TECHNoCULTURE

Authorship attribution

Episode 25

Full transcript

Guest: Mike Kestemont [Mike]
Host: Federica Bressan [Federica]

[Federica]: Welcome to a new episode of Technoculture. I'm your host, Federica Bressan, and today my guest is Mike Kestemont, Assistant Professor at the Department of Literature at the University of Antwerp in Belgium. Welcome, Mike.

[Mike]: Hello, good morning.

[Federica]: Thank you for being on Technoculture. I'm very interested in your work because Technoculture is not just interested in how digital technology impacts our lives, our experiences, but also how it changes the way in which we do research, in which we acquire new knowledge and make sense of it. Your work is situated in the digital humanities, so quintessentially multi-disciplinary. To begin with, can you tell us a little bit about your background and the context in which you work now in Antwerp?

[Mike]: Sure. So, I started out as a student of language and literature, and then I wanted to go on and do a PhD, and the PhD was going to be about authorship attribution, so can we use modern technology to identify the authors of medieval texts? And often, we don't know these authors. And it turned out that people were making a lot of progress in that area using computers, so that's why I started studying what was called applied computer science back then, and I used it in my PhD, and that's how I got involved with what is now called the digital humanities. And from there, I got more interested in artificial intelligence in general, and so in my work now at the Department of Literature, basically what I do is explore how artificial intelligence can be used to push literary studies, really, and how we can use AI to process or learn about texts in fundamentally new ways, I think.

[Federica]: I first learned about you through a project, or I should say a video documentary of a project you participated in. The project revolved around Hildegard von Bingen, and it caught my attention first and foremost because I'm a true Hildegard von Bingen fanatic. For those who don't know, Hildegard von Bingen was a Benedictine abbess. She lived between the 11th and 12th century. She was a writer, composer, a visionary and a polymath, and how I fell in love with her is that I heard Ordo Virtutum, her most famous composition, performed live many years ago. So I watched your video with interest, and the project was not really about her music, but about her writings. Even if it was a few years ago, would you like to tell us a little bit about that project? To begin with, what was the research question and what your role in the research team was?

[Mike]: Yeah, so I was actually contacted by two colleagues at the University of Ghent in the History Department, Jeroen Deploige and Sara Moens, and they were editing text by Hildegard, but the thing is that these texts were a bit off stylistically, and they started doubting whether Hildegard actually ever authored these texts. And then they knew about my work in authorship attribution, and they said, 'Hey, can we collaborate and find out what is going on with these texts?' And there we, so we applied stylometry to these texts, and we got very strong and solid results actually showing that if she ever wrote these texts, they were significantly reworked by one of her secretaries. So there was a very nice project and then I submitted it for a conference overseas, and it got accepted, but I didn't have money back then to pay for the travel, and so we saw that as an opportunity to make a documentary about this, and then eventually this documentary would air on the conference for the first time. So it was a very nice thing, yeah.

[Federica]: This is the video I saw, right?

[Mike]: Yeah.

[Federica]: We will link it to the description of this episode so other people can see it, too. So how was the research done? I mean, to answer that question, what kind of methods did you use? What did you do to tell, is this text by her or not?

[Mike]: So basically what we do in authorship attribution is, we look at small insignificant words which we call function words, so these are these small little words that don't have a lot of meaning, that are useless, that have a grammatical meaning only more than an actual semantic meaning, and in Latin this would be like the word et for 'and'. So it's a word that is often used, but it isn't linked to a specific kind of topic or what have you. And the idea is that if you look at the frequencies of these texts, of these words in these texts, often you see that authors have very individual distributions as to how they use these words. So, for instance, for

Hildegard, one word is very typical, and that's the word in in Latin because she always sees things in visions, for instance, so whenever a text has a high frequency of in, basically that's a marker that it could be written by Hildegard, and then you start combining all these little pieces of evidence, you can build up a very solid attribution on the basis of these function words.

[Federica]: When you say that, that you can attribute a text to a person or probably not attribute a text to a person, say, 'No, they didn't write that,' do you mean that you could bring this as final proof into a courtroom? I guess what I'm asking is, what is the degree of certainty of attributing authorship like that?

[Mike]: Well, so I know that in the US and the UK, stylometry, or what is called forensic linguistics, is actually admittable as evidence in courtrooms. And there's this very nice series on Netflix, the Unabomber, that actually shows like the early origins of this kind of research in a forensic context. I don't think in Belgium that they allow it, although we often get contacted by the police, for instance, asking whether we can help out, and we're always willing to help out, but often you see that the conditions are not met, so there's some certain conditions like texts have to be long enough, you need to have enough example material from an author. And so whenever these conditions are not met, we always say that we cannot collaborate because we wouldn't be able to give or provide evidence with a sufficient, 'significant', as they call it, statistically.

