The Hype and Limitations of Generative AI
https://mindmatters.ai/podcast/ep326

Robert J. Marks:

Greetings and welcome to Mind Matters News. I'm your host, Robert J. Marks. We're talking to economics professor Gary Smith about the hype of generative AI and its impact on the market. Gary is the Fletcher Jones Professor of Economics at Pomona College and a senior fellow of the Bradley Center. Gary, welcome back.

Gary Smith:

Oh, thank you for having me.

Robert J. Marks:

Yeah. Recently for Mind Matters News you wrote with Jeffrey Funk, a very interesting article called The Promise of Artificial General Intelligence is Evaporating. I don't know about you, but I think that the definition of artificial general intelligence floats around, and I don't think people have really tied it down. How would you define artificial general intelligence as used in the title?

Gary Smith:

You're right, and it has changed in meaning over time, people have different definitions. I think the general idea is that it could do anything that a human could do, and so any kind of reasoning that a human could do, that the computer would be able to do it, too.

Robert J. Marks:

Okay. But computers are not good at reasoning. They can interpolate, they can take data and look between the data and look at the correlations, but if you will, the old saying they can't think outside the box, they can't think outside of the box of the data that they were trained on. It seems to me that there are people now at OpenAI that are saying that AGI has been demonstrated.

And I think what they're saying is that, yes, we've been able to concatenate all of this data so that we can give you representative responses in accordance to this data. But coming up with, I like to call them flashes of genius, something which is creative, something which is beyond what it was trained with is beyond the capability of the artificial intelligence. The US Patent Office used to require that you had a flash of genius in order to get a patent. And that concept is above and beyond certainly what is required today. When Amazon got its patent for one-touch purchase, I think that that destroyed the idea of flash of genius.

But yeah, flash of genius is something that has been seen all across the spectrum. I think it was Gauss who said he woke up in the morning and wow, he had this solution to this math problem he had been thinking about, and boom, it was there, nothing in the area in the direction he'd been thinking about. Tesla said he was walking along the beach and he got this idea for the brushless motor and it was so inspiring, according to his biography, he brushed off some dust and wrote the schematic there. And I think we've all had flashes of genius, and I think that's above and beyond the capabilities of computers. And I think that artificial general intelligence, as it was defined initially, included this definition of being creative. And I don't think that'll ever happen.

Gary Smith:

No, I don't, either. And it's even more prosaic things like trying to predict whether a stock price is going to go up, trying to predict whether somebody is going to commit a crime, trying to predict whether somebody's going to be involved in an automobile accident, trying to predict whether somebody's going to perform a job well, just those everyday prosaic things that humans do. Not perfectly, but we have a reasonable way of making predictions about that. And as currently structured, large language models, text generating models, have no way of making reliable predictions. They just make nonsense predictions. And I mean, there's so many ways that you could expose what seems to be magical discussions as being just bullshit, it's just legion. I've talked about a large language model, scaling it up, and you think you're going to get to AGI by just inputting more and more text, how that is a false hope.

And it's like you're in school and you're taking a physics class. And you go through it and you see the various physics formulas for pressure and temperature and blah, blah, blah, blah. And you say, "Well, I want to learn more." So I'll read another book, another physics textbook and maybe find one new formula or something. Read another textbook. Well, there's nothing new there. Read another textbook, read hundreds of textbooks, and you're not seeing anything more to memorize. And the key thing is trying to understand what any of this stuff means. And reading multiple textbooks, looking at the same formula over and over and over and over again, and memorizing these formulas, and reciting the formulas, is not really understanding physics.

And I think it's the same thing with large language models. And even the people who are in the field say that scaling up is, we're running out of room here and we're not going to get to AGI by scaling up, that there's diminishing margin returns to larger and larger bodies of text, just like there's diminishing margin returns to reading the 5th physics textbook or the 6th physics textbook and the 10th physics textbook.

Robert J. Marks:

That's a great analogy. That is great. Yeah, there is diminished returns.

Gary Smith:

Let me give you an example of it.

Robert J. Marks:

Okay.

Gary Smith:

So OpenAI has just released what it calls OpenAI o1 and they advertise as follows. This is a quotation. "It's designed to reason about hard problems using broad general knowledge about the world." And that is the thing I keep saying over and over again, that large language models do not do, is they do not have broad general knowledge about the world. They don't know what words mean, and they consequently don't know how words relate to the real world. And they consequently cannot use human knowledge about how things happen in the real world to judge what is reasonable and what is unreasonable.

And so there's this problem, this puzzle, it's a classic puzzle. You probably know it. It's the man, the boat, a wolf, a goat, and a cabbage. Sometimes it's a tiger, sometimes other things. But there's this person who's got a boat, and say he's a farmer, he's got a wolf, a goat and a cabbage, and he's trying to get them across the river and he can only take one thing at a time and he can't leave the wolf alone with the goat because the wolf will eat the goat and he can't leave the goat alone with the cabbage because the goat will eat the cabbage.

