Al Language Models: Real Intelligence Or Creative Thievery? <u>https://mindmatters.ai/podcast/ep342</u>

Pat Flynn:

All right, everybody, welcome back to the Mind Matters podcast. This is your favorite guest host, I think I can say that now, Patrick Flynn. I'm regularly over at Philosophy for the People, but I love hopping in here on the Mind Matters podcast to discuss all things philosophy of mind and artificial intelligence. I am joined today by Dr. Eric Holloway and Dr. Robert J. Marks. We're going to talk about AI, plagiarism, and the illusion of intelligence, specifically why AI-generated language is not truly creative and depends on kind of a little artistic thievery, I guess we could say.

So gentlemen, thank you again for joining me. It's a pleasure to be joined by both of you once again.

Robert J. Marks:

Great.

Eric Holloway:

Yeah, great to be back.

Pat Flynn:

Yeah. So we'll skip the introductions this time because people can just hop back to the previous episode where we got to know each of you a little bit and get right into the good stuff. Dr. Marks, if you wouldn't mind, let's start with you. What are we talking about today? What is this subject and why is it important?

Robert J. Marks:

Well, we're talking about AI, plagiarism and the illusion of intelligence. It turns out that these chatbots, such as ChatGPT and Grok and Perplexity and a number of the other ones based on large language models, are assumed to be creative. And indeed, they're not creative. They are a result of plagiarism and capturing the ideas of other people.

I like to think of it as... well, let's go back to Isaac Newton. Isaac Newton said, I have done what I can do because I have stood on the shoulders of giants and he had a big library, by the way. When I was a boy, we used to go to what was called the stacks and the stacks of the library were where they kept all of the journal articles and you go up there and you look for a journal article and you'd find the volume, they had them all bound, and you'd take out the volume and you'd leap through it and some jerk had ripped out the paper, you didn't get it. The internet came along and all of a sudden all of this stuff was available. So it became easier and easier to stand on the back of giants, just like Isaac Newton said.

Today with these large language models, oh my gosh, it's so easy to climb on the back of giants because you have all of this knowledge which is available to you in a very understandable way. Yet just like Newton said, he was creative because he was able to stand on the backs of giants. He added the creativity. And just like today, the creativity is added not by ChatGPT, not by Grok, not by Perplexity. It allows us to stand on the shoulders of giants and have the ability to create, and that is a human attribute which is not resident in the computer. And so that's where we are. These large language models are astonishing what they can do. But nevertheless, we're just standing on the back of giants. We don't see them creating. We talked about AI doing creation last time and not that.

Pat Flynn:

Yeah. And as well, when we talk about being a thinking thing, an intelligent thing, a comprehending thing, this is another matter as well, including being a rational thing or a conscious thing. And I spend a lot of time playing around with these large language models, and you're right, they're extremely impressive what they can do. But the more time you spend with them, I think the more obvious it becomes that, yeah, they're not thinking things, they're not comprehending things. I mean, just today I was doing some research on weight loss actually, and it just kept getting references consistently wrong. And I kept saying, this is wrong, this is not right and it was clearly not comprehending. I had to keep prompting it. I had to put it into deep research mode eventually. And it kind of got it right eventually, but it was just very clear with the amount of correcting you have to do, especially when you start getting into specialized matters, this is not a thinking thing. It's an impressive thing, but it is not a thinking thing. Maybe it will be able to fool us later on because it will just get better at its approximations and its magic tricks. But even right now, I think if you're paying attention, there's some very obvious signs that what you're dealing with is very impressive, but not a thinking thing. Do you guys agree with me on that statement?

Robert J. Marks:

Yeah, I would certainly agree with you.

Eric Holloway:

Yeah. Well, it sounds like, Pat, that you have some background philosophy. So you probably know about Searle's Chinese room experiment. Have you heard of that?

Pat Flynn:

I do. Yes, of course. Yeah, we've discussed it on this podcast several times actually in previous episodes. But yeah, go ahead and re-summarize it for people.

Eric Holloway:

So basically the idea is you have this librarian, he's sitting in a giant library full of translation books where it gives you a Chinese character and it tells you what Chinese character would come after that. So then he gets a message that's full of Chinese characters, but the thing about this librarian is he has no idea about how to read Chinese. He has no idea what these messages say. But what he can do is he looks up each Chinese character in his massive library of books and then just writes down the next Chinese character that his rule books reference, and then he passes that out the other side.