[Federica]: How much text did you have for Hildegard?

[Mike]: Yeah, for Hildegard, she wrote so much, so that really wasn't an issue. We had hundreds and hundreds of thousands of words because she's been active all her life in so many genres that it's actually very interesting to look at her oeuvre because it is so large and varied, so that was never an issue.

[Federica]: Is there a minimum amount of text you need to have to do this analysis, or it depends?

[Mike]: Well, there are some papers, but so... There's two sides to that coin, because I often get that question, 'How much text do you need?' There's, on the one hand, the anonymous text, and one number that is often given is 2,500 words, but the longer it is, the better, of course. And then on the other side, you also have the size of what we call the training material that you have for [an author 00:07:41], and there, the more you have the better, and you would need, I would say, at least 50,000 words to have a reliable model of somebody's writing style. So text length is very, very important. Also, genre. It's very hard to do authorship attribution across different text varieties, so if you would train a system on an author's oeuvre in novels, for instance, it's very hard to recognize the same hand in theatre plays by the same author.

It's one of the big challenges in the field right now.

[Federica]: So authorship attribution is a specific application of stylometry. We're in a technical scientific domain here, and words are often taken also from the everyday language, but they mean something very specific in this context, so if you don't mind, I would like to ask you to clarify these terms a little bit for us. They are certainly fascinating concepts but not the easiest, probably, to comprehend, so if you don't mind, I'd like to ask you about artificial intelligence to begin with, as a very broad area, if I understand correctly, of which machine learning is subset, a sub-domain. Deep learning is a subset of that, and distant reading is a subset of that. Do I get the hierarchy right?

[Mike]: You could say that, yes. You could say that. Distant reading is very hard to define because it comes in many different forms, so it's sort of a container term, but it's hard to give like a definition of it. It basically comes down to researching a lot of text, using some sort of automated techniques nowadays, and wouldn't have to be deep learning or any kind of thing. So sometimes they have very simple analysis that could also count as distant reading, you could say.

[Federica]: Like indexing web pages?

[Mike]: Yeah, like the classic example of a simple application is the Google Books paper that was published in Science a couple of years ago, and there, so they just had a lot of material, so they claimed that it was like, I think, 4% of all books ever published in the history of mankind. What they basically did was look at word frequencies throughout the years and how these shifted. That's a kind of distant reading because they process a lot of text using computer tools, but what they did was simply word counting, so that's a very simple application, you could say, but it also qualifies distant reading. So distant reading is something that can be done without fancy machine learning.

[Federica]: It drives me nuts that these labels have all words taken from the everyday language, but they don't really mean what we mean, like what does it mean that a machine learns and all of that? In the case of distant reading, what's 'distant' about it?

[Mike]: One term that I like is, was suggested by a colleague of mine, is panoramic reading, so instead of being on the ground looking at the flowers, what you do is, you take a drone and you fly over a literary landscape. That's another metaphor that explains what distant reading could be.

[Federica]: What is it about the deep learning movement that makes you enthusiastic about it? You're a self-declared enthusiast of the deep learning movement.

[Mike]: So I was using machine learning for my PhD, very simple machine learning. It was only after my PhD that you had this emergence of what is called neural networks, and this is actually a very generic and powerful way of doing machine learning, and for me, basically changed the way I did research because I was able to use much more complex models and actually understand what was going on. And people often say that neural networks are black boxes, but I never agreed with that. For me, this is a very intuitive learning algorithm, and I actually understand what is going on in what people call a black box, but it isn't, because it's actually very simple math that is behind it. So that's why I like this kind of machine learning because I understand it and I think it's intuitive.

[Federica]: Moving up in the hierarchy I outlined before, we find machine learning. What does it mean that a machine learns, and what is it that a machine learns? First of all, by 'machine' we mean a computer. Learning involves the concept of intelligence, like you mentioned before. You change, you adapt yourself. How do these things go together?

