I'll do another interview

in 29 years then and we'll see. Alright, great. So can you tell us a little bit about what

your company does, I know in your case is university, would you tell us

a little bit about the sort of stuff. I mean over 21 years,

I'm sure the roles evolved a bit as well. >> Sure, yeah. Well, as I said, I work in

a natural language processing group which is quite a large group within

the computer science department, who do all kinds of things related to NLP. And within that we have

a specialized team of around 12, the number changes periodically,

plus a few PhD students. Well, as part of the GATE team,

which all works with the, in some form rather with the GATE toolkit, which is a tool kit we've been

developing for the last 21 years for doing all kinds of natural

language processing tasks. So within that I help to manage

the team of researchers. We work on all kinds of

different NLP tasks. So we have a number of projects,

several European, big European projects. We also do some consultants there for

various companies. We do training as well. So we train people to use the GATE

toolkit, which I said we've been developing, which is all open source and

freely available. But sometimes we train

people to use the tools and do specific consultancy

work with them to help them develop their own tools and NLP projects. So, my role is a bit of

a mixture of team management, I do some NLP research myself,

and a lot of outreach. So part of what we do is taking NLP and

our tools and projects to the wider world, and

obviously acquiring funding as well. And we work in all all kinds

of domains with NLP tools. >> Great, okay, thanks Diana. It'd be interesting actually because you

mentioned GATE and I'm a great fan of GATE, and it was the first NLP

toolkit that I ever seriously learned. And suppose for a while, for the first 10

years of maybe you've been at Sheffield, there wasn't that much other tool

kits out there, probably maybe MLCK. But obviously in the last few years

there's been the hugging face stuff, all the various google flavors,

all the various OpenNLP, Stanford NLP, CoreNLP,

any thoughts on those? I mean, I know this platform differences

and so on, but any thoughts on, do you ever get to use them or

do a bake off or anything like that? >> Yes, I mean the whole

philosophy of the GATE toolkit, I mean it was really the first

toolkit of its kind, so it came before really the Stanford

tools or NLT Care, or anything else. And the whole aim was just

basically to make it easier for people to put together different

NLP components in an easy way, rather than having to adapt everything

to work with everything else. So, a whole philosophy is

never based around any kind of notion that GATE is

the best NLP tools as such. So we take very much the philosophy

that you might want to use different tools for

different tasks, and we try and make it as compatible with

other tools as we can. So we've actually integrated GATE

with a number of other tools. So for example you can use Stanford tools, you can use OpenNLP,

some of the IBM tools, various other tools you can use

within GATE, they're already in. Some are already integrated, and

other ones you can integrate yourself. And we've recently incorporated

a whole Python line as well, which means you can now integrate with

Python tools as well [INAUDIBLE]. So, as I said, yes,

I mean every tool has advantages and disadvantages depending

on what you want to do. The nice thing about GATE is

that it creates a very nice environment to do stuff visually,

especially for people who don't necessarily

have computing skills. So it has a nice visual environment. And even though it does have a bit of a

learning curve, it can very easily be used by, for example, humanities researchers

who don't want to know what goes on under the hood, whereas for

them even just using some Java tools or Python tools might be

a little bit beyond them. So it's quite nice in that respect. But it also provides a sort of framework

where you can test out a whole bunch of different tools. So for example you can take a bunch

of different or whatever and try them all within GATE and

see which one works best and which ne connects best with other

components you might have in your toolkit. >> Yeah, excellent, thanks. So you know I've around the NLP community

a little while, and over that period we've all developed our own favorite

go to resources for various things, so what would you say is your favorite book,

or blog, or resource for NLP? >> So,

something that I've come across relatively recently in the last I guess year or so, is a weekly newsletter that's

produced by Bob Dale from the Language Technology Group,

it's called This Week in NLP. And basically it highlights

all the latest news and trends in the world of commercial NLP. And it's something I find

really interesting, because for researchers, these kind of news,

and trends, and so on and tools can easily bypassed you as

an academic researcher, unless you spend a lot of time following things

on twitter and news sites and so on. It's very short, it's just to kind of

a quick digest where it highlights, That's the most interesting and

relevant of NLP themes that the author seen in the last week, and then gives

you links to more sources and so on. So it's really nice,

it comes on a Friday afternoon and it's a nice way to just keep track

of what's going on in the world and to kind of keep that link

with the commercial world. Which is often quite important and

said and something that sometimes, as an academic research you get

a little bit separated from. >> Yeah, absolutely, it's becoming

increasingly important, isn't it? I remember the days when NLP was

perhaps less than 100 people in the UK, that was the community and

it was all researchers. And of course now it's probably

number of researchers hasn't changed a number of practitioners is

going to significantly, isn't it? >> Yeah.

