00:00:00

Gea: Welcome to "Never Write Again?", a podcast about artificial intelligence and writing skills. Thanks for listening. My name is Gea Dreschler. I am assistant professor of English linguistics and academic director of the Academic Language Program, or the ALP, part of the humanities faculty. In each episode of this podcast, I talk to someone who is involved in some way with the topic of ChatGPT and AI tools. The conversations are about how these tools actually work, what they can (and cannot) do, and most importantly, what we should do with them in education. Ban them, test students differently, don't let students write any more? Together with the guests, I examine the subject from all angles.

00:00:54

Gea: Today's guest is Piek Vossen, Professor of Computational Lexicology in the Faculty of Humanities here at the VU. Piek is going to tell us mostly about the technical side of ChatGPT. So what are the language models behind ChatGPT? Why is ChatGPT so popular? And do AI detectors actually work?

00:01:17

Gea: Welcome Piek, nice to have you! I first ask each guest a few questions about their own writing process or writing skills. Are you a Mozart or a Beethoven, so do you come up with the whole composition and then write it down once, or do you write in first draft and keep endlessly polishing it until once you have something which is a final version?

00:01:40

Piek: Yes, it depends on what purpose I'm writing for, I guess. So for scientific articles it is kind of more planned, structured. But if I just write an email, or a letter or a poem or something else or a story, then well, it's just: one word evokes another, one image evokes another. Yeah, then I don't have an idea beforehand.

00:02:04

Gea: What is the best place to write?

00:02:07

Piek: Yes, I prefer actually on a laptop anyway, sometimes, by the way, I just use pen and paper again.

00:02:12
Gea: Do you listen to music when you are writing, or no?

00:02:16
Piek: No, no.

00:02:17
Gea: So no music, absolute silence?

00:02:18
Piek: No, I don't do that anymore, I used to. Yeah.

00:02:21
Gea: And what tools do you use yourself when you are writing?

00:02:25
Piek: But yeah, I think the spelling and grammar checkers, you can hardly turn those off so yeah, I do use those a lot. I don't use tools that generate texts, or, like, I think half the time it's also wrong suggestions, which I then have to correct again because they changed something, or a wrong word. Or then, then I read it back and think my god have they filled in something else again which I didn't write.

00:02:50
Gea: You want to be in control of your own text.

00:02:53
Piek: Yes. I don't really need something to make up words or pieces of text for me. Well, if it, especially if I'm writing in English, if I don't know exactly how to say something in English, then it's helpful. But then it's really more about, for example, the right verb, preposition, verb combinations/ constructions and not so much lexically frankly. And also if you're writing about your work, then you know all the terms, you actually know all the words. So it's really about formulations then, and then in English. Yeah, I'm not a native speaker, so then I do use those kinds of suggestions and so on.

00:03:30
Gea: And we're going to talk about this at length later. But: AI in education, should we ban it or embrace it?

00:03:38
Piek: Well, yes, prohibition, you're not going to succeed I don't think. I wouldn't know how you're going to manage that. You can't observe students permanently, you don't want that at all. In my opinion, that's not a pleasant atmosphere, that's not how we should be in education together. Besides, it's just part of the world. So you also have to accept that it is there and take care of it together: How can we use it in the learning process? Because that's it, right, a learning process, not just tests: Learning is not just testing, but testing is necessary for something, apparently. But some forms of testing, they don't work anymore. You can forget about them, I think, in the future.

00:04:21
Gea: We are going to talk more about this later. Now, maybe for people who are not into this every day: what is Computational linguistics, or lexicology I should say, sorry.

