

Into the Headlines_ Episode 2: Writing Robots

Speaker 1: (00.08) Science fiction has planted the fear that machines will one day replace us. It starts with some of our more labour-intensive jobs being taken over and inevitably slides towards a robotic uprising and the human race enslaved by a dominant species of super intelligent, sentient and virtually indestructible lifeforms. Artificial intelligence took another leap forward recently with the launch of OpenAI's ChatGPT, a next generation conversational AI chatbot, sparking debate among writers, journalists and other copy-based content-generators that the end, for them at least, may be nigh. But is it? From the University of Aberdeen, I'm Laura Grant, welcome to Into the Headlines.

Intro music: (00.51)

Speaker 1: (01.07) Episode 2 – Writing Robots. Today I'm joined by Dr Georgios Leontidis, Reader in Machine Learning and Director of the Interdisciplinary Centre for Data and Artificial Intelligence at the University; and award-winning Scots writer, illustrator, singer/songwriter, storyteller and honorary teaching fellow at the School of Language, Literature, Music and Visual Culture – Sheena Blackhall. Welcome both.

Speaker 2: (01.27) Nice to be here.

Speaker 1: (01.28) Georgios, language models in the form of chatbots aren't new. What have they looked like to date and what have we used them for?

Speaker 2: (01.36) That's a very good question, yeah, you are perfectly right. We've seen this type of system for a while now, they have been developed and deployed in different settings which we are more conversational types of systems where you ask a question and you get a very simple response back and some of them are also able to do some very tasks for example if you ask what the weather will look like tomorrow and then it gives a very basic answer. But actually to go back to the very roots of this system we have to look back at the Sixties so when we had the first system, it was called ELIZA which was a very basic first system of this type of models but as you can imagine back then everything was very simplistic, you were writing something very well structured and then the responses were also very well structured so it was pretty refined. Now fifty, sixty odd years later we've seen a huge development in systems and that has culminated in ChatGPT which is a most popular system nowadays and for us that are in this field, we have seen also many technological developments that have led to having such systems today. I would say there is not one factor that is more important than others but if I had to choose one that would be the computational resources that we have available nowadays that are enabling us to scale the systems and actually come up with systems like ChatGPT, for instance.

Speaker 1: (03.06) ChatGPT is being called a step-change. Cards on the table, I'm not the world's most tech-savvy person so, in fairly simple terms, what does it do and how does it do it?

Speaker 2: (03.18) To be able to answer your question I think we have to look at the predecessors of ChatGPT, the likes of Alexa and Siri that have been with us for a number of years and I think now we are familiar with these systems in a sense that we know that we have to ask them something, so we have to structure a sentence and ask them something, and we expect a response back from them, either like a verbal response so they are responding with dialogue or with some sentences also or with actions, so they may be doing something and responding with a sentence or a paragraph or whatever that is. But then, we

have now the ChatGPT which is such a system but plus plus, which means it has managed to be trained with a huge amount of data that is available on the internet and I think we speak about more than three hundred billion words have been used to train ChatGPT and that amounts to more than five hundred gigabytes of data. Now if you think three hundred billion words, it's a huge amount of words and then, so this is kind of the data that the model has been trained with and then we have the model itself which, for me, means the architecture of the model, like the heart of the model, which is just a neural network, it's a specific family of neural networks called transformer models, and that model itself has a billion parameters and we can think of its parameters in a sense like a neuron, like an artificial neuron. Now why I mentioned before computational resources, the reason for that is that action to train such a model, so to train ChatGPT we speak about humungous resources and it is very hard for anyone to calculate exactly how much ChatGPT training has costed but that is in the low millions range, so it could be two, three million dollars that training such a model has costed and that means that, for us even to be able to retrain this model, to develop a competitor of ChatGPT, we would maintain a huge amount of cost, it's very hard to do that unless you are a large corporation that has such investment to put into these models. So I think it is a combination of factors, but the amount of data and the availability of additional resources has actually accelerated research in this area to be able to come up with ChatGPT and also the successors of ChatGPT that are going to come in a few months' time.