>> So can you tell us about an NLP project

that you felt was like a real milestone, I guess in your career or real

learning experience where you thought, yeah, I've got this, this is

something I really want to pursue or this is something that I

think I can do as a career. >> Well maybe not so

much a project as such, but if I can go a little bit beyond this,

but I think the whole world of

social media analysis and sentiment analysis, there was a point, I guess around ten years ago or so. Where social media was becoming

more popular, it was gradually, people were starting to use things

like Facebook and Twitter and so on. And I started working on

sentiment analysis and we started on a project just developing sentiment analysis around things like, well first of all things like reviews,

online reviews and then around political opinions and so on. And suddenly it was I think

NLP it's really hard to explain to other people and

it's something I always used to feel a little bit devolved

from not reality as such, but a very obvious use that

everyone could see in this. And when we started working on

sentiment analysis I thought, and there was very little being

done on that at the time. And I thought this is

really going to take off, this is going to be one of the things

that is everyone is going to need to use. And so

we started working on this topic and I thought this is an area that

we really want to get into. So I really pushed the development of

at the time, quite simple tools for doing this kind of thing with

simple machine learning and so on. And it then led to

a whole sort of strand of work that we've now developed and is one of our core areas of research. So looking at all kinds of sentiment,

hate speech and it also took us from a research here which was very much about just NLP. So basically trying to extract information

in different ways from documents and make sense of it to something that

became much more multidisciplinary. Because as soon as you start looking at

things like social media analysis and sentiment analysis, you start getting

into the whole sort of psychology of why do people say the things they do? How do people respond to it, what sort

of person says this kind of thing? What's the relationship between

how people act on social media and how they act in real life and so on. And it suddenly takes you into

a whole world of psychology, of social science,

of all kinds of other topics. And for me that not only is it really,

I think it's really interesting, I find it absolutely fascinating

trying to kind of unravel all this. But also it has very

important consequences and it opens up all kinds of doors for

what you can do with it. As well as the number of people that can

make use of these kind of technologies. So that was a sort of real I

guess I hope no instant sense, it was a door into a new world

[LAUGH] that makes sense. >> Yeah,

totally I mean I would agree with you. I think the sentiment analysis was

in some ways that kind of not quite the killer app, but it was one that I

think brought a lot of startups were founded around sentiment analysis. Well and it was a niche in its own right,

startups that didn't necessarily say, we're an NLP company and we've got

NLP school's watch, we apply them to, it's kind of almost the other way around. Sentiment was recognized as a problem and

companies started to address that problem that recognize that

NLP was the technology they needed. >> [CROSSTALK]

>> I think it's something that sort of brought NLP into the-

>> Yeah. >> Into a world where

it hadn't been before. >> Yeah, exactly, so you mentioned that you manage a team

of researchers up at Sheffield. So, what kind of skills or

competencies do you look for when you're looking to hire

another NLP specialist? >> So this is quite interesting and

complex question I think because, I mean most people don't

come directly out of university having done an NLP degree for

example, which I actually did, but

I'm quite rare in that sense. So NLP traditionally has attracted people

from a variety of different backgrounds. So in my group we have people with more

linguistics backgrounds such as myself, people have core computer science

backgrounds, people from maths backgrounds and then also people from

physics and other kinds of backgrounds. But things are changing quite a lot,

especially in the last few years and especially with the rise of machine

learning technologies and so on. So NLP now is becoming,

it requires more and more of a strong maths background

above all else I think. >> Yeah. >> The competent, but

it's a sort of high level, if we're looking at researchers

rather than PhD students, they need a good competency and

typically a machine learning. They don't necessarily

need all these things, but good competency in machine learning,

deep learning as I do, but also a little bit of

background into NLP is good. I mean it can be learned, especially if

you come from machine learning background where often people have sort

of dabbled in NLP tasks. But I think more importantly,

a kind of core skills and interests, so I always look for someone who's

actually got an interest in language. And this, as I said, with the way that

NLP is going at the moment is becoming more and more multidisciplinary and

this means that you need a kind of. Like I say an interest in and

the curiosity and not just being able to write

an algorithm to do whatever. But also to think a bit wider

to think a bit of a little bit more outside the box and

to think about these things like the sort of social science aspect,

the psychology aspect and so on. And to be able to sort of go beyond just a straightforward

programming kind of skills so that you have this ability to

go down different avenues. So this way I guess this means

a certain bit of a bit of curiosity and and adaptability to because the field

also it's changing very very rapidly. So people really need to be

quite adaptable in that sense to be able to learn new things. And yes new tools new

technologies all the time. >> Which actually leads me neatly on

in a moment I'm going to ask you about the future of NLP. But just before we do that, what do you

think is hot at the moment in an NLP. Well it's kind of flavor them on. >> Deep learning gives you, it gives you

it has a lot of power right to do things that just were previously impossible. But the flip side of that is that