So the experiment shows that you can have something that looks like it's intelligently processing language, but actually has no idea about what's actually going on there. And the funny thing about systems like ChatGPT is it's pretty much a programmer read that philosophy mind experiment, and then decided to write a program based on it. So that's basically what ChatGPT is, it's a huge library.

Pat Flynn:

Put it out there at large, right? Yeah, 100%.

Eric Holloway:

Yeah. It's basically a huge library of just lookup tables. There's zero understanding going on there. It's just looking at all the characters you've written and then based on its huge library of lookup tables, it predicts just the next character. It's not even predicting a word or a sentence or a cohesive unit of thought. It's just predicting a character and then it just puts that character down and then uses it to predict the next character and so on. So there's zero understanding going on there. It's just like the poor librarian looking up letter after letter.

Robert J. Marks:

Eric makes a great point. Not only does AI not understand what it's doing, but it will never be creative in the sense of generating something which it doesn't know before. I think we have to be careful in pointing out the flaws in ChatGPT because chances are they're going to be improved someday. They're going to be improved, but they're going to be improved under the limitations of not being creative or not understanding.

The current chatbots are based on a 2017 paper called Attention is All You Need, it was written by some Googleites. Like Eric mentioned, it's auto-completion based on steroids. They use something called multi-head attention, and they not only look at the next letter, but they try to put it in context. And the people that published this paper actually put in an attention aspect into it to pay attention what were the important words and what were not the important words. So this is what came out in ChatGPT-2, ChatGPT-3. But it's all based again on syntax and not semantics, it doesn't care what the meaning of the word is.

So you've probably heard of the word hallucinations, that these ChatGPTs have hallucinations. I hate the word hallucinations because it anthropomorphizes the ChatGPT, it tries to make it human. It isn't. It responding the way it was programmed for Pete's sake. So it's a mistake. And so they have a bunch of people, all of the chatbots have a bunch of people that are trying to put band-aids on all these cuts. They sit around, they said, okay, it hallucinated here, let's fix this. It's hallucinated here and let's fix this.

Plus, with the modern machines they have actually added on more stuff. One of the things they've added on is mathematics. If you go to Google and you give it 62 plus 4 plus 22 equals, it doesn't go to a web search, it kicks into a math mode. And we have a similar sort of thing working on these large language models that they kick into a math mode. And man, I use ChatGPT, the paid version, it can do calculus, differential equations, probability problems. It's really, really amazing. But that's not because of searching on the web and doing all of these things.

Eric Holloway:

And it's not because of the neural networks.

Robert J. Marks:

No, it's not because of the neural networks.

Eric Holloway:

Yeah, it's cheating basically. They just call out to well-established libraries to do all the math-y stuff.

Robert J. Marks:

Well, yeah, exactly. Eric and I are familiar being engineers and fellow nerds that there's this program called Mathematica that could do all of these wonderful things in mathematics. And the guy that

authored Mathematica, a guy named Wolfram, suggested that ChatGPT should merge with his Mathematica in order to do better. Well, ChatGPT did not merge with Mathematica, but they borrowed from Mathematica and now can do the mathematical sort of stuff.

I was surprised. I teach a graduate program in stochastic processes and I copied a PDF file of a problem in the book, and I fed it into ChatGPT, the paid version. I said, solve this problem. You know what? It did it perfectly. It read the PDF file and it was able to solve it. I mean, these things are amazing. But it's not because of this original 2017 paper, it's the bells and whistles that have been added.

Eric Holloway:

Yeah. And also, actually, real quick on that point too, there's another reason why Als can have this amazing ability to solve things, and that's because the solution may actually be in their training data. So again, they're cheating.

Pat Flynn:

Yeah, good points. And Dr. Marks, it's right that we shouldn't base our assessment on the obvious flaws of ChatGPT. I guess my point was there is that these cracks do give you a bit of an insight, that this isn't the real McCoy, so to speak. But very, very likely these cracks will not be discernible in the future because it is just that impressive. And it's more important to understand these deeper philosophical points like the Chinese room experiment, to understand from that perspective why in principle, this is not a thinking thing, it's not a creative thing and never will be no matter how impressive it gets in making the impression otherwise.

Robert J. Marks:

Exactly.

Eric Holloway:

Yeah. Well, on that point, you can actually demonstrate that these AI models cannot ever possibly be creative. And that's because of something called model collapse. So if you compare it with normal human creativity, if you have, let's say, a poet or some kind of artist or just a writer, they can read other writers work, and that's how they become better writers. They think themselves, are creative themselves. Deal with their own stuff, deal with other writers stuff. And they just keep getting better and better.