[Mike]: So intelligence has one very important component, and that is learning, so it's basically what humans do. What you do is, you adjust your behavior in such a way that you anticipate future rewards, so on the short-term or on the long-term. So basically, you change your behavior because you think that this will pay off in the future. And what we call AI nowadays, artificial intelligence, is often machine learning, not necessarily robotics. So when you mention the term 'AI', people often think about robots, but that's only one part of the story. I think much of the progress nowadays is just realized in software, so there's no robots that are involved in this. And when we say 'machine learning', you have two kinds of learning, and there's one learning that is one kind of learning that right now is very successful, and that is what they call supervised learning. And basically what you do is, you have a neural network look at examples of tasks that were provided by humans, and the simplest examples is, for instance, image classification. So you give a computer 1,000 images of dogs and 1,000 images of cats, and based on these examples, the computer learns to discriminate between dogs and cats, and that's something that is actually going pretty well right now, so and this — and 20 years ago, this was pretty difficult, distinguishing between a cat and a dog in an image, but right now, that's a solved task. The problem there with this kind of supervised learning is that we as humans still have to provide all these examples, and when collecting thousands of images of cats and dogs, that's something that's time consuming, it's laborious. And that's why people are now also very interested in the second kind of machine learning that is called unsupervised learning where instead of giving a computer these examples where you have labels that have been applied to images, you just give the computer images and have it learn just on the basis of raw data. That's a very exciting paradigm, you could say, in machine learning, because then the human doesn't have to do much anymore, doesn't have to provide all these annotations, and the machines are actually learning for themselves. And there's also impressive progress being made right now in that area. And that's, of course, the holy grail, you could say, of

intelligence where machines teach themselves to be intelligent.

[Federica]: Besides research applications like stylometry, for example, what are other applications of these ever more sophisticated algorithms, more intelligent, in the real world? I don't know, advertisements suggested to me on Facebook?

[Mike]: Yeah, so the simplest application, I think, right now is face recognition. You have that on social media now so that computers can actually recognize who is depicted in a picture. I think that is the most visible application right now. Recommendation systems on what have you, Amazon or even Netflix, that's also a form of AI that is very popular right now, and there you see that the way we consume cultural content is actually increasingly being guided by machines and what a machine thinks that would be good or interesting for you to read or watch.

[Federica]: CAPTCHAs we sometimes get where we are asked to identify street signs, for example, that's using us as labor for supervised learning. Is that correct?

[Mike]: Exactly, and of course, now in the Silicon Valley, and these self-driving vehicles are the next milestone that people have identified on the way to true artificial intelligence, you could say, and there, you just need a lot of labels to be able to recognize road signs, because you often see them in very different conditions. So that's why firms like Google have come up with these smart ways of actually collecting annotated data. These RECAPTCHAs, you used to have them with letters as well, and so often Google would actually harvest data from there also for their OCR of books, so it's a very interesting application, you could say.

[Federica]: In the early days of artificial intelligence, I know that it was believed that it would be rather easy to teach a machine how to process the external world, the input from there, but then we actually learned about how complex the processes involved in our cognitive abilities are. Having a body, by the way, is something very much related to processing the environment, interacting with it. You just talked a little bit about where we are now, and it's indeed, remarkable, but there are, for sure, tasks and critical cases in the examples you just mentioned that are still unsolved, so to speak, that require an advanced cognitive ability. To discriminate between, well, for laughs, a cat that looks like a dog, I don't know how that can be, but an example. A street sign that is depicted in an advertisement board by the road. Now, self-driving cars should not want to interpret that street sign, but it's a street sign. So it doesn't mean to be the question of a skeptic. I'm really asking because it's all so fascinating. Do you think that we will get there, that we will be able to have machines that can process information and discriminate between critical cases like with the same sophistication that a human can do, if not better?

[Mike]: I think autonomous driving is something that we will still see before we die, but

that is something that will happen, that is bound to happen, that is perhaps a good thing to happen, because I think that machines, if they are trained properly, can be much better drivers than we are. But yeah, I could see that happen. We're not that far from it, I think, but this is something. . . This is a move that we have to do with the society as a whole, I think.

[Federica]: Is there something in your field of research today that is not possible yet, something that you would like to see happen?

[Mike]: So I already mentioned OCR, optical character recognition, and that's something that works pretty well for printed texts. Something that everybody thought was impossible when I started in academia was handwritten text recognition where you would give it a photograph of, I don't know, medieval manuscript or perhaps a manuscript of Samuel Beckett, and you would actually have a computer read the handwriting, which is much, much more fuzzy than printed text. And now, there's actually research showing that this is feasible, and so I suspect that in the next 10 years, research will be able to get to the point where we can also read handwritten texts, and that's very interesting in my specific field, because in history most of the books are handwritten instead of printed. As a very specific example, for somebody in my field, that's very exciting because it means that we would actually be able to search, for instance, through all the Latin manuscripts in the Vatican. That would actually be feasible. It's not feasible right now. Right now, we still depend on transcriptions of people, which is very laborious also, as a task, to transcribe these manuscripts. So that's very exciting.