we're getting into a sort of dangerous area of black box NLP where no

one's got any idea what's going on under the surface and

actually that's still really important. So the ideas around the explain ability

of a of AI of learning models and NLP is becoming more and more important? So I think that's one area where there's

a lot of research going into developing tools that are also explainable. I would say there's there's a couple

of domains where NLP is particularly coming to the fore, one is the biomedical

fields and that's been progressing for a few years now, but again with

the tools that are now available. This really has a massive potential to

actually change the whole field of modern medicine and really make new discoveries,

but also I think something we don't actually work on ourselves particularly is

things like chatbots and conversational agents which are really moving beyond

sort of simple scripted tools. And I think in today's society these

kind of things are becoming more and more important. So I think there's a lot of new work and a lot of scope for NLP research there. >> Yeah. Okay. So I mentioned that I'd

ask you about the future. Predictions are always tricky,

especially five and 10 years out. But I'm just thinking back on it,

I think that NLP now, not many people have predicted the world,

we're in now a deep learning and transformers and but and the rest of it

10 years ago, not even five years ago. So, but if you were to stick your

neck out, what do you think would be what would we be talking about if we had

this conversation five or 10 years time. >> Yeah it's such a difficult one because

I mean technology itself is changing so fast and NLP is changing very

fast at the moment as well. So I'm not very good at predictions. >> [LAUGH]

>> But what are things interesting? I mean the if you look at the the Gartner

hype cycle I think NLP has, well according to the most,

well the most reason one I've seen which I think was 2020 NLP is now passed

the peak of inflated expectations. And it's kind of starting to fall

all again towards the trough of disillusionment. And I think part of that is linked

to what I was saying before actually about deep learning which although

it's gained massive momentum. And until it's very easily able to tell

cats from dogs on twitter or whatever. But I said it's also this whole effect

is also starting to be a problem. People are realizing that

actually it's not I can't do everything that people thought it might be

able to do especially in the NLP Field. So in some sense people are getting

a little bit resolution but they're also moving more eventually towards a more

realistic view of what NLP could do. So if we look back say 15 years ago or

so I spent so much time trying to convince

people why NLP was important and everyone's like we don't need NLP,

what use is it to us? And I would spend every project I did,

I spent about the first year of it convincing everyone else in the project

why we needed NLP in the project. And the same with companies and

now we've kind of gone in the opposite direction where everyone wants NLP and

they want all these tools, but they don't really have a good

idea of the limitations. So it's kind of it's sort of

flipped a little bit in the other direction in that sense. And now we're starting to get gradually,

we're getting to the point where people are realizing to some

extent the limitations. So I think there's quite a lot

of uncertainty now around what exactly NLP can do and

where the limitations are. So this is a long winded answer. Sorry,

>> That's true >> I think eventually, what's going to happen is we're going to

come out of this in 5, 10 years, we're going to come out

the other side of this. And if we go back to the hype cycle again, I think we'll be moving

more towards this plateau productivity where we really

see we really get a much more kind of stable situation for

where NLP is at, but I think the certainly things

like conversational agents and so on and the speech side of things

will be massively advanced and I think NLP be will be very much

a normal part of life in many ways. Even if people don't. Yeah okay because it's under the surface

people don't see it you know? But I think it will become more and

more of a very a very natural part of everybody's life

not just sort of specialists. >> Okay great. So one final question to wrap us up,

it's a bit of a tongue in cheek one really about your experiences of NLP over the

years and the community, if there was one thing you could change a bugbear

a skeleton in the closet or whatever. One thing you could change about an NLP. >> Or the profession what would it be? [LAUGH] It's a hard

question to answer I think. I don't know if it's exactly

answers the question. But I would say the fact that

no one outside the field of NLP actually really has any clue what it is. I think it's one of the hardest

things to explain to people in a way that they can understand and

you kind of you have people who think they know what it is because they've

heard of GPT three or whatever. And they either think it can

sort of magically do everything or that it's some you know really niche field

that no one could possibly want to use, but they really have no concept of sort of

how complex the field it really is. And it's something that sort of,

it's really impossible to explain except with very concrete examples, so

I don't know how you would change that but [LAUGH]

>> [LAUGH] Well maybe that maybe that's the secret source that keeps us all in

a job, I don't know, we'll find out. Alright, great, thank you very much. Diane, and that was a very wide

ranging and thoughtful interview. So thanks once again, and I wish you

another successful 21 years at Sheffield. >> Thank you. [LAUGH]

# **Richard Jackson interview**

Great. So I'm very pleased to say. We're joined by special guest today who

has a ton of a practical experience in natural language processing. So to kick us off, can you tell

us who you are and what you do? Hi David, yeah. Richard here. I'm NLP data scientists that

a global pharmaceutical company. >> Right. Thank you. You can tell us a little bit yeah about

what your company does all the profession more generally if you like and

what your role involves. >> Yeah.