Piek: Yes, even more difficult, Lexicology yes so Lexicology that is the science that tries to build knowledge of words. And if you put that knowledge into a computer, you could already call that computational lexicology. But then that would be a database of words, so then first you need to know: what are the words of the language? Nobody knows that. Sometimes expressions are also words, also combinations like "strong coffee": is that a word or should that be eh, "Strong coffee". It could also have been "bitter coffee", but we use "strong coffee" now so those kinds of collocation-like things might also be words. So to put or not to put a space between words, that's all convention. And if you go on like that then we're talking about millions of millions of possible words. And if you want to learn a language, then you know those combinations too. Someone who learns a language, they make typical mistakes, that- they know the words per se, but not the right combination of words. Well, if you want to register that, already that's a huge task and ideally you want to do that by linking to how those words are actually used. Written texts, but also spoken language in interactions between people, all that data we collect of course. So texts, you find on the Internet and interactions, you can record those and then you can say these could be the words of language. But here you see how people use those words, and these people use them this way, and those these people use them this way. Those are not used, or very often or not a lot. Everybody knows a lot of words, but only uses a small part of them. So how does that work exactly? Well, so you actually go on and on and on, and then successively you get to the point where you say: but wait a minute, you can also make that computer use those words itself. So then you make a system that imitates or models human behavior, of how people use words, well, then we get a bit in the direction of the current AI-like applications which then involve a computer that has learned to use words, just as we use them.

Gea: When did that development actually begin? Within linguistics? because by themselves lexicologists, of course, we've had them for much longer. That goes back much longer. But that that move toward more AI application, is that really something from the last five years, ten years?

Piek: No, no, no, I remember I was a master's student at the UvA with Hugo Brandt Corstius and he came with enormous boxes and in them were the proofs of the dictionary "contemporary Dutch". I think he got it from Van Dale. And he also had a digital version. So then we said: Hey, but wait a minute, now we can try to discover in that digital version on a large scale all kinds of patterns and such and so. That was I think in the early 1980s or so. At that time they were also working in Cambridge and also at the UvA on the Longman Dictionary of Contemporary English, which was one of the first dictionaries available digitally, in English, that could be used for research. They were also working on it in America. So a number of groups that started doing that and started imitating that from each other, and then all at once -- where you used to have in linguistics, research about a few examples, a typical article where somebody uses their contrastive analysis about one word, this or that. Well, then you have five pieces of text, maybe ten words. Now in an article being studied, we could look at and analyze 100000 of them. So that went at once, from a scale of a handful of examples to yes, almost more words than a human being knows at all. And so in

addition to that you had corpus linguistics, which was a bit older, from the 1950s, I think. Yes, they were going to collect texts first and actually, I think in the 1980s, then that combination also came about and it also became empirical. So actually that whole empirical side that came with corpus linguistics but also with that Computational lexicology in the 1980s, where you are now able to start analyzing large amounts of texts, large amounts of words on a large scale. Yes, and so then it turns out that those theories are actually wrong, completely.

00:08:54
Gea: Yes, and of course that's very different from researchers who might try to create a dictionary on their own based on well, manuscripts or examples or bills with examples of words. Now, with that much data, of course, you can say a lot more about what is really being used.

00:09:09
Piek: Yes, and so then you get dictionaries in which they illustrate the information in the dictionary using real examples from the corpus but that's still not based on empirical analysis. And actually it never got there because those lexicographers -- they're not lexicologists and nor are all lexicologists really doing this -- The tools understand how from the data you also have to derive your models right away and that's where those worlds kind of diverged. Actually, the whole machine learning era, which really started a little bit in the 1990s and is still going on, people working on that are not linguists, they are really computer science-like mathematicians, who made all those machine learning models, they still look at language in a different way. They followed their own path and they ended up making very powerful models in which people's language behavior is imitated by a computer. And you can do that by what they call "supervised learning." So then there are still a lot of people who say to the computer, yes, but if you look at these words or look at these phrases, this interpretation goes with that. It can be a sentiment, but it can also be all kinds of other information. And then you can teach a computer to make the same association. And then you give a random text and then that interpretation can stick to it. Why and how it does that is a whole other story. Whether that is the same as we humans do it is also another story; Probably very different. And that took on a life of its own, and that actually became more successful than anything that was done before it. So if we were to write a grammar in the classical form, with rules: This and that and that are the word sequences of Dutch, these are the rules that go with it. Here you have a, a dictionary or lexicon with the words with their properties. Well, go ahead and try to analyze any text with that, it will do it correctly to a certain extent and wrong for the rest. And if you let a machine learning program infer that in an indirect way, it does it much better than any grammar can. So yes, so then you stand with your theory, well really, I was trained in the classical linguistic way of thinking. And yes, you just have to admit that that whole stupid computer science behind approach to a language ultimately works better, is better able to analyze a text than all that theory. And then you've lost.