Now, you can demonstrate conclusively that these AI systems don't do the same thing, and that shows they're not actually creative and in fact are completely parasitic on human creativity. And that's because if you take just a bunch of AI models and it doesn't matter which ones just use all the AI models you have, and you have them generate all the content and then train on their own content, and you keep doing this over and over again, you get the exact opposite effect. As opposed to a human creative artist who keeps getting better as they practice and train on other material, these AI systems are provably always going to go downhill. They'll just generate either really repetitive stuff or complete gibberish, but they won't do what a human does, which is continue to get better. And so that's why you can conclusively and mathematically say these AI systems are not at all creative in the way that humans are.

Robert J. Marks:

Yeah. So what Dr. Holloway is talking about is the inability of AI to be creative. Because if you train one AI on another AI, you get what is known as this model collapse. And it's astonishing to see the results of

model collapse. A number of papers have been dedicated to looking at model collapse and describing it mathematically.

Pat Flynn:

That's sort of bad news for artificial general intelligence, isn't it?

Robert J. Marks:

Oh yeah. Well, it depends how you define artificial general intelligence. If you define it as having all of the intelligence of mankind, yeah, I think you can do that. If you ask it to be creative, that isn't going to happen. Because all the knowledge of mankind is the mountain that Newton stood on. And I think that it's working towards accumulating that. But yeah, it'll add to that mountain. It will never be creative. Just like Newton did not, he stood on the mountain but he did creative stuff above and beyond the mountain, and that's evidence of human exceptionalism and we're the only ones that are going to be able to do that.

Eric Holloway:

You can think of AI as like the anti-Newton. The anti-Newton that just instead of climbing to be on top of the mountain, the new AI just keeps digging away at the bottom of the mountain to build little piles of rubble elsewhere.

Pat Flynn:

That's a good analogy. Right. Yeah.

Eric Holloway:

So that's also a problem. If we start relying on AI too much, it's actually going to destroy human creativity and lead to also the collapse of human endeavors.

Pat Flynn:

Yeah, we'll be worse off in a sense. That's interesting, right? Yeah.

Eric Holloway:

Yeah.

Pat Flynn:

I mean, I'm already concerned about that with how much people rely on AI for so much. I mean, especially even just basic writing. I mean, who's going to be able to write or edit anymore in the future? I sound like the grumpy old man now, I guess. But dang, it's a little too much of a crutch for people these days I'm afraid. Right? Yeah.

Robert J. Marks:

Well, I mentioned the idea of the ChatGPT solving this homework that I give in stochastic processes, and the students can use ChatGPT in order to generate it, but when I give them quizzes I'm not going to ask them the same question. I give the quizzes in class without access to any computers at all, and they're going to have to understand, they're going to have to know the content of that result in order to apply it to other problems. So that's at least encouraging.

Pat Flynn:

Yes. All right. And we'll just stick to verbal and oral exams I guess too, or something. I don't know. I'm not a professor, so that's your guys' problem, not mine.

The next thing I want to talk about again is this plagiarism issue. Chomsky's got this quote, he calls it high-tech plagiarism. This has been an issue. My understanding is there've been some legal issues around it. So talk to me about this, either one of you or both. What is Chomsky getting at by calling these things a form of high-tech plagiarism?

Robert J. Marks:

Well, indeed they are. And one of the things that you're not hearing reported a lot is all of these plagiarism suits being played out in court. The New York Times filed a lawsuit against ChatGPT's OpenAI for using content that fell under the Times paywall. So you've got to pay to its content. And they're also being sued by the Authors Guild in a class action suit and this guild represents authors like David Baldacci, John Grisham, who used to be one of my favorite authors in his early books, The Lincoln Lawyer writer Michael Connelly, and the guy that wrote The Game of Thrones, George R. R. Martin, they're all in this class action lawsuit against ChatGPT for stealing their material.

Another one is Getty Images. You can also use these large language models for generation of images, except they're not large language models anymore. And Getty Images, if you've ever used the Getty image, you pay them and you're either going to pay them voluntarily or in court because they are very litigious. So Getty Images has taken legal action against Stability AI for employing more than 12 million of its copyrighted images in the training of an AI image generator and the New York Times in making their case show incredible images about the generation of Iron Man, the Joker, Joaquin Phoenix. And then it said, generate an Italian video game person and it gave the Mario brother exactly. So there is clarity that there is plagiarism.

