[Automatic captions by Autotekst using OpenAI Whisper V3. May contain recognition errors.]

[SPEAKER_00]
You are now listening to a podcast from the University of Stavanger.

[SPEAKER_03]
Hello, everyone, and welcome to the episode seven of the reginno podcast by Center for Innovation Research University of Stavanger.
I am Muzammil, joined by my co-host Raj today.
And we have with us Milad Abbasiharofteh, who is Associate Professor at Business School Aalborg University, Denmark.
and working on innovation and applied data science.
Milad, welcome to the podcast.
I'll start right away.
What's your story in the field of economic geography?
How did you land here?
And what's your fuel?
What drives you?

[SPEAKER_01]
Okay.
Hello from my side.
Thank you for having me.
It's a lovely place to be.
And thank you for your invitation to be on this podcast.
Well, I'm originally from Iran.
I studied bachelor program in urban planning, geography.
In Iran and in 2011, I guess I moved to Germany.
I did a master program also related to urban agglomeration.
And I remember at the late phase of my master program, I attended the second geography of innovation conference.
And it was a milestone decision in my career because there I met Tom Broekel, who became my PhD supervisor after my master program.
And that brought me to the, in my view, very interesting and vibrant field of economic geography.

[SPEAKER_03]
Yeah, speaking about conferences, I had some similar experience.
Probably if I talk about my first PhD paper, it was a pure result of attending a conference, which was a global conference on economic geography in the US.
And these conferences, I believe they're like, it's super interesting to attend and probably the type of networks you get in these conferences.
Yeah, so...

[SPEAKER_02]
Yeah, I can bring us right to the topic and explain a little bit on the background on why Milad is at Stavanger University to begin with.
So in our department, we're very curious about one method, which is text analysis.
We have a few colleagues who are experts in text analysis within the

group, and we have so many other researchers who want to go into text analysis.
And that's where we thought Milad would be an excellent help to our department.
So we invited him over.
for a two-day workshop and we talked a lot about text analysis how to do it what it can be applied for and everything and that's why we now also think that it would be a great idea to record this on a podcast so milad what do you think are the advantages of recent technological developments on research how can we do research differently now that was not possible let's say five years ago and and what can we learn from whatever you are doing at albergue

[SPEAKER_01]
Very good question.
So let's see that like different waves, methodological waves of how collect and analyze data.
So what I put under the first wave is, okay, you go to the field, you interview people, you collect information, so-called primary data.
and you analyzed.
This has been done before and it's still a valid method.
The problem is, however, you cannot scale easily.
It is associated with the high cost, resource intensive, and simply you can't do that for a specific case.
Could be a set of companies or a place, a city or region.
So what I put under the second web is using secondary data sets.
So at least in our field, a lot of researchers have done interesting research with patent data, trademark data, R&D data, so on and so forth.
This is nice.
We also learn a lot from this type of data collection and analysis.
Again, in this context, the problem is if you are interested in a certain group, like startups or individuals that are not well represented in secondary datasets, then by default, you overlook these groups.
The next wave I would simply label it as using textual data.
So those startups, I believe, have some digital fingerprints on the web.
And
They show and talk about their products, their activities.
And the problem is that these are not reflected in secondary dataset, but we should find a way, find a set of methods to extract this information from unstructured data.
And this has been an interesting set of projects I've been involved in with some colleagues from Utrecht University, University of Groningen, and now at a business school in Aalborg.

[SPEAKER_02]
Yes.
Can you give us some examples of what can be done?
What creative ways have you seen recently and what more can we do in that field?

[SPEAKER_01]
Yes.
One example that is a broader project and also part of it is the idea of web AI and the web AI paradigm of innovation.
So the idea is that companies' web pages provide a very interesting source of information.
If you study innovation, if you study how regions diversify into new activities, whether it's related or unrelated terms that we use a lot in
economic geography, evolutionary economic geography, and
since smaller companies, especially smaller companies or startups, are more active on the web compared to well-established data cells like R&D and patent data, we can capture their activity much better.
And then we can have a better approximation of innovation, a broader definition of innovation, and integrate these new insights, more detailed insights into our empirical work.