00:11:42
Gea: Yes, we did a lot of work anyway, which maybe in this case a computer can indeed do better. Now you've already explained something about machine learning but if we look at ChatGPT, because of course that is what we are talking about all the time now: How does that work? What's behind that?

00:12:00

Piek: Yes, ChatGPT is actually a a product of yet another form of machine learning, which is what I just explained and that is then, as it's called, unsupervised learning. With supervised learning, then you give an example, so to speak, and stick a label to it. And then that computer is going to associate that label with the properties of your example. And what those properties are, you can think of all kinds of things and then they also.... anything you can think of, they've already thought of oh maybe we can also represent that as a feature of that example. So that's been well, something like ten, 20 years of development in which they figured out: How can I represent that text in the right way, so that that computer can generalize about that text and apply that to another text. In unsupervised learning, you don't say how to interpret a text at all. No, you're not going to give a few and I'm talking about maybe 10000 or 100000 examples, which have been the numbers so far. No, you give a computer millions, billions of texts. And all you specify as a command is: Well, if you're given a series of words in a certain order and I would occasionally mask a word, "mask" is what it's called, then I ask you, computer, what do you think is the right word that was there? Well, then I, I don't have to do anything, because out of all that data I just grab all the all the phrases that I can find, and I just paste off some words every now and then and I know what the right word was and a random word is the wrong word. Well, what that unsupervised learning is going to do is you're going to "represent" the other words that are in that text. And what that is, I'll tell you in a moment, that's another story. Those words they are going to give them a representation that allows them to predict the correct word. Hey, so I woke up this morning and I made a.

00:13:58

Gea: sandwich?

00:14:00

Piek: Yeah, or a cup of coffee or or a smoothie, or actually we're cliché, animals, right, so we fill out that list very easily, very quickly. And that computer is going to learn that, too. By just looking at: What have all those Dutch people filled in on that little space that is masked there? It does the same thing in English, and French and German and any other language. And we are all about the same cliché animals all over the world. And so we're all going to fill in the same kinds of things. Then what happens? Getting up, I had said. I get up this morning and first thing I did, I "made one," "make," it's all words. Those now get a kind of representation, which enables them to predict the next word in this sentence, or a word back, left and right. But that representation is now going to replace the word. So that means that the representation of make or get up for a moment is going to look like other words that we would also use in that place, that mean more or less the same thing to predict such a word again. And in the first example I said yes in machine learning, they take their sentence and then they start to see those words as properties. And then they have to make an association. And if I have a new text, it's going to see if it has the same properties. And if it does, then it's going to predict the same label. But now, in unsupervised learning, it has learned to extract from those words and learn some very bizarre representation. That represents all the words that have a more or less similar meaning or usage, in terms of meaning, but also in terms of grammatical linguistic properties in that place, in that sense. So, based on that representation that model, if I go back to supervised learning now. There's a set of words that substitute

by that representation and that generalize over anything that could be in there. So instead of now having an association with my label, with one specific example, I have generalized, as it were, that representation to all the possible texts that you can have, that more or less represent this and then suddenly that learning becomes much, much, much more efficient.

00:16:16
Gea: And then can you actually say that, such a program and ChatGPT, that it doesn't understand what it does? That it's all based on frequency and predictions about combinations that are most common, and based on that an expectation of what follows this word or what comes in this order? Is it all kind of flat actually in terms of structure, so not hierarchical or some kind of sentence analysis, but really based on words that you see?

00:16:45
Piek: These models, those are then, say transformer models and GPTis one of them and you have many more. Bert is another one. You have all kinds of variants of those models, you know, and OpenAI was actually not the first to do this and are not the only ones to do this. Those that are pretty much, they are in themselves, they are large, let me put it that way, not complex.

00:17:09
Gea: You just said that those that language models, those that ChatGPT is built on, are not actually new. There may be parts of ChatGPT so packaging with the chatbot which is new. But does it surprise you how much attention there is now for ChatGPT and what it can do? Or where does that actually come from? Why are we so enamored with it because of AI?

00:17:29
Piek: Well, I think, the reason is because OpenAI has built some kind of layer. The model underneath it is GPT. And that's one of those transformer models we were just talking about.

00:17:43
Gea: I have already forgotten the abbreviation, well transformer but yes.