And the question is, what's going to be the future? What's going to be the future after these court cases are decided? By the way, just an update, I looked it up before we talked about it, a US district judge, Sidney Stein, he ruled that, yeah, this suit could go on. It was referred to the US Judicial Panel on Multidistrict Litigation, I've never heard of that. But what they've decided to do is take all of these lawsuits, including from the Times, this Independent Writers Guild, and they're going to just put it in one big lawsuit and we're going to see what happens there. And it's going to be interesting to see what the outcome of that lawsuit is. I have some ideas, but you never know.

Pat Flynn:

So a question for the two of you I have would be more of a practical one of how you might be counseling your students on the use of AI. So to kind of frame this a little bit, my understanding, and can certainly be fact checked on this is when it comes to chess, it's not that humans or computers are the best, it's humans plus computers are the best at chess right now. I think that's still right. And that seems to me like that's probably going to be the case across the board in lots of efforts. Like the best writers won't just be the human or the computer, but it'll be the human working with the computer. The best artists will probably be the human working with the computer. Even in music, we're seeing a lot of AI advances in music right now, probably be some sort of cooperation between the two. And so I think it's probably obvious that those who are really good or skilled at the craft before using the computer will be the best at using the computer in cooperative efforts with it.

So yeah, I'm just kind of throwing that out there to get your general thoughts on whether you think that is the way things will go and how you would counsel people to, on the one hand, yeah, get good at the things that you want to pursue in your life, but also find ways to maybe work with AI not to do it for you, which is the problem that I think we're realizing with people just plagiarizing or cheating on tests or writing papers, but to make you more efficient in a sense.

Dr. Marks, start with you, what are your thoughts on all that?

Robert J. Marks:

Well, I do think that AI is going to be incredibly disruptive. I mean, we no longer see toll booth operators. It's all done by technology. We no longer have travel agents. We all book our own airplane tickets on the web. And so what are the jobs that are in danger? They're the jobs that follow algorithms, a set of procedures for doing something. Those are the ones that are in danger. So we see these ChatGPT and this Grok and this Perplexity generating astonishing art, and it's going to be a gut punch to graphic designers, there's no doubt about it. But I think that the final results are going to be good for hotel rooms and doctor's offices.

Pat Flynn:

Sure.

Robert J. Marks:

And so that's where they're going to hang these things. And if you look at actual art, it's actually the relationship to the artist. And if we look at, say The Beatles, for example, The Beatles weren't the greatest musicians in the world, but they had a lot of promotion and stuff like that and stuck in their fame and they just got celebrated. If you think of the Mona Lisa suppose, for example, that we had the ability with the 3D printer to generate an exact replica of the Mona Lisa. And on that back of that painting, you put copy, and then you put the original Mona Lisa, and you put them up for auction at Sotheby's. Which one is going to attract the greatest bid? It's going to be the original Mona Lisa. Why? Because of the connection to Leonardo da Vinci. The other one we know is an Al copy, even though they're identical. And we see this in music, we see this in paintings and art, and it's the relationship to the creator, which really establishes the value of art. But for just everyday stuff like in graphic design and in pictures, in hotel rooms and doctor's offices, yeah, I think that ChatGPT and Grok and Perplexity are going to generate the images of the future. Yeah, absolutely.

Eric Holloway:

Yeah. I think this gets back to your point from the last podcast, Pat, where you had that scenario where random processes spelled out a message, but that destroys its meaning. I think it's the same thing with AI-generated art. If something's purely AI-generated, then people actually tend to find it fairly uninteresting and boring because they know it's not actually meaningful, there's no real, or at least no significant human intent that went into designing it.

Pat Flynn:

Yeah, that's a great point. I certainly feel that way. And I already know, at least just in my small realm, a number of people who are intentionally seeking out paying human artists, paying human musicians because they feel that, right? They feel that lack of meaning and significance from the auto-generated stuff. Now, is that going to be widespread? No, because I think at the end of the day people would rather save money if they need something for the doctor's office, right?

Robert J. Marks:

Sure.

Pat Flynn:

So I'm not banking on that. And yeah, like anything else I've always been cautiously optimistic about the advent of AI. I think I still hold that. I don't think I'm as pessimistic or cynical as other people. But certainly, yeah, there already has been massive disruption. I think there will be considerably more disruption to come. I hope that it's a net benefit overall, but it's always hard to predict these sorts of things, isn't it?