Speaker 1: (05.56) There's a huge amount of buzz around it. I think I read somewhere that a hundred million people have used it in the first two months after its launch. It feels potentially like a huge traction of those people have been journalists and people like me who write for a living and want to see how good a job it does in terms of what it's generating, and also a lot of people just having fun. Why is there so much excitement about it, who is currently using – properly using - this technology and what sort of things is it being used for?

Speaker 2: (06.26) If I go back to the first part of your question I think that one of the main reasons that we've seen a huge amount of users using this type of system is the fact that the company behind the development of this model, OpenAI, they decided to make this model available by a very simple browser URL link, you know, through an API how we call it, so anyone and everyone can, as we speak, go and write something and use this model which is, I wouldn't call it unprecedented but it's a bold move from OpenAI to have it open source and release this model so users can use it, so I think the main reason is the easiest behind being able to sue this model. Now, going back to other parts of your question, I think that, as I mentioned before also, given the fact that this model has been trained with so much data, inevitably they will be kind of performing well in some cases and bad in some other cases, that's inevitable and the reason behind that is that we have to understand that these models are limited in their capacity to actually reason or think like us humans, so they are more like parrots, to some extent, so they learn some and try to find patterns in the data and then when you ask them something, they try to see what combination of words and previous information that they have seen in their training data can be used to generate some new sentences that look new but actually they might be a combination of other data that this model has seen while training. Now, I would say that, depending on how much they try to stress test this models they will exhibit different levels of intelligence, if I can put it like that, which means that some contents might be easier to imitate a human response. So then if you then interact with this system, you might say okay then, this is a very simplistic response for a simplistic question, so yeah, I cannot be certain whether that is ChatGPT or a human giving me the response. But they also suffer from a problem called hallucinations, so that means that sometimes they are making up stuff and they are providing sources that don't exist and that's a big problem at the moment, so that's why if you go to the other part of your

question about who are the users nowadays, I would say, especially because I am a frequent Twitter user, I think that pretty much everyone is using ChatGPT, from you know, I've seen cases where people have tried to see how good in a medical domain the systems could be, you alluded before about journalists using these systems or academics, anyone has tried to use that I think nowadays like from all different disciplines, but I don't see that process phasing down, so I see that we are just at the very beginning of using these systems and one of the reasons is that these systems are constantly updated, so it's not like you have a static system where, it's there, you can interact with it and get bored of using it and then just forget about that. These models are constantly updating so OpenAI, the company that has developed this model is releasing new versions of this model I don't know every week, every other week, so you have new versions of that, and that will culminate with a new generation of this system in the coming months. So then you are going to see even better and better models coming out and then what we thought that was impossible before in terms of asking and getting a decent answer back I that's going to get better and better as we move towards the future.

Speaker 1: (10.01) So I did share a press release with you that we'd asked it to write, and it was a fairly basic press release that we tasked it with, but it did any alright job, which was a little bit worrying. It had the structure there, it understood kind of how to pitch a top line and how to craft a quote from a spokesperson – but I'm wondering what you think the influence ChatGPT or whatever comes after it could have on our lives long term, what sort of jobs might actually change as a result of its existence?

Speaker 2 (10.32) That's a very good question but given I'm sharing the stage with someone like Sheena I think I'm the last person to comment on anything about linguistics or structuring the English language...but I think that speaking mostly from a technology perspective, I think as with any other technology that we have seen over the past few decades I don't see these systems replacing any kinds of professions or jobs, I can see them being a useful auxiliary system that it can provide you with some baseline content and then you have to put your own creativity and more artistic flavours on that but I don't see that being good enough to say, yes just trust them to deliver something without having any human input. Also it could be that these systems could be deployed in a way where it can allow you some more space to do something else. You know, all of us we are living some very busy lives, we could do with some sort of help sometimes so I see them being for us more of a helping hand in certain cases but I wouldn't say that it's going to be replacing anyone in the very near future or in the near-term future but who knows what is going to happen in the long term, right!

Speaker 1 (11.47) Well Sheena, let me draw you in here. You are a highly accomplished writer. What do you do when you approach a piece of writing? Do you write from the heart, from the head, a bit of both?