[Federica]: In a previous conversation with me, you mentioned the word 'serendipity' speaking about the application of machine learning techniques in your field of research. That was very fascinating to me. What do you mean by that? How does serendipity fit in the picture?

[Mike]: So the thing is. . . So in my field where people often say is the advantage of distant reading is that you don't have to do a preselection anymore, and that means that. . . So human reading is something that takes time. You have to be attentive, etc., so you will always limit your reading activity, and often you will limit it on the basis of bias because you read the text because other people have suggested it might be interesting as a text. So basically you're limited to a preselection. With distant reading, it's actually able to read all English literature in the 19th century, so you don't have to kick anything out anymore because you don't have the time to read it. And the nice thing with these computational analyses is that often you have these outliers that would show up where the computers tell you, 'Hey, this could be interesting.' And often, you make these surprising findings that I would call serendipity. So I think serendipity is a very nice side effect, you could say, of applying digital methods in the humanities. We didn't have that much serendipity in the humanities before the advent of the digital humanities, so I think that that is one of the added values of the age.

[Federica]: So you're saying I understand how the process works, but when I run it on a large dataset, something I could not process by hand, the results can still surprise me.

[Mike]: Yeah, yeah.

[Federica]: I'd like to go back to function words a little bit. So there's small words like prepositions, conjunctions that we use all the time, define how I speak more than the nouns I choose. I wonder if the use of function words would give away the fact that I'm not an English native speaker. That fascinates me because it's counterintuitive the first time you hear it. So can you explain how that works? Meaning, you count the words, you have two people and you say, 'Okay this person uses this function word this number of times and this person this number of times.' What do you do then? What kind of analysis can you do or how do you compare?

[Mike]: Yeah, so that's difficult with function words because they don't mean anything, it's often hard to interpret why an author would use this preposition often, for instance. So I don't have a good answer there, actually. It's actually kind of black magic, I always say, the fact that it works so well. We don't really know why it works so well. Part of it is just information theoretic. We use much more function words than content words, so that means that even in a short text you simply have much more measurements for function words than for content words, and that might actually also explain why they work very well. So it doesn't have to be an interpretive reason.

[Federica]: Are differences among genders and across languages factors that are taken in consideration in the analysis?

[Mike]: So gender definitely, so often in gender studies, people would also use function words to differentiate between the biological or cultural gender of an author. Languages obviously, there it's also very useful to do automatic language identification.

[Federica]: Languages with less prepositions, are they harder to study?

[Mike]: So the thing is, there, you often benefit from applying some sort of morphological analysis to the words because in these languages, like Latin but also more extreme cases like Finnish, people would actually glue their function words to the content words. There, it's interesting to chop them off, you could say. So there, it's also possible to apply this idea of function words, but you apply it to morphemes inside words rather than isolated words.

[Federica]: The algorithms that process a large amount of data, such a large amount that one human being couldn't do it, are they as intelligent, or more intelligent or less than a human being, if ideally?

[Mike]: Definitely not as intelligent as a human. There's nothing that beats a human when it comes to interpretation, etc. They have advantages that humans don't have, of course, so scope is one of them. They can simply read more, so they do it in a more shallow way, you could say, but they can process more. They're explicit about what they do. Often, humans find it much more difficult to make explicit why they like a certain book, for instance. It can be very hard for a human to express that. And they're consistent, so you can make sure that if an algorithm analyzes a text, that it does so in a completely analogue fashion in the future. So if I would read a book that I also read when I was 15 years old, the way I read it now would be different, right, because I have changed as a person, too, you have different associations, etc. So computers don't have that issue. When they process a text using a specific algorithm, they can do so multiple times always in the same way, and that kind of consistency is also an interesting advantage of algorithms. So scope, consistency and explicitness.

[Federica]: It's interesting because you said that learning brings as a consequence the fact that you change, that you adapt your behavior, and you read a novel differently because you have changed in, say, 20 years, and now you just said that a good thing about the machine is actually that it reads the same way regardless of time that has passed.

[Mike]: Yeah, and that's also a very nice research direction that I see for the field is that we try to model how human readers actually read. So right now, what we call distant reading isn't really reading. It's just word counting or whatever. It would be interesting to look into ways of training machines that they would actually process texts from more individual perspective. That would be very interesting.