So my company is involved in all aspects of drug discovery and development. So this goes from primary research all

the way through to clinical trials and manufacturing that all

respects of getting. >> Great. >> Into patients. And how does the NLP fit in with that? >> Yeah.

So companies like of this size, they tend to have enormous

warehouses of data to manage. They've been running for often decades and

have gone through many different cycles of kind of like data management,

data, warehousing data etc. And my role is to use NLP to try and get the best out of the unstructured

text aspects of this. So for instance,

this might be built in a document or a sentence classifier to find important

piece of information in a large corpus of data may be working on

relationship extraction to try. And build a knowledge for our key

information around a certain domain. >> Yeah.

Actually just you mentioned you know you have a large corporate of unstructured

data that mainly scholarly works or patterns and news items. What kind of text. So we tend to work a lot

with the medical literature. Okay. So abstract that we won't get.. There's also a pre prints and

there's an awful lot of internal data so as well so when our lab scientists

record results from the laboratory work. We have like electronic systems

to capture that kind of free text data you know

all kinds internal and external are potentially sources

of interesting information for us. Yeah. Okay great. Thank you. >> So

let's just think about the NLP profession. You mentioned that your background was

data science as well as NLP which I guess gives you a sort of unique

perspective on the topic. What would you say is your favorite

book or block or resource on NLP? >> Yeah so I think it's not difficult to

nail down just one particular resource. But I guess typically I'm looking at

proceedings of all the major conferences every year. So the ACL, etc obviously

intend to the yield an enormous number of interesting papers

were very difficult to digest. And the one thing I found is

Sebastian rule that works Google. He produces a great newsletter

several times a year. This is a really fantastic read for summarizing all the major

developments in NLP. >> Yeah. >> So yeah it's really well written

puts a lot better in to it and sometimes I just cheated me that. [LAUGH]. >> Yeah. >> Yeah and I guess other than

that like I really I really love the hugging face library. This is a pipe torch based in the library. What we've managed to do is really

coalesce the NLP community I should say the post. But then it'll be community around a

simple life every and doing a good job of keeping up to date with all

the latest that went in for models. >> Yeah. Okay. Excellent. Thank you.

So you mentioned that your background

was I think he said data science, your data scientists originally

before you moved into NLP. I guess I have mainly been in NLP

rather than we just use different types of I guess like Donovan we

mainly focus on the NLP expect. >> Yeah. >> Anyway the reason I ask is that

would you say that over the years that you've been in NLP? And I guess going back also to your

academic days I think was was at King's College or? Yeah. Well what would you say it's with a

particular project that you feel was like a milestone in your career. Where you felt as if actually you know

this NLP think there's something in it and this is something I want to pursue. Yeah I can't have

misinterpreted this question. I really was a real learning experience as I kind of read that

as okay what went wrong. Can be, yeah, that's a better answer. Actually.

A better way of interpreting the question. Now let's go with that,

let's go with that. >> All right, okay. So I guess like as as

a practitioner in NLP more often than not your confronted with

data sets that aren't ideal. So for example, you might be

working with PDFs for instance, which don't have text encoded in them

exactly how you'd like it to be. Or maybe you're working in a special

specialist technical delay which comes with his own sub language. These can break a lot of assumptions

that come with a lot of the kind of pre trained models and

that are already out there. This can present huge challenges and trying to get something useful

out of your data set using NLP. So for me personally I work a lot

with a biomedical literature. This tends to involve processing long,

very complicated sentences that quite hard for

a human to read let alone the machine. So you captured enough context of

your system to do something useful with it can be quite tricky here. So now normally like we

read a lot of people, we try to stay up to date

with the latest research. But we tend not to bother with anything

that doesn't come with his own code base. Because what we find that

it's very hard to validate. You have to invest a lot of time to

actually validate the research yourself before you can even think about

getting anything into production. So that that presents quite

a risk to us in a kind of a job. So but

especially a hard lesson came to us one so long ago when we were trying to

implement a paper that we go about this. So this paper described very impressive

results that came with the code based on a pre trained models. So all the green flags are there. We came up with this plan where

we wanted to extend it and adapt it to the biomedical domain for

our needs. So after about 10 weeks of kind of

hitting her head against the wall, we kind of discovered that all was

not appeared what it appeared to be. And the actual Kobe sit still,

I wasn't able to read. He produced the results

that we obtained father, rather pretrained model that was supplied. This is very disappointing for us because

you're really not lost a lot of time with the assumption that it was our fault

that we were doing something wrong. Well, within the co

based that was provided. So having worked in research myself, I suspect what happened was that

the code base was made public. More than afterthoughts to

the original paper possibly to please some of you during

the period due process. And maybe the original waters

were probably working up another one somewhere that to me. So I think the big lesson here