Speaker 3 (12.01) Well any piece of personal writing is filtered through experience. So initially the language that I write it in can vary because I may choose to write in what's known as the 'Mither tongue' or Doric, which I was brought up speaking. That's limited in its vocabulary but leaving that aside its generally said that if you write in Scots that's from the heart and if you write in English its cerebral, its coming from, you know, the pure intellect. I wouldn't agree with that entirely because as you grow older you acquire more language. Again people say well youngsters don't write much in Scots, well but they are only acquiring language. I'm 76 this year so I've acquired a lot of language down that time, but I wouldn't have been rattling off reams of Doric when I was five. I was hearing a lot of songs, Scots songs, mainly Burns or traditional Scottish songs because it was a musical house I grew up

in, and they also spoke a lot of rhymes, traditional rhymes, so that came through as well. The other thing that's helped me with writing and thinking about writing is that for 20 years I attended one week a year a Buddhist meditation centre in Balquhider, run by two Buddhist order members but they were also published poets and they ran the creative writing side of it so you had meditation, alongside production of work and I found that extremely useful. Of course, quite a lot of that was in silence so they would seed the meditation, it was the opposite of being bombarded by stimuli, it is a withdrawal from that, so instead of going out you went in and down. So that's a totally different way of looking at writing. I mean I can just follow people on the bus, not like a stalker but if I hear interesting conversations, I've seen me staying on the bus for several stops because I'm desperate to hear the end of the story. So anything can help me write.

Speaker 1: (14.25) Quite right. I shared a bit of AI generated poetry with you. It was quite basic, but what did you make of it?

Speaker 3: (14.33) Well, just put it like this, if you get the chance of instant coffee. You know, the cheap instance coffee, just sling in the teaspoon and stir it up and it tastes like sludge, fine. But if you go for the connoisseur's coffee, you know you use a cafetiere...because it's got to be filtered through experience, human experience, that's when you'll get something genuine. An elephant can do painting when it feels like it but no-one classes that as art, do they?

Speaker 1: (15.06) Now Sheena you've travelled extensively over the years in Thailand, India, Sri Lanka, Egypt, China, researching material for your work. How important do you think that personal experience, or life experience, is when it comes to writing meaningful copy that strikes a chord with people?

Speaker 3 (15.24) Well what I found was interesting from what we heard previously there was the word heart, family and generation. I wouldn't put the name heart, family and generation anywhere near a machine because it doesn't resonate with that. It's interesting that they are doing the same with art now, they are managing to fake pictures from famous artists, a machine produces this now. We can produce sheep, eventually I suppose we will be able to clone people and we mightn't even be necessary at all, we can just stand aside and let the machines take over.

Speaker 1: (16.06) Do you think people want that though? Do you think people want machine-generated poetry or music lyrics?

Speaker 3: (16.12) No. The reason I travelled extensively is that I'm interested in how other cultures operate. I'm not quite sure how the machine would cope with different cultures. It's a difficult one I mean to start with I absolutely loathed word processing, I hated it, I said it will never catch on. I just like to write by hand. It took me a long, long time to dip my toe in the water and now, yes, it's very, very useful and I wouldn't be without the word processors and all the rest of it, so I have been dragged kicking and screaming into the 21st century but not entirely.

Speaker 1: (16.55) Well I guess these things are tools for us though, aren't they?

Speaker 3: (16.58) There is a type of writing called therapeutic writing. It was started around the Sixties by a man called Roberto Assagioli and it's called psychosynthesis and he developed techniques for dealing with grief and all sorts of issues and the nearest to that is a machine-operated thing called the therapist, or something. And it just repeats what you say to it. So you would say to it 'I'm terribly frightened of trees, I've got a terrible phobia about trees', and it will go 'A phobia about trees?' and there's a pause and this encourages you to

speak back as if you were talking to a psychiatrist. I don't know how many people would care for that instead of a real live person to speak to, but you know, why not. You could stand at the well at King's and bawl down there and get the echo coming back to you and get the same effect.