00:17:47
Piek: Generative Pre-trained Transformer. Generative means that in that what I said that "masks," right, so that a piece is taken off and it has to predict that. If it's generative, it just masks off the next word, the sequel. And if you have Bert, which is bi-directional, then you can tape to left and right things because of that Bert actually has a better representation of the sentence, because sometimes information at the end is also important to represent something. But yes, for those generative-like applications, GPT is very good, so it really just knows: if you give this piece to me, I know exactly how it continues.

00:18:29
Gea: He can finish it and he doesn't have to have the end of the sentence.

00:18:32
Piek: And then it does a one word completion and with that word the next and the next: So they can generate a very long text. And the generative property of this model, that what it's very good at then, they've exploited that by building all kinds of

behaviors above that. Because at the bottom of that model it can actually generate anything, the most bland, nonsense, garbage and and all the ugliness you can think of. The model can generate anything, so generated also that, so I have no control over that, but so on top of that they just built what they call intentions, which are those chat-like features. And they focused on summarization, translation, question-answer it and some of those.... In itself not that much at all. If you say of I have 30 strawberries in my hand, how heavy is that? Then, that will generate something.

00:19:22
Gea: He will say something.

00:19:24
Piek: Always.

00:19:24
Gea: Whether it is true then, of course, again remains to be seen.

00:19:26
Piek: Yes and the more absurd the input is, the more absurd the output. But often people ask a lot of standard, simple, basic things.

00:19:35
Gea: But he does say now and then,or he/she/it.... "I don't know this. I'm a model, so I can't do this... I don't know what the weather will be today, because I have no info about it."

00:19:45
Piek: Yeah, that's that that behavior that they create on top of it, and they actually do that through supervised training again, so to show if you, if you get this in you have to react like this. Underlying you have GPT and that generates all kinds of things, and they just show 20 variants of GPT, they present them, and then they say, yes, that's the good one, that's less good and that's even less good and this is completely wrong, and so on and so forth. So the system doesn't know exactly why well one is better than the other, but it's going to imitate that behavior, it's going to imitate it eventually, and there's also in there of if you're asked this and this, you shouldn't answer.

00:20:22
Gea: There's so there's actually still quite a lot of human work there, including that training.

00:20:26
Piek: Yes, so they made that very efficient. That's OpenAI's claim, that that's their contribution, because they didn't come up with the GPT itself. They did engineer it, that very well. But others have actually figured that out but they say that they superimpose that behavior in a very efficient way and then you also understand, because that behavior is so human and so easy that you can ask something in your own words to a system and then get an answer back that has made it a hype. So they've actually... It's just become very approachable to use this complex technology that you might have had to get a master's degree for before, or something. And now everybody can use it, because you don't have to go directly to GPT anymore. No, you just simply ask your questions to that thing sitting on top of it, and because of

that now the general public can use this technology with all its good and bad features.

00:21:21

Gea: Yes, because do you think we actually... Do we as the general public have any use for that? Or is it actually having some kind of trick? Thingy, look what fun and in a month we will have forgotten about it.

00:21:31

Piek: Well, some things are really useful and helpful, so I would definitely have it make summaries of all kinds of things, for example. This and this and this, just summarize it for me. I think that could be very good. So if the information, that's what they call the prompt, that's the input you give and if you don't give input, you don't get output. So you have to give something and those are your instructions. Well, if, if if you ask the system to do something with that information and it doesn't need any information other than what it says, it can often interpret that very sensibly and associate something with that. But if you have to get something from the outside, you don't know where it gets that from. It is hidden somewhere, in everything he has seen and read, and then he hallucinates.

00:22:25

Gea: I've heard that term many times now, what does that hallucination mean?

00:22:29

Piek: Well, you have to imagine that as it works that underlying model, and that's just hard to have control over that, and that's what they're trying to do with those functions. You have a series of words there. The next word, what that's going to generate there is actually a representation of the next word. And then that one is going to look at it again we have the word lexicon/vocabulary which are all the words that that one has ever seen in its representation. He has made a kind of list of them. And every word also has a representation. And he's just going to see how high a score each word gets as possible on what I expect, what should go there. So the word that ends up being chosen there by the system there, then it can think: yes, this is really going to be the very best of all. Unfortunately in our general world model it has nothing at all to do with what we had specifically asked for in our input. But based on everything he has read before, I really do think that this is the best word that fits there. And then he's just going to put that in there.