is firstly obviously don't believe everything you read

when it comes to papers and be wary of kind of poor quality co faces

when you're trying to implement research. Unfortunately researchers don't tend

to be incentivized to make good code. And I think a lot of opportunities

are probably mixed because of this. >> Yeah it's a really valuable lesson in

fact the whole issue of reproducibility is important and is becoming increasingly

important, biomedical sciences. I think it's always been important but

in NLP as well now. So, yeah it's quite a topical issue and so let's turn a little bit to the work

you do with your employer and I'm guessing that you're

probably part of a larger team. So can you tell us a little bit about

the skills or competences that you would look for if you wanted to hire,

let's say a junior NLP specialists. >> Yeah. So I tend to see a lot of

CVS with from people that came to be claims to have

done a lot of NLP work. But what we actually tend to find

is they just run a few kind of out of the box classifiers on standard NLP

datasets mainly for their own learning. They need to have

the experience on the TV. So what we really look for is evidence

that they actually care about language and specifically with the main in

which they will be working. So I'd like to see like a real passion for

linguistics and the father and will be filled because

this is how you really get to grips with how your system is before him. Simply like knowing how to provide

good code is really important for us as I needed to just now

with the previous anecdotes. And more generally just understand the way

to kind of computer science ecosystem. So everything from how cloud platforms were stable networks have a broad

knowledge of all of this stuff is really important to make sure someone

is going to be efficient network. Because we don't spend all of our time

doing the really fun stuff in and it will be unfortunately training

system models a lot of it is kind of a hard slog to actually get things

to talk to each other correctly. And things to scale and yeah. >> Yeah. Good old classic

engineering skills I guess. So I was in short supply on me. So in a moment I'm going to ask you about

the future and ask you the tough question about predictions for the future but

just slightly easier take on it. What do you think it's

hot at the moment in NLP. >> Yeah I see a big trend around

trying to understand the data aspects of NLP like something

you alluded to just now. So we've seen them in the recent years. We've seen a lot of

development work on NLB models. This is being like a normal people. So we have birds and a whole stable for language models that

followed birds actually, what we tend to see is people still use

birds a lot in their primary research. So if it may be trying to

new classification here, they still use quite

frequently as a base model. I interpret this as, but it's

generally good enough as a base model. But what's less clear is how to

kind of really tune it properly for your specific task at hand. So in many cases the we've seen

only fitting on datasets quite significantly and it's kind of the poor

generalization performance and less than optimal real world performance. >> Yeah. >> So

what I really love to see is the emergence of these interprets ability tools. >> Yeah. >> It's really useful trying to uncover

the biases that come with data sets. One of these,

it's from google is called LIT. >> Was that LIT Or M I T. >> L I T. >> Okay. >> Yeah. >> So yeah like everything the whole

around this kind of interpret ability and models because especially in

highly regulated the main desires. It's important that we

can show like why certain a certain always made

a certain classification, it becomes very hard to get it through

the the red tape and put it into. >> Yeah well you could always just

switch back to logistic regression. [LAUGH]. But let's not go there. Okay. So that the tougher question although

it depends I mean it's not entirely we're not going to hold you to it but if you would if we were to have this

conversation 5 or even 10 years time. And I was to ask you the question

that I just did, what do you think? Well what do you think we would be talking

about if we were to do this interview in say 5 or 10 years? What would be hot then? >> Yeah so yeah obviously a tough one. I think we've seen some really amazing

accomplishment in NLP recently. So lots of people have heard of GDP 3. It seems to get really

impressive results but I don't think we've really seen

like an AI revolution yet. Certainly not what's been promised. So I think like one particular area that holds a lot of promise will be NLP systems that can translate human speech into code. Because I think when you have people be

able to speak about what they want machine to do it it actually translate into some

process that can automate some tasks. We've been quite complex tasks without

having to train the developer for the years to do it. >> Yeah. >> This opens up program into

a whole range of people that we wouldn't have even

considered using it before. So that that could be huge. I think that really like change

how we interface of technology and a whole range of interesting ways. I'm less optimistic about

bowing completely to be honest, because I think even defining the right

question is still a big challenge. The data sets seemed to

be still quite small that examine tater agreement

on these is quite hard. At least the loan schools,

I think that ultimately biology is just very difficult thinking and language

is a portal to express biology and with its complexity,

we just don't have anything better. >> Okay, great. Thank you Richard. So just to wrap us up there and I just got one final question which

really is a bit tongue in cheek. So just thinking back about your

previous experience with NLP or even your current role now. If there's one thing that you would

change about the professional or the way it's perceived other way

its taught or anything really. Any sort bugbear, what would it be,

what would you change? >> Yeah,

I do have a personal bugbear actually. [LAUGH]. Quite often, we see like really

interesting paper comes in the code base and

whether they're forced to or they decide to the authors a not for

commercial use license on the code. So it's not that we have a problem paying

for products that we think are useful. But what tends to happen is is that is

kind of well thought out about how like this commercialization of research

will work in a lot of cases. So typically will be directed to

the commercial wing of an organization. That they could be, won't well,

very frequently they won't really have much of a clue where to start in

terms of getting the conversation going. And ultimately we give up because

the procurement process just isn't worth the effort for

what essentially something that's not a product is just a set of upset

of code on get hub or somewhere. So yeah, there's a, there's a lot

more to commercializing research then kind of put in restrictive

license on the code bait and hoping the money is

going to flood in I guess. So I feel like if a group

is serious about it and they should give organizations like us,,

an interface where we can work with them. And as I said, like it's certainly not

a question about paying for the thing. It's just bringing the product to

market is very different from right to the research people. >> Yeah. Okay, excellent. Thank you very much Richard for

a very, this wide ranging and thoughtful interview. So yeah,, thank you again and all the best with your future

career in biosciences and in NLP. >> Thanks David wipe it out.

# **Udo Kruschwitz interview**

Great, so I'm really pleased to say we

are joined by a very special guest today who has a ton of practitioner experience

and has spent many years working at the intersection of academic NLP with

practitioner and industrial practice NLP. So to get us started, could you tell

us who you are and what you do? >> Well, thanks for

inviting me first of all, my name is Udo Kruschwitz, and

I have been working, as you said, in the area of information retrieval and

processing for 25 years or so. I work at the University of Regensburg

where I'm a professor for information science now and prior to that I was

a professor at the University of Essex. I'm still visiting academic now and as you

said, perhaps just too advanced thing, I am deeply rooted in the academic world

but I have worked on a lot of projects with industry and I particularly like

that anger of knowledge transfer, transferring knowledge from

academia into industry and back. >> Yeah, I think that's fair to say

that's a pretty unique perspective, and if you'll excuse me for saying so

it makes you a little bit of a hybrid, which is quite unusual I

think in the NLP community. So can you tell us a little bit about

what your current company does? I mean I know that you mentioned

that you're an academic, but obviously you have a lot of advisory roles

and a lot of interests in practitioner activities, you can tell us just

briefly a little bit about those. >> Yeah, so I mean working at

the university obviously the main idea is to somehow balance out

research teaching and admin. Being at a research driven

university of course there's more of an effort of getting

research output organized, and applying your findings

from projects from working with industry in your teaching and

that's great. So that's a nice thing apart from that, yeah of course there are projects and

collaborations with companies. So for example I have long standing

collaboration with committee in London used to be a startup signal. They are no longer start up I think right

now they call themselves to scale up. So five years ago so when we started

working with them was three people in a garage, and now it's 160 and there

are all sorts of collaborations there. So most of it has to do with

taking students they experience their knowledge and then throwing

them into the deep border and getting them to work on

real world problems. And it's beneficial for all sides, so

the company gets some ideas out of it, students get some experience. And for me, it's always a nice experience

as well to see this transition from being a purely academic reminded

person to someone who then says look, this is how we can apply it and

this is what we can't apply. >> Excellent, great, so

you mentioned a minute ago that you've got 20-25 years experience in NLP, [LAUGH] I guess over those years

you've seen a lot of books, blogs, resources come and

go in the NLP community. Is there a particular resource that you

find as you know stood the test of time or perhaps even at the moment

you find a particular book or blog is it is really

useful in your own work? >> Well with my NLP had on I would

say that you're a skier mountain speech language processing

book is the main milestone. It was a milestone when it first

appeared I think the first edition came out in 2000 or so,

because it was the very first book that comprehensively covered all aspects or

the main aspects of natural language processing in an up to date and

raw as well as deep fashion. And since then the second edition

I think came out 2010 or so, since then the third edition has been

in the making still hasn't come out, because the field is moving so fast that

as soon as you think the edition will come out it can't come out because

there's another breakthrough and there's another thing that's happening. So right now I still think it's

the best possible starting point to get to learn something about

natural language processing and all also to teach it, and

it's being maintained on the website. So although the book has

not come out in print, you find a lot of the stuff online and I would say that is for me,

really the main reference. Looking at from slightly different angles,

because here I am working in information science, I would say another reference

which I see as highly influential and relevant important is very

different is Donald O'Connorman. Donald O'Connorman thinking fast and slow because I come across a lot

of problems where I look at and I think we haven't you read

the O'Connorman's thinking fast and slow? Because the answer is no, and that bias

was described in detail with many examples in there, and for me, these I would say

are probably the most important and most influential reference

points in my work. >> Okay, excellent, thank yeah, I can

particularly well both of those actually, but the O'Connorman book as well as yeah,

they're both great resources. So again, reflecting on your

experiences over the years, and we've all had this sort of, well both of

us have been around in NLP a long time, and I think we've all occasionally take

a moment to reflect on our career in NLP. Would you say there's a particular project

that you felt was a learning experience either in a good way in the sense

that you thought, that's a milestone, didn't think I could do that and

I've achieved it? Or perhaps the other way around where it

didn't quite turn out the way you wanted, but not that nonetheless was a pivotal

learning experience for you. >> I would say for me the main

project in that respect was probably the first one I worked on

when I came over to England. So I applied for a job in at

the University of Essex in 1996, and then I went over an interview,

they said yeah, you can have job come away next month. And then I started early in 97 and there was really a massive

project was with British telecom. So, an industry driven project, [COUGH]

and the idea was to build a system, to build an intelligent

assistant that would help you to answer questions around

the yellow pages topics. So, for example, the old people among us, we still remember these yellow pages

books that came through once a year, and they were listing essentially all the

businesses, goods and services, and so on. And so typical question, would that be

something like I need a plumber, and the system should then be able to guide

you through whatever services are listed in there, so that you pinpoint the ones

that are most relevant to you. So for example the system should then say

in some sort of dialogue system okay, plumbers that's a bit

brought a lot of plumbers. Do you want to boiler cleaning,

do you want general coming, do you want this in the net? And by the way we're about do you live? And all of these things that in the end

you would say, okay here this is a company that you should call or here are the small

at that you might consult now. The reason why that was

a learning experience was because it combined the whole

range of NLP problems. So it was a question of

turning unstructured, partially structured text into

some sort of structured format. So the yellow pages, they came in some

sort of partially structured form where you had information that was very

clear, so the name of the business, but then the text that came

with it that was unstructured. So you have to somehow make sense of that. You had to build a dialogue system

which would be able to navigate you in as few steps as possible from

the users question to the answer. There was an information retrieval

element of course because in the end it was all about finding and

something in a large or larger scale database

of potential answers. And there was also a speech elements. So for the prototype we built, we bought

what was then the state of the art system, the nuance speech recognition tool which

came on a little tape, $5,000 I think. We paid for that and

you had to write your answer, you have to write your expected input

in some sort of grandma form and so on. But it had all of these elements in it and

that was 25 years ago. So for me that was a major

reference point in that respect. >> Yeah, it sounds great project. And of course dialogue systems have been

on a slow burn I think for many years, but it's interesting to see that over

the last five years they've really gone out of the lab and

into the mainstream again just sort of reflecting on your your

experiences over the years. And I know you've been involved

in hiring research assistance and also evolved in growing companies

like signal and hiring and acquiring, human talent and

resources for these organizations. What skills or

competences would you say you look for if you're hiring let's say

a Junior NLP specialists. >> That's very good question

actually because I have been through that process and

had seen success and failure. And what I would do nowadays is,

I would look at what sort of skills can you train and what sort of skills

can you acquire on the job and what sort of things are actually

more the mindset of paper. And in one particular case we hired

someone and we were looking for somebody who essentially is not available. So somebody with the user experience

with the design background, natural language processing knowledge,

game, identification experience and

all sorts of things. And it became clear that

we wouldn't find anybody would fit all of these requirements. So in the end we decided and p stuff,

if you have some basic background, you could learn a lot of these things,

you could actually learn it on the job. But user experience is something

which I found out you cannot teach, you cannot train easily, it's something

which you need to acquire in years and years of working in the area and training. And so this is for

me the main guideline really. What sort of stuff can

you train people on? And one of the things can you not,

you cannot change. You see that a lot when you

look at computer science demos where people say look, here's a demo and they clearly have no idea how the user

might want to in the interactive system. And fair enough because it's not their

job because you would have to have a user interface expert in there. But that essentially describes

roughly what I would go from A right, excellent, thank you. >> So in a moment, I'm going to ask

you about the future of NLP and you invite to make some predictions. Always tricky, but let's just reflect

a little bit about the present day. And you mentioned that with the giraffe

skin marking, but everything changing so rapidly and I couldn't agree more. What would you say is hot

at the moment in NLP? >> Well, NIP I mean we've seen

a massive paradigm shift in the over the last few years. So what we see now is

that many of the problems are addressed with

transformer based architect. And as long as you have enough data to

train on the original language models, then you will outperform

any other approach that you might reasonably

take as an alternative. And the funny thing is that

if you look at the original birth paper because birth

is the main reference point net respect that didn't come

out until 2018 on archive. And it was really probably published in

2019 Knuckle, that's only two years ago. And if you look at Google Scholar,

last time I looked, I think it was something like

17,000 references within two years. So that gives you an idea of

what is really hot in NLP. And obviously this is something

which I would say is the single most important thing that has changed

the world of NLP obviously there were a number of reasons why this happened

because more data became available, tools became available

accessible to anybody. Pretrained models were made available

by companies like Google, Facebook etc. And also have the architecture CPUs etc. Without all of these elements,

this change would not have happened, but that to me is currently the biggest thing that we still see coming

through to its full potential, really. >> Yeah, excellent, great. Okay, so here's the tricky one. And again, it's worth pausing a moment and just saying if we'd had this

conversation five years ago, could we have predicted that we'd

be talking about transformers now? Probably not, and

with that caveat in mind, what if we were to do this

interview in five years time? What kind of things do you think

we'd be talking about then? >> A good question? Ask me again in five years time. [LAUGH] I mean looking at it really, it's only been 2.5 years for

this to happen. It will be interesting to see what I do, I think it's interesting to look back and see what sort of NLP problems were

considered hard and impossible to solve. And I mentioned earlier

speech recognition. So when we bought this nuance to it, Which I still have some on my shelf here,

so I can see it was impossible to do something which would be robust

enough even for a single speaker. So it worked reasonably well for

me as a non-native speaker, funnily enough, but whenever my English

colleague tried to say something. The system fall over even though

we had to describe the input in some form of grandma structure. So speech recognition,

I don't think anybody could have predicted 10 years ago that this would be so

easily handed nowadays, let alone speaker independent

speech recognition. And because of these changes,

I would find it very hard to predict what obviously what I do

think is that right now, one of the biggest problems is that

you need massive amounts of data. And I think one of the things that

will that we will see over the next years is a more sophisticated,

more selective way of using data. So rather than throwing tons and

tons of data at the problem, you might actually get to the same

solution in a more sophisticated way. But speaking about active learning for

example, so thinking about ways that because active

learning is already being deployed in meaningful ways where it's

expensive to prepare a no data data. And I think we will probably see more

of that you do no longer need these massive amounts of data as long as

you pick the right data points. But I think what we will see and

what we are already seeing is NLP becoming ubiquitous in every day,

like we already see that. I mean think about websites

when we grew up there was no web search of

course shortly before you retired the web search engines came along. And nowadays, you type something

into talk into your device. And it's giving you answers, it's no

longer giving you 10 blue links, but it's giving your answers. And these things have really being

introduced fairly rapidly and nowadays, it's normal that you get

an answer of course you expect an answer, but think back five years ago and

it was not the case. So I think ubiquitous NLP architectures, solutions, all nets,

that's something that we'll see dialogue systems mentioned earlier,

etc. >> Actually it's just it's worth

just kind of pausing a moment as you mentioned there about the tasks,

what we consider difficult. And reflecting on the squad leader board

and other benchmarks a lot of these things are topping out close to levels

of human performance, aren't they? So I mean, that in itself kind of out

raises the question, where do we go next once we've achieved human levels of

performance, what is beyond that? Yeah.

>> You're right in a way, but the interesting thing is that if

you look at some recent papers that have appeared in the top NLP conference,

like ACA for example. You see that there is some real

concerned about how to interpret these performances because you also see examples

where the test data is slightly modified. And then these systems that

appear to show beyond human level performance suddenly drop in

performance quite dramatically. And so there's a whole issue about

how robust are these data sets and is it maybe just too closely

tailored to training and test data, and it doesn't really tell us something

about the real world obviously does. But I think some of these things should be

taken with a pinch of salt and more and more people actually raising this

concern saying look should we really believe these? And we probably shouldn't,

>> Yeah, okay, excellent. So just to wrap us up then and this is a kind of slightly

tongue in cheek question really. So if you could change one

thing about the professional, perhaps your own career within it either

past or present, what would it be? >> That's the most difficult question

perhaps because I'm perfectly within my community. One thing that I do see is that

you have different communities. So even just looking at the academic

community, you have the energy community, you have the information between the

community, you have the speech community. They all have their own conferences,

they all have their own meetups, they all have their own little bubble. And we see more interaction between these

different communities happening now, maybe a bit more of that will be good. The other thing that I observe, but I don't want to change that

because it's quite funny. You often have when you look

at how academics pitch their solutions when I'm reviewing papers and

when you go to conferences. You often see this sort of pattern where

they say, look imagine if this problem now here's the solution and then you

sit the audience you think yeah, but you don't have that problem,

[LAUGH] that's made up problem. And in the way I quite like that,

I don't want to change that. I mean some people might argue that we

should be more driven by industry needs, but I think people coming up with

solutions to problems that don't exist, I think should also be there,

I quite like that. >> Curiosity driven research,

I think the term I'd give for it. Great, all right, thank you, Udo. That was a wonderful journey through your personal experiences

with the NLP community. And thank you for

joining us from Regensburg and yeah, all the best with your future career in

in the information Science department. Thank you again. >> Thank you very much