[SPEAKER_03]
Yeah.
Milad, before your workshop, I was quite used to see and read economic geography done with the patent data.
But you talked about using trademark data, and it was quite interesting to me.
So what can trademarks tell us and what can we do with the trademark data?
What can it uncover?
For example, it can uncover a range of insights.
According to you, how are you using the trademark data?

[SPEAKER_01]
Very nice question.
So let me give you a little bit of background.
So during my PhD project, I work with ThumbVocal and I work with, well, widely used data sets like pattern data and lead data.
And in one paper, scientific publications, these were nice.
And then I kind of diversify into related field of research if I want to use this terminology.
And it started when I started working with Carolina Castaldi and Sergio Petranio at Utrecht University.
And Carolina has done very interesting research in using trademark data.
I would argue that not all innovation can be patent.
And basically patents are designed to capture either technological breakthrough or technological solution, technological advancement.
But
Other aspects are missing, so commercialization capability of companies, or if companies are innovative in bringing new services to the market.
And trademark can nicely approximate commercialization part.
So if you see innovation as a process, well, to make it simple as a linear process, so it starts with some scientific breakthrough leading to technological innovation or technological solution, and then it would translate to commercialization, some market applications, if successful.

So in this simple linear model of innovation, patents can capture technological solution or technological part, and trademark, I would argue, would be a better data set to capture what companies bring to the market, what is market innovation.

[SPEAKER_02]
But now you've been able to combine the patent data and trademark data.
Can you walk us through the process of how you did it and what new range of activities now is possible for policymakers to look at?

[SPEAKER_01]
Yes, so this is basically the result of my postdoc projects at Utrecht University.
The question is pretty straightforward.
So I name a product like a smartphone, and I ask you what technologies are used to develop this product.
And you might do a little bit of search on your own and come up with some technologies needed to develop this product.
And you can ask the question the other way around.
So I give you a technology and I ask you, what are the technology you have to combine with this technological capability to develop certain products?
This is the basic question of this project.
However, the complexity is in the fact that we don't have one or several products.
We have thousands of products on the market.
On the other side, we have hundreds of technologies on the patent side, on the technological side.
So we use some statistical techniques.
and text analysis to find meaningful relations between technologies and market applications.
To do so, we analyzed millions of patent records and about two millions of trademark data, trademark application.
So to address the second part of your question, what would be the application?
So one simple application would be, and I think is very policy relevant, how we can capture technology diffusion in the market.
So let's say...
Last year or five years ago, we had a technological breakthrough happen in, for instance, Stavagno.
So we want to know how this would translate into market application.
And we are geographers or in economic geography, we are interested also in how and where of economic phenomena.
So if technological solution brought to the technology world in Stavagno,
where we expect to see the market application of this technology.
So this concordance would help to actually track new technologies into the market.
And then when we have this, we can model the diffusion of this technology into the market using well-established techniques.

[SPEAKER_02]
That is super interesting.

[SPEAKER_03]
Yeah, and talking about GenREDO AI, initially when GenREDO AI came, we had a question whether to use it or not to use it, which is not a question anymore.
Now the question is how we can use it in order to make maximum out of it.
How is GenAI helping this text mining approach, for example, or applied data science techniques for innovation research?
And what role can GNI play in this?

[SPEAKER_01]
Yeah, it's indeed a hot topic.
And a lot of colleagues talk about that, whether we can use it or not.
I mean, in the context of economic geography and innovation research, of course.
And if we use it, whether we can rely on the result or not.
So to answer your previous question, I talked about three waves of methodological approaches and collecting data.
So I would put it under the fourth wave that you do not need necessarily programming.
You don't need programming skills.
And you can, relatively speaking, at a good price, get access to different APIs.
And through API calls, you can ask research-related questions.
So to give you one example, I did explore some aspects of my research using generative AI.
For instance, in economic geography, it's interesting for us to know about the geography of products.
So where specific products are being brought to the market.
And earlier exploration showed that using GenAI actually provides us very interesting and nuanced understanding of where products are being brought to the market by what companies.
And since we saw this, observed this interesting result, we sampled around 1,000 companies' LinkedIn information.
specifically startups, which are not well represented in patent data or trademark data.
So in this data set we created, we can see what companies put on their LinkedIn web page showcasing their products, basically goods and services.
We asked GenAI
Given the name of a startup, location, and website of this startup, if available, what is the main goods and services this company brings to the market, and then compare it with LinkedIn information.
And we found across these 1,000 cases, actually what GenAI provide us is reliable, and it's a good approximation of what these new companies bring to the market.