[Federica]: This field of research is highly multidisciplinary, and I'm sure so is your research group in Antwerp. I want to ask you something about the communication within your research group, how hard it is or actually just how does it work, the communication flow, among people with different backgrounds? Do you have to spend a great deal of time explaining the state-of-the-art of your field of expertise, like some concepts you've explained for us here today, just so that together with your colleagues you can formulate the right research questions as a group?

[Mike]: It's interesting, but it's extremely difficult, of course, the work across disciplines because everybody has their own vocabulary, and there's a lot of misinformation generally. So if I look at my specific field, like people from the traditional humanities, so either they underestimate what machines can do today or they overestimate what machines can do, and often there's a lot of negotiation going on there when you try to come up with a good research question, trying to manage expectations, basically, and make people aware of what computers can do nowadays, so nothing less, nothing more. So that's often the negotiation there.

[Federica]: That's interesting. Could you give an example of that?

[Mike]: Yeah, so one example of being misinformed would be, for instance, this handwritten text recognition, so people don't know that it is possible right now already. That is interesting because then they also won't look into the possibilities that this technology already has for them. On the other hand, where people overexpect, you could say, from machines is whenever people try to find an answer to a very interpretive question, so questions like, 'How have Marxist ideals been present in post-war Irish fiction in the 20th century?' So that's a question that is way too vague for a computer to answer. What you can do with a computer is, of course, spot passages in a corpus that mention Marx, for instance, but it isn't going to do the interpretation for you. So that's where people might overreach.

[Federica]: Excuse me, sorry if I interrupt you. You said 'vague'. I think that that question is actually not vague, but simply high-level and sophisticated, so one day a more sophisticated algorithm could actually answer it.

[Mike]: Yes, and perhaps I should have used a different word. It's too high level, indeed, for a computer, so we would have to think much more about sort of the subquestions that have to be answered before you could answer that high-level question with a computer. So it's true it's not necessarily vague, but it is very high level, yeah.

[Federica]: Do you think that artificial intelligence as a field aims to go there?

[Mike]: Sure, yes, I think so. That would be very interesting if a computer could answer questions like that. I think that that is the aim. It's an aim that we will not reach very soon, but yeah, that would be the aim.

[Federica]: Would you define that as a human-level intelligence, or that's the goal, like as good as a human can get or beyond?

[Mike]: I think we have, the game is to go beyond what humans can do. That would be nice.

[Federica]: What's wrong, exactly, with humans, with the human brain? Is it fallible? Would machines be infallible?

[Mike]: Yeah, okay, so now we're reaching a very philosophical discussion, but. . .

[Federica]: Absolutely.

[Mike]: The truth is, what we want to reach in the end in science. Right? That's a very

naive ideal, but I guess that that is the ultimate aim of what we try to do in academia.

[Federica]: I read from the text on your website. 'Recently, truly inspiring pieces have been published in this field (artificial intelligence) about cats, kings, and queens or music, introducing powerful techniques that simply beg to be applied in the humanities.' Please tell us, what's with cats?

[Mike]: Yeah, so it's reference to three papers that I found very inspiring in the past decade, I think, already. And one about cats was a paper from, I think, Stanford and Google where researchers collected like, I think, 100 million images from YouTube, so just random images, and then they had a neural network just look at these images for a couple of days. And what they saw, or when they inspected the network after it had been trained, they saw that specific regions in the network had become sensitive to very specific kinds of images, and there was this one neuron, the cat neuron, that actually had learned to recognize cat images because apparently in these 100 million images that they selected from YouTube, there were simply a lot of cats, so the network told itself to recognize these cats because there were a lot of them. And I thought that that was very inspiring because it's one of these examples of unsupervised learning where the computer didn't know anything about these images, it only got the images, and still it was able to detect these very high-level concepts from these images, like cats or human bodies, etc. So I found that very inspiring.

[Federica]: Define 'neuron'.

[Mike]: A neuron is one very tiny piece in a network that is sort of an information processing unit, you could say. So these neural networks, they consist of stacks of layers, and these layers are composed of individual processing units, you could say, and one unit is one neuron, they call it. And as you get higher up in the layers in this network, you see that these neurons also become sensitive to more high-level concepts, and that's actually why people call it deep learning, because there's many deep layers, many layers of abstraction. And of course, this cat neuron was a neuron in one of these higher-level layers in the network.

[Federica]: Has the system been able to recognize that this is a cat, or it's actually more accurate to say that it simply grouped all these pictures together, like there is a common element, and that's a cat for us?