Speaker 1 and 2: (17.52) <Laugh>

Speaker 3: (17.55) But therapeutic writing is something that's quite personal and people use it to deal with inner problems. It's very much an individual thing so how a machine can cope with an individual because it's very much...you know...you don't expect a Hoover to have guilt or innocence or joy. It's just a Hoover, it sucks up dust.

Speaker 1: (18.25) Well Georgios, you were saying that systems like ChatGPT are only as good as the information that's fed to them. Does that mean that they also reflect inaccuracies and bias in terms of the data they are using or have access to?

Speaker 2 (18.40) Exactly, definitely, that's one of the main limitations of such systems at the moment and also one of the reasons they are criticised for. As you can imagine when they consume all of the data that is available online, the data is not filtered online so you have any sort of inappropriate content that these systems might have been developed with. That means that both the output that this model actually gives you is actually inspired by the infodata and infodata can be biased, discriminating, exhibit all sorts of unfairness and discrimination and everything like that. So it is expected that they will generate content that will imitate the types of data that they were trained with. Having said that, although the complete datas are not really disclosed there has been some sort of curation of the infodata that OpenAI used to train this model and there has been some news that I actually read a couple of days ago about outsourcing the job of doing this type of activity and that meant that some humans had to go through this type of data and they had to curate data, so imagine the cases where people are reviewing very toxic content to filter this out, and I read some story – I don't know if they are true or not – but they were describing cases where people might have had some sort of mental issues because they had to review this content and this content was unfiltered so they had to review anything and everything that is out there. So toxicity of these models is one of the biggest problems I would say at the moment and bias and sexism and all of these bad attributes of such models is something that is an ongoing issue I would say.

Speaker 1 (20.31) So how do we address that?

Speaker 2: (20.32) These technologies, they are a positive step to improving how we do stuff but we have to be mindful that there is a lot of stuff that has to be improved yet and toxicity and filtering the outputs and making sure the systems are used appropriately and responsibly, these are the foundation blocks that everything has to be built up on otherwise we will be having systems that are free agents, doing whatever they want, providing any sort of output and everybody can consume this output with all sorts of consequences of that so I think that will generate a lot of interest from ethicists and people that work around bias of such systems so I'm hopeful that technical people like myself will work with people like ethicists and people who are looking at other elements of technology to be able to come up together with some solutions that will mitigate the problems that the systems have at the moment and provide something that's better suited for a mass adoption, I would say.

Speaker 1 (21.32) AI has significant potential benefits and risks for universities – in relation to ChatGPT here at Aberdeen we are continually monitoring developments to inform our

approaches going forward and reviewing any relevant policies for both staff and students – what are your thoughts around its use in the education sector, Georgios?

Speaker 2: (21.53) I would say that my personal opinion is that we have to be open minded with every sort of new technology that comes out. Of course we have to be prepared to respond to such technologies and put some policies in place that will deter people from abusing the system, so using them in the wrong manner. I'm supportive of having such systems used in an education setting. I can imagine cases where perhaps these systems can accelerate learning processes, they can be used as a means of finding information quicker. I guess, all of us we grew with technology in some form or another and we've seen how the internet changed how we do stuff, we saw how YouTube that was launched back in 2004 has changed the way that we even see content, even many courses are available through YouTube and many people are educated through YouTube, so if we go back 20 years then we might have said the same thing about YouTube. You know you have a platform that gives you content, it is unfiltered, kids might be watching that – all the discussions we are having now we could have had back then and even worse, so I think that technology has space in every setting, education included but we have to be mindful of perhaps the consequences of having this technology available to everyone. In my mind you cannot just say I'm banning a technology because students might be using it to plagiarise or write essays for them. I think we have to develop and evolve and find ways of incentivising students not to use these tools, or maybe incentivise them to use them but in a specific way. Perhaps, write an essay and then compare with what they would have written and then write an essay about this comparison between those systems so I think we have to be creative about technology but as a University we are, of course, we have to make sure there is a policy in place to ascertain that students abide by the rules that we have available and that the education that is offered to them is appropriate in terms of assessing and educating them. But I think ChatGPT has a place in education, or similar systems.

Speaker 1: (24.10) Are there tell-tale signs that a piece of copy has been system-generated?