00:23:37

Gea: And does that also explain that who can produce, can deliver a lot, which you then think: Yes, it sounds, it sounds perfect, it looks very convincing, but content.... So a made-up biography of someone or a made-up story about where a certain last name comes from, or.

00:23:57

Piek: Yes, but that's because it's say with that an unsupervised learning has, he made that generalization from what that read. So there's no index of the facts that he's read. It's not like oh, yes, I've read that and there and there so many people live and they do this and that and that. And with the knowledge itself, factual knowledge, which we then still somehow kind of remember. When we learn things and read things he has, there has only sort of conceptual layer there, but the factual facts

directly has not indexed that. So he is actually going to generate something that conceptually fits with that based on everything he has read.

00:24:41
Piek: If he would just read a lot of stories about people with a left leg first boarding a tram as -- no idea if that's the case, I don't know – then he would think, yes, that that's it, that's what I've actually clearly deduced, all that stuff, so I think that's what's going to be with this. Then he's just going to add that.

00:24:58
Gea: Yes, so so in that sense there's no filter on the input and and what it delivers.

00:25:04
Piek: No, the only filter is the control of prompt, as I mentioned. So that's technically called the prompt these days, and that's why you have a new profession now, the prompt engineer, that's the one who tries to keep as much control as possible over exactly what it's going to generate through that prompt in a very clever, very explicit way. Very clear you have to be to that system like a little child. You must do this and you must not do that and you must do it this way, because otherwise it can make something up just like that.

00:25:38
Gea: Yes, so this is already a very clear example, I think, of the limitations of the system.

00:25:44
Piek: Yes, but it depends on the goal. If you are interested in the creative, yes then fantastic, let's make it up all, nice, doesn't matter.

00:25:53
Gea: Maybe for a scientific article it is not.

00:25:56
Piek: No, I wouldn't have it written as a scholarly article, nor a journalistic article and actually, students, not your essay anyway.

00:26:03
Gea: No-yes, that's a nice bridge, of course, as teachers, we've all come across this, tried it out, and maybe seen student essays where we've thought, Hm, did the student write this themselves. What should we do with this in education? But maybe also, what can we do with it?

00:26:22
Piek: Yes, I actually think then: what can we do with it. Yeah, what there should be no more is having students write essays and just say okay. I'm now going to evaluate the essay per se, but.

00:26:32
Gea: Should that teach people to write at all?

00:26:35
Piek: Well, they maybe learn to write using these tools so how do I write an article that I, that I support myself while using these tools? It's kind of the question of what it, what was the purpose of the of the article. Look, if the purpose is testing whether you are able to write a scholarly article, really in terms of quality and style of writing, then

yes, that may be something that we don't do anymore, just as much as we don't test manuscripts anymore either.

00:27:09
Gea: Well, there are people who go back to handwritten exams.

00:27:13
Piek: Yes, that's one way, but I would venture that that's also going to be a ChatGPT there, generating handwriting, it can probably already do that. I wouldn't even be surprised by that actually. But yes and oral tests anyway, yes. But you're quite welcome to have an article generated as far as I'm concerned. But it seems very difficult to me when it comes to the content of what's behind the article, A: whether ChatGPT is capable, is capable of doing that correctly and B: do I think you may have to do more than just turn that in and review it. Maybe you should have an oral test.

00:27:47
Gea: What you're saying is: you have to think very carefully about what is actually the purpose of the assignment you're giving students? And then how do you test that? But but kind of a standard assignment write an essay or or write a summary, we're not going to do that anymore. Is it actually, because there are these ChatGPT detector or AI detectors, do they work? Can we somehow find out that a text is from an AI?

00:28:15
Piek: Yes, but that's... ChatGPT/OpenAI have such a service themselves, to detect if something is AI-generated, and they are going to use that data again to take care of making it even harder to distinguish, even more natural.

00:28:31
Gea: Can you see it as a human being?

00:28:33
Piek: Well, there are some things that I know that, you can check for that at the moment and one of the things is thank God most of the data is English and there's very little Dutch for example, and we know that it uses English-like, grammatical structures for Dutch, which we just don't use like that.