[Mike]: Yeah, so it wouldn't be able to say 'cat'. It would say, 'This is category 18,' or something, so it wouldn't be able to call a cat a cat, so to speak. Of course, there are techniques of training systems in a way that they would be able to label that kind of information, but here the system only had access to pixels. It didn't have access to words, so that's why it never learned that cats are cats.

[Federica]: What you've been saying in the course of this interview so far triggered a thought in my mind. I didn't think we would talk about machine uprise against humans, like in some sci-fi literature, but the thought that was triggered in my mind is this, that, actually, a machine uprise is just not a risk. It's not on the table as a possibility, no matter how sophisticated machines can become. They can spot cats in pictures — that's pretty impressive — but they don't know what to do with the result. They lack a will, so no matter how sophisticated they become, I think that intention is something that they will actually never possess. Now, a machine could be dangerous, yeah, okay. It's badly programmed or in the one task that it's designed to do, something goes wrong and it hurts a human. Sure. But the machine uprise, speaking really of machines as something that could be injected with enough intelligence to outsmart us and take over — which is completely sci-fi scenario — but I would not have put the possibility off the table completely until a few minutes ago, but you just convinced me with what you've been saying that that's never going to happen.

[Mike]: Exactly, and there was a very interesting paper recently published by, I think, the head of research at Adobe on the question whether machines can create art, and this person was very categorical in saying, 'No,' simply because the machines do not have the intention of creating art, so there's no social function there, and he says art is essentially something social, and that's why we would never credit a machine also with the authorship of art.

[Federica]: Well, some lounge music is automatically composed.

[Mike]: Yeah.

[Federica]: There are algorithms that can generate, well, original music that's, for all intents and purposes, it's called a music composition.

[Mike]: But it's without intention, so I guess it's if it's music that isn't very important, like the kind of music that plays in restaurants, that nobody really listens to, I think that's okay, but like, for many people wouldn't actually like that kind of music simply because of this lack of intentionality, so they want an artist to be behind the music, I guess, somebody trying to express something.

[Federica]: This is a philosophical question, indeed. You know that to test how good an automatic music expressive performance is, it's normally compared to one or more performances by humans, and if the subjects like the automatic performance better, then it means that the model is very good. However, I think that as a subject, as a listener, if you find out that the performance you liked was actually not played by a human, you might have a moment of cognitive dissonance. I see what you're saying.

[Mike]: And I think the admiration would be gone in a second, so if you hear that it was actually a machine who composed that music, I think it would change your aesthetic experience just like that.

[Federica]: So if I understand correctly, you say that I could be sitting at a restaurant and I hear a music and I like it, so I try to Shazam it, and it doesn't work because it's not listed anywhere. I find out then that it's music that was just put together by an algorithm; nobody came up with it intentionally. Do you think that my liking for the music would disappear, would be compromised, would go down?

[Mike]: There is a chance, yeah.

[Federica]: I don't know if this is related, but there's this concept of the uncanny valley which has to do with artificial intelligence, but more with androids. Basically, your positive response to an android goes up as it looks like a human more and more, the better it's designed, the more sophisticated the behavior, the more you are attracted to it. It's interesting, but when it comes so close to a human that you can't actually distinguish it from one or it's just so close, your liking drops. You have a moment of rejection, like it's creepy. And that's called 'uncanny valley'. And I don't know if this is related to what we just talked about in the music, because I like a music, and I like it because it's so good that I believe somebody composed it. When I find out that nobody did intentionally, a machine did it, then I might feel a moment of rejection, whereas maybe if the music is not so good, less, and they tell me, 'You know, computer did this,' I might be impressed. I say, 'Oh, that's remarkable.'

[Mike]: And those are issues that are becoming more relevant already. So I'm using also a lot of text generation in my research right now, and sometimes you have example...

[Federica]: To submit fake papers to conferences?

[Mike]: No, no, no. No. No, no, no. I would never do that.

[Federica]: No, no, no, like to test if the review process is sound, like it's been done before.

[Mike]: Well, there was this very interesting thing on Twitter this morning where there was somebody saying that perhaps we should make that the default: We have conferences, and at least 5
[Federica]: Can you mention some applications of machine learning and deep learning to music?