Robert J. Marks:

It is. Smoke is out of the bottle, AI is going to be here, and we have to learn to live with it. We're seeing now AI being used in the classroom. I see things like Grok and ChatGPT being used in classes on creative writing. They actually encourage the use of these things in order to clean up things. Shoot, I use ChatGPT to clean up my language all the time. I'm an engineer. I never learned how to write good. I even spell good, G-U-D. I don't know how to write very well. So I think a lot of things are going to be changed by this.

There's an interesting thing. It used to be that the US Copyright Office would not copyright anything generated by AI. So this is good news for us. So if we go to one of these AI places like Grok and we generate an image, that image cannot be protected by copyright. We can use it anywhere we want to. That's US copyright law. Now, there was an interesting case about a guy that won an art contest and his final art was generated by AI, and they wouldn't give him a copyright. But he said, you don't understand. He said, what I did is I went into, I think it was Midjourney, which was an image generation piece of software that's available online, he said, I gave it a prompt. I didn't like that. So I changed it a little bit more. So I did all of these changes and finally with, I forget the exact numbers, but it was like 200 prompts over a period of 24 hours that he did. And he says, I'm sorry, I think I should be given a copyright on this even though the final product was generated by AI.

It turns out that the US Copyright Office a few weeks ago at the time of the recording of this podcast, came out in favor of this, and it says, you can copyright stuff generated by AI as long as human creativity goes into this creation of the process. So this guy's original art now should be able to be copyrighted because he used it iteratively. And if you think about as an engineer in design, anytime you design something it takes a number of iterations before you get it right. It's just that we can use AI in the generation.

If you've heard of WD-40, the reason it's called WD-40, water displacement, the guy that invented WD-40 took 40 tries before he perfected it. Same with Formula 409. Formula 409, which is kind of a household cleaner, was perfected only after 409 iterations. So we in engineering know about this iterative process. And thankfully the US Copyright Office has realized that AI can be used as a tool in this iterative process of creation. So I think that that's going to be an interesting future of artificial intelligence.

Pat Flynn:

All right. That's great perspective. Gentlemen, any final thoughts? This has been a fascinating conversation. I've certainly learned a lot, as always. But what would you want to leave the listeners with today?

Eric Holloway:

Yeah. One final thought is actually this is nothing new, this whole concern about AI disruption, it's actually just the standard refrain of the entire industrial revolution. People have always been worried that some form of automation or another is going to destroy some human endeavor. But you can think of it like a plant and leaves. If you have just a little bit of automation, you have like a little shrub and people have lots of leaves on that shrub. But if you have automation, then you can grow into a huge tree and you can have many, many more human leaves on that tree. I think the paradox of automation is it increases the number of jobs. Like in my own field, the ChatGPT or Copilots, all those different helpful programming aids, I think actually lower the barrier to entry for programming and mean many, many more people can become programmers than were previously possible.

Robert J. Marks:

Oh, yeah, I agree. I think to Dr. Holloway's point, the Luddites, if you remember them, they were rioted and they broke up a number of looms, as I understand it. Somebody came out with punched cards, and it used to be computers were programmed by punch cards, but they were using punch cards in order to do weaving, and the Luddites didn't like that, and they did a bunch of protests and destroyed it because they were afraid of the new technology replacing their jobs.

We have a similar sort of situation today where people are concerned about AI taking their jobs. If your job is algorithmic, yeah, you're probably in danger if your job requires creativity or understanding, probably not. Artificial intelligence, it may impress, but it stands on this what we call this mountain of human effort. And the real question isn't can AI create, but whose creativity is it really using? And that's the bottom line. And hopefully I think that it will be demonstrated more and more that the only source of creativity, true creativity, is due to human exceptionalism.

Pat Flynn:

Excellent. Thank you gentlemen, both, so much. And to all the gentle-listeners, please stay with us next time, we're going to take a look at the computationalist theory of mind and explore the question of whether the human mind can truly be reduced to a machine. We hope to see all of you there. Please subscribe. Be sure to leave a review if you like what we're doing here, and we'll talk to you next time.

Announcer:

This has been Mind Matters News. Explore more at mindmatters.ai. That's mindmatters.ai. Mind Matters News is directed and edited by Austin Egbert. The opinions expressed on this program are solely those of the speakers. Mind Matters News is produced and copyrighted by the Walter Bradley Center for Natural and Artificial Intelligence at Discovery Institute.