Speaker 2: (24.15) Oh, that's a good question. So, yeah, there are some cases where you might be able to understand that the content is machine-generated content. Now, again, I might be the wrong person to ask because I'm not a native speaker so the way I use language is most likely different to how a native speaker would use language but sometimes the machine generated content is much more structured, well formulated, almost artificially looking sometimes so when you read that you say, hmm, that is very formal language and sometimes a bit awkwardly written. You know I wouldn't say that you will find something extremely obvious that is wrong but if you look on Twitter there are quite a few users who will try to stress-test these models and they are trying to pose some very tricky questions. For instance, questions around mathematics or try to kind of query them and ask continuous questions and then asking another question and then you see that the system contradicts itself on a previous answer which was correct before and then it became a wrong answer. So then you say why did you give me a good answer and a correct answer before and now, if I slightly change my question, you give me a wrong answer for the same thing and you provide different sources and there were some cases where they system apologised for making a mistake and then went back to correct itself. So suddenly you have this situation where the way the cadence evolves is unnatural at times. So I don't think you'd have a hard time finding these cases, there are quite few of them actually.

Speaker 1 (25.49) Is there any equivalent in human-generated writing Sheena?

Speaker 3 (25.53) You can get performance against content, and by that I mean you might get a writer who reads in a monotone so when they are trying to produce their own work through speech its very boring and just kills it stone dead. On the other hand you will get a performance poet who gives a very vital, vigorous performance but when you read it on the page you think, no.

Speaker 1 (26.32) But it's an evolution, isn't it? Snapchat has recently launched a new ChatGPT-driven app function and I'm sure others are exploring its use, what do you think the future of this technology looks like?

Speaker 2 (26.45) Actually I think quite a few browsers like, for instance, Google Chrome or Microsoft Bing, they have started incorporating these systems onto their normal search engines so you will see that you can enable this type of extension which means it is a functionality of your browser, and while you are searching for, I don't know a recipe for, having friends over for lunch and you want to create something new, then you will see at the same time ChatGPT generating suggestions on how you could go about providing them with a three course lunch at the weekend, so then you can see that you have ChatGPT providing out in parallel with your normal search processes that you go through whenever you look up something on the internet. And if we extrapolate that, we can see many cases, like you mentioned Snapchat and others, that ChatGPT will become a feature, like a component of other systems. And ChatGPT is just an example, Google have their own equivalent that they are developing that's called Bard, and many other companies will do the same. That's why I mentioned before that I'm sceptical to focus specifically on ChatGPT in general in our discussions because this is just an example of a system but there's going to be many more such systems deployed in the near future, so why do we focus on ChatGPT and its limitations when this is a very fast evolving domain when ChatGPT tomorrow might be obsolete. If we think that ChatGPT was released a couple of months ago and there have been so many cases and discussions about that, imagine what's going to happen in the next five months, or six months, things are really progressing quite rapidly in this domain.

Speaker 1 (28.27) Are these systems going to change the world?

Speaker 2: (27.30) I hope so, for the better, as technology has done in my opinion. Of course we have to be mindful that nothing has only positives, I would say, like technology has had positives and negatives but I think, I'm quite positive in saying it about technology so I think it is going to change the world and its going to, perhaps, make us rethink how we do some stuff and perhaps refocus on other stuff also as humans, and that goes back to the question you asked me earlier on about whether these systems will replace humans in some certain cases or jobs. I will insist on what I answered before that are going to see cases where these systems will support humans but not replace humans, or evolve the types of jobs we are doing, but not replace anyone from the jobs we are doing but yeah, hopefully I think systems like ChatGPT are well placed to change and improve the world around us – with some negatives at the moment that we have to fix of course.

Speaker 1: (29.24) This has been a really interesting discussion. Thank you for your time and your insights both but I'm afraid our time is up. And thanks also to our listeners, I hope you've enjoyed today's look into the headlines. I'll be back soon to unpick more topical news stories with experts from the University of Aberdeen soon, but if you just can't wait, visit abdn.ac.uk/slash/news to find all the latest stories and announcements.

Outro music