[Mike]: So these recommendation systems that we already talked about, so a couple of years ago, these were still ignorant of the content of songs, so if you would be recommended a song, that would basically be because one of your friends also listened to that song, for instance, and then they would know, 'Ah, perhaps you'll like this song too.' The thing is that these systems, they wouldn't know which song they were actually talking about. They wouldn't know anything about the tone, or the notes, or the actual music that was in there, the audio wave, so to speak. But now you see that deep learning has also been applied to actually make recommendations on the basis of the actual audio that is in a song. That's sort of a big shift because what you used to have back in the days was the cold start problem, so whenever a new song would be published on Spotify, for instance, you wouldn't be able to recommend it to anyone because nobody would have listened to it. Right? That's a cold start problem, so it's only when people start listening to it that you could actually get data on who this song might be relevant.

[Federica]: Snowball effect.

[Mike]: Yeah.

[Federica]: And the more popular, the more popular.

[Mike]: Exactly, and now with this new technology, you can actually also recommend a very new song to users because you just listen to it with the computer, then you try to model to the actual music taste of your users. So that's very interesting.

[Federica]: This might piss off some artists that claim they don't want to be labeled, their music doesn't really belong in any category, and here comes a machine that analyzes your music parameters and just fits you in this box, in this box, and this box exactly.

[Mike]: Yeah, or the machine might be able to say, 'You're right, you don't fit in this genre, or you have come up with something entirely new.' The machine could be the judge of that, but I guess, yeah, in many cases.

[Federica]: What's the equivalent of that for recommending books? Besides what my friends have read, I might want to analyze the text in a way that learns more about it than the number of function words used in that. So far, we've mostly talked about counting function words. What else can you do with the text to extract information that's useful for recommendation? I would imagine you need to understand the plot, for example.

[Mike]: Exactly, yeah, so it could also be very relevant there, because you can look into what a book is actually about. So if you like, for instance...So Netflix has this category of

series with a strong female protagonist. Right? That's metadata that we typically don't have for a book, right, but if you would have a machine look at the content of the actual book, you would be able to model such phenomena, and on the basis of the knowledge that you distilled, make better recommendations for readers.

[Federica]: Can you tell us about something exciting you're working on right now?

[Mike]: I'm working on fan fiction right now, so that's been a very interesting development in my research. Lately, I discovered fanfiction, which I didn't know before, because I'm teaching a class now on Harry Potter and how we can use digital methods to study the Harry Potter novels, and so apparently there's a lot of fanfiction which is basically imitation Harry Potter that is produced by fans online who, for instance, come up with new stories of Harry Potter or who come up with alternative endings of the stories.

[Federica]: Apocryphal versions.

[Mike]: Yeah, and this is huge, so I have megabytes of imitation Harry Potter now on my harddrive just waiting to be analyzed. It's very interesting to see what people do with this. A lot of it is pornographic, which is both interesting and...

[Federica]: Harry Potter.

[Mike]: ...curious. Yeah, so people are doing very weird things to Harry Potter and other protagonists, but it's interesting because it gives you insight into the kind of reader response that you wouldn't get from traditional inquiry, you could say. So what I'm looking right now is, for instance, the way characters interact and which characters get more popular in the fanfiction and which characters don't reappear in the fanfiction, like something very sad is that Hagrid, who is a very central character in the Harry Potter book series, is very unpopular in the fanfiction for some reason, whereas other characters like Draco Malfoy, they have this surge in popularity. So it's very interesting to study these patterns because it's a way of modeling reader response on a scale that we weren't able to study these phenomena at before.

[Federica]: What is a research question there? It's not obvious to me, how much the audience liked a novel or a character?

[Mike]: So to give an example with this, Hagrid...Do you know Hagrid?

[Federica]: I know nothing about Harry Potter.

[Mike]: Yeah. I also didn't know anything about Harry Potter until like a year ago or so,

but it's been very. . . It's a lot of fun, actually. So for some reason, Hagrid is not popular in the fanfiction. Why? So what is in the Harry Potter books. . .

[Federica]: Excuse me, for those who don't know, who is this character?

[Mike]: Hagrid is a big friendly giant, you could say, who lives and works at the school that Harry Potter goes to, and he is often seen sort of. . . Because you have two groups of people. You have the adults, the teachers, at Hogwarts, which is the school where they go to. Then you have the children or the network around Harry Potter, you could say. And what you see if you study this literary universe using networking techniques, for instance, you see that Hagrid is sort of the hub between the children and the adults, but for some reason he's left out of the fanfiction, and people don't find it interesting to include him. So I'm interested in how that happened, so how did JK Rowling constrain this figure in such a way that people weren't interested in introducing him in the fanfiction?

[Federica]: Very original approach.