00:28:54
Gea: Yes.

00:28:55
Piek: But yes, then you just have to come across by chance.

00:28:57
Gea: Yes, what do you actually think of the output anyway because, is it.... does it write well?

00:29:06
Piek: I think it's a very annoying style that that uses. A bit of that hyper correct, always friendly, helpful AI.... that's going to kind of bother you the more you use it, frankly.

00:29:22
Gea: He's not very concise either.

00:29:24

Piek: Brevity is indeed a problem, and even if we say very clearly that it has to be concise; you can't use more than so many sentences or words, it still occasionally goes wrong there. Yes, and it is true, I think, that it is one of the most impressive things, well, that if you know, very well all kinds of find the right style with certain kinds of... Let me say, for instance, if you were to describe a certain state of mind, he knows how to find the right style for it. So the stylistic talent is quite great and you can tap into that or steer it to bring that out.

00:30:10

Gea: So you're basically saying of: we have to learn to cooperate with with the AI tools, forbidding doesn't make much sense; We have to learn to work with it somehow anyway. Also the students: Should we start giving them some kind of extra training. Like, how do you recognize an AI?

00:30:24

Piek: I definitely think students need to know what the possible errors and problems are anyway, but also what the capabilities of the tool are. The issue of whether something is generated by an AI or by a human: I think there's a much bigger societal problem than that being a problem for students. Honestly.

00:30:44

Gea: Yes, in a sense, of course, we can still direct a little bit what they use and don't use, or or what they do with it, what they use it for.

00:30:52

Piek: Yes, some students will be inspired to cheat with that later and but I assume most will not.

00:31:00

Gea: There are also students who say, I can do it much better than ChatGPT, so why ask such a program to do it?

00:31:06

Piek: Yes, well, I think we should also have ethical discussions, but that also applies to when you, for example, if you teach students how computer viruses work, then the students also know how to make a computer virus. Right?

00:31:22

Gea: Looking ahead for a moment: Ten years from now, will students still be writing papers or will we only do oral exams that?

00:31:34

Piek: Yes, well, that's kind of the output; the form of the output will certainly be different. Yeah, yeah, I don't know, maybe it's not. Maybe it's more the teacher replaced by AI frankly, so maybe the AI is testing the student, whether the student conceptually has the knowledge and has the understanding.

00:31:57

Gea: It seems that you can also make very good rubric with AI.

00:32:02

Piek: I certainly believe that. Yes, yes, yes, and I think.... That's obviously the most important thing, what you want. You want students, for example, to have the right mix

of both a piece of knowledge that you have to have, because you can't just look everything up, you have to have a piece of knowledge yourself and have the understanding, but also to be able to be able to make sharp analyses precisely, to have creative solving ability, for when something arises for which there is no answer. So suppose, nobody has the answer to it. Then how do we find the answer? And I think that's what you have to learn. Ultimately. So problem solving ability is the most important thing, so not where do I find the answer, but how do I find the answer myself?

00:32:45
Gea: Yes, so making learning objectives very clear, not just writing the text, but what is involved.

00:32:52
Gea: And what are the developments in the field of AI that you will be keeping a particular eye on in the coming months?

00:32:58
Piek: Well, over the next few weeks… I have to say: it goes almost by the week. Then again you get, "there's another model," you can't read up on it.

00:33:12
Gea: Maybe ask GPT to make summaries after all.

00:33:13
Piek: Well especially, you should especially not ask that ChatGPT to tell something about themselves because then they get lyrical about themselves. They have apparently finessed on that as well.

00:33:29
Gea: Will this continue at this rate, do you think?

00:33:31
Piek: I have no idea, but I I hope not.

00:33:37
Gea: We are not done with it, for now.

00:33:40
Piek: No, no.

00:33:42
Gea: Thank you for this conversation.

00:33:44
Piek: You're welcome.

00:33:44
Gea: I was able to create this podcast thanks to a grant from the Center for Teaching and Learning's BKO SKO alumni fund.

00:34:01
Gea: Also contributing to this podcast were: Jens Branum, Abby Gambrel, Joost Canters, Megan van der Vorst and a host of AI tools.