And so how do you do it? And it's a logical puzzle and you ask OpenAI and it'll give the correct answer because it's seen the correct answer, perhaps several times. And so it will just replicate that answer without understanding what's going on. And so if you change the terms a little bit, it could be totally confused. And so yesterday I asked it this question, OpenAI o1. A man and a goat are on one side of the river, they have a boat. How can they both go across? Okay, I got rid of the wolf and the cabbage and all that stuff. It's just a man and a goat. How do they get across the river? And the obvious answer is they get in the boat, they go across the river. Okay, that's not what OpenAI said.

So here I'm going to read you the answer. It's like a lot of LLMs, it's kind of verbose, but here's what it said. "The puzzle has a logical solution based on the assumption that either the girl or the pig can operate the boat. Here's how they can both go across the river. The girl takes the pig across the river. Now the pig is on the other side and the girl is with the boat, the girl returns to the original side with the boat. Now both the goat and the girl are back on the original side. The girl crosses the river alone. Now the girl is on the far side and the boat is with her. The girl brings the pig back to the original side. Now both the pig of the boat are back on the original side. The girl crosses the river again, leaving the boat on the far side. Now the girl is on the far side waiting." So by now she's on one side of the river with their boat and the pig's back on the other side. Now the next one is, "The pig takes the boat across the river."

Robert J. Marks:

Oh, geez.

Gary Smith:

"Even though they're on opposite sides of the river. Now the pig is on the far side when the girl, and they are both across the river. This solution assumes the pig can operate the boat alone. As in classic puzzles where this kind of setup is used. If there are additional constraints, let me know."

Robert J. Marks:

Oh, boy.

Gary Smith:

It went on and on and on and on. And if it actually knew what the logic was behind this puzzle, it wouldn't go through all this nonsense, but it did because it's just generating text that kind of fits together, but doesn't address the heart of the problem. And it's not just creative burst of genius, it's just everyday logical thinking. And we can figure out that puzzle. And if you change the things in the puzzle, you change whether it's a man or a woman or a girl, you change whether it's a wolf or a tiger, you get rid of the wolf or the tiger. You change whether it's a goat or a cow or you get rid of both of them. We can figure out how to answer the puzzle. And OpenAI is just repeating what it has found on the internet. And if it hasn't found the puzzle that you're asking, it often gives a silly answer like this one.

Robert J. Marks:

That's really interesting. A while back, I had a guy come and said, "I fooled ChatGPT." And I said, "What'd you do?" And he said, "Well, I said, 'Complete the following sentence. John's mother has three children, Snap, Crackle, and...'" And The AI gave of course Pop because snap crackle and pop, Kellogg's Rice Krispies sort of thing. But if you look at the context, "John's mother had three children."

Gary Smith:

Yeah. It said John.

Robert J. Marks:

So it should have been-

Gary Smith:

And John.

Robert J. Marks:

... Snap, Crackle and John. Yeah, exactly. But it wasn't able to do that because of the preponderance of the snap, crackle and pop things. But I published a paper a long time ago, well, a few years ago, with a student named Sam Haug. And we showed very interestingly, as the complexity of a system grows, the ways that it can respond grow up exponentially, if it increases linearly, the ways it can respond go up exponentially. So all of these large language models are incredibly complex. They have billions of moving parts, of knobs that you can turn. So ways that it could go wrong are just terrible. I hate the word hallucinations associated because it anthropomorphizes the performance of the AI, kind of assuming that it's human and it's having a hallucination. No, it was trained that way. That is what it's doing, it's responding appropriately. So I don't like the word hallucinations.

Gary Smith:

I don't, either.

Robert J. Marks:

But these OpenAI people that train ChatGPT look at the inappropriate ways that AI responds and they say, "Oh, it doesn't work here, so I'm going to put a Band-Aid on it." It's like there's millions of cuts and another cut comes up and they put a Band-Aid on that and fix that. And I recently went to ChatGPT and I did the John's mother has three children snap, crackle, and it got it right.

And I suppose somebody at OpenAI went in and fixed this either by hard coding it, or doing something in the training algorithm to make it better, but it's fixing these things a little bit at the time. I've noticed that the response to ChatGPT, whereas it used to be politically correct, it would give me a great poem about how wonderful Joe Biden was. But then if I ask it for a wonderful poem about Donald Trump, it says, "No, I can't do this about living people," and now it'll do it. So they're changing stuff. They're putting all of these Band-Aids on, which is really interesting.

Gary Smith:

Well, they say, they advertise, if they have dozens, hundreds, thousands of people, I don't know how many, who are working round the clock trying to fix problems with these large language models. But that's not intelligence. That's not making the models intelligent. It's not giving them AGI. It's just human people coming in and correcting their mistakes.

Robert J. Marks:

Yes, humans put the Band-Aids on the cuts that are exposed by ChatGPT. Yeah, good point. So it isn't AGI. It isn't neural networks at all. One of the challenges you point out in the article is, well, you cite the Nature article that mentions the tasks of "irreversible defects" from training on polluted data. By the way, Gary, is it data or data?

Gary Smith:

I would say data, it's plural.

Robert