[Mike]: It's a lot of fun, yeah. It hasn't been done before, so there's not that much research yet into transformative fiction, as they call it. It could be much bigger in the future, I think.

[Federica]: I'm learning about fanfiction from you right now. Tell me more. Does it exist also for Batman and other things?

[Mike]: Yeah, it's huge. They have it for Marvel. I even found an instance of The Wasteland fanfiction, which is a very hard poem by T.S. Eliot, and somebody produced fanfiction of it, which is insane. Very cool. It's also, it gets very interesting when people start fusing fandoms, as they call it, so you have the fandom of Harry Potter, but you also have the Twilight fandom, for instance, and then people might for instance take Harry Potter and have a sex scene with Bella from Twilight, for instance. It's very interesting, also, there, the kind of patterns that you see in how people fuse fandoms.

[Federica]: These cultural phenomena seem to call for sociological analysis. What do you do as computational text analyst here?

[Mike]: I don't do social science stuff. I simply. . . I haven't done anything that is relevant for social science, I think.

[Federica]: Okay, no social sciences there. So you have to explain to me what counting function words have to do with this character having sex with this other character.

[Mike]: So there, you forget about the function words. In this kind of analysis, you actually, you go for the content words, which are very interesting, and it's also very interesting to look at the kind of topics that emerge in the fanfiction that you don't see in the canon or the authentic stories of Harry Potter. And you see, for instance, that if. . . Silly stuff sometimes. Vehicles. So vehicles are much more important in the fanfiction than in the Harry Potter books, so there you can see that's actually the real world bleeding into the fanfiction. Military stuff, also. That's very interesting. Yeah.

[Federica]: Mobile devices.

[Mike]: Yeah, so a different kind of media that also bleed into the fanfiction that you don't see.

[Federica]: Music.

[Mike]: I didn't see that. Yeah, could be. And now I'm focusing on Harry Potter together with two colleagues. We're doing a book on this. But there's many other fandoms that are equally interesting, like, for instance, the Lord of the Rings fanfiction, also huge. And the nice thing is that you have it in many languages also, so it's huge in English, but it's even more huge in Spanish. There's lots of Spanish Harry Potter out there as well, Russian fanfiction, and that's nice because you can actually study these fandoms across languages and cultures, lots of Asian fanfiction as well. That's very interesting.

[Federica]: Just because I'm not familiar with this, let me ask you something more. Where do you find this material? That is, is it published? Is this random people writing novels, short novels, or blog articles and putting them out there how?

[Mike]: So there are these community platforms where people author it, like fanfiction.net or Archive of Our Own, and it varies hugely. So they have these very short pieces, I think they call them drabbles, of only 500 words, but there's people who do book-length fanfiction, and some of it is just huge. There's this very famous fanfiction author, Norman Lippert, and he was reading Harry Potter to his children, and when the series ended, the children were sad, and he said, 'Okay, I'm going to do my own Harry Potter series but focusing on James Potter, the son of Harry Potter,' and he basically has book-length continuations of the Harry Potter saga just for his children, and he also publishes it online, but these are serious books, very long books. So people spend a lot of time on this. That's actually also amazing. The kind of energy that people invest into fanfiction is. . . Yeah.

[Federica]: That's the cutest thing I've ever heard.

[Mike]: [chuckles] Yeah, yeah. It's very cute.

[Federica]: Wow, that's the professional version of, 'Dad, tell me a story.' 'I'm going to write a book.'

[Mike]: Yeah. It's actually, it's pretty nice. He did a very nice job. People are reading it, and JK Rowling actually is okay with it, so she said, 'If you do Harry Potter fanfiction, you can't earn money with it,' which is fair, I think. 'You can't sell it, and it can't be pornographic or racist.'

[Federica]: Which it is, pornographic.

[Mike]: Not the James Potter. The James Potter is actually what we call affirmative fanfiction which basically continues in the same line and style as JK Rowling.

[Federica]: Okay, so it's like Rossella, the sequel of Gone with the Wind.

[Mike]: Something like that, yeah, and people actually like it.

[Federica]: So thank you very much for this conversation. I have learned something new, and I hope that some of our listeners have learned something new too. As I try to unsee some pornographic images of fictional characters, I would like to thank you very much for being on Technoculture.

[Mike]: Thanks for having me. It was a lot of fun talking to you. Thank you.

[Federica]: Thank you for listening to Technoculture. Check out more episodes at technoculture-podcast.com, or visit our Facebook page @technoculturepodcast and our Twitter account, hashtag Technoculturepodcast.