[SPEAKER_02]
Yeah, that's super exciting.
Yeah, but I have one more question for you, which is for someone who has never worked,
in text analysis before or large language models before, not using

it, but by using it for research.
How would one venture into this topic?
Where can they start to look into?
And the second thing, which is something that you mentioned in the beginning and just recently as well, the cost associated with it.
What can you do without having any costs associated with it?
And how much would it cost for you to do?
What kind of things?
Could you give us a bit of a background?

[SPEAKER_01]
Yes, definitely.
So I think there are a number of very nice review pages.
I think our field of geography is lagging behind a bit.
If you do a bit of literature review in management, economics...
computational data science, of course, these fields, you always find very recent literature review that provide you a nice overview of current state that you can start with.
I believe to start, you don't need to know all nitty-gritty things and programming, you don't need to have programming skills.
And Genotype AI actually kind of democratized access to this type of tool.
This is the first step.
And one thing I want to point out, and it's very important, is validation.
So
I review sometimes, review some papers that it has a nice approach, very smart, empirical setting, but authors do not provide any systematic validations, validation exercise.
We still don't know how these results are created by GenAI.
If you use GenAI or other
AI related techniques.
So it's really important to make sure by very well-defined transparent validation exercise that the output is reliable.
And it would be great if you repeat the second part of your question.

[SPEAKER_02]
But there was the tools that weren't used to learn.
Where can they learn it and how much will it cost to begin doing this?

[SPEAKER_01]
Ah, okay.
So I think in terms of cost, it's much cheaper now compared to pre-region AI.
era, like five years ago or so.
And if you want to start with the well-established, well-documented procedure, I think starting to have chat with the generative AI would do the job.
It's really important to have that in mind.
This makes sense if the topic is well-documented online.
But if you want to start from a niche area, then it works.
I think it makes more sense to go to conferences, talk with

colleagues who have more expertise or experience in this field.
Another thing that I want to point out and is related to the cost you mentioned, we should know, of course,
These AI tools, generative AI specifically, are very interesting to use.
But in terms of research, I have the feeling sometimes some research questions can be addressed and done in a transparent manner using much simpler, older, I would call it, old-school NLP techniques.
So this is important to keep in mind that we don't need generative AI for all types of research questions.
And also important to mention that using LLMs has ecological footprint.
So when we enter a prompt, normally we hit up some server somewhere in the US or Asia.
So it's also important to keep in mind that using LLMs has some good environmental impact.

[SPEAKER_02]
Yeah.

[SPEAKER_03]
Yeah, thank you very much, Milad.
So I think we'll move to our last question, which is related to the keyword that we asked previous guests.
So the last guest, he left you a keyword, frugality.
So would you like to give two or three sentences related to frugality?

[SPEAKER_01]
Okay, this is a difficult one, but a good one.
I know about frugality via a professor from Hanover University, Ingo Liffner.
He has a number of nice papers on this topic.
To, well, loosely say something about frugality, for me, it's doing more with less.
I think it becomes relevant when we put sustainability in the context.

[SPEAKER_03]
Yeah, thank you.
And what's the keyword you would like to leave for the next guest?

[SPEAKER_01]
Okay, to also advertise our new paper, it's a joint work of 23 authors on web AI paradigm of innovation research.
I would use web AI and I hope it's also relevant for your listeners.

[SPEAKER_03]
Yeah, thank you very much, Milad.
And thank you very much to our listeners as well.
Until we meet next time and keep uncovering stories and ideas behind academic papers that shape our understanding of innovation and regional economic development.
Thank you very much, Milad.

Thank you, Milad.
Thank you for having me.

[SPEAKER_00]
You have been listening to a podcast from the University of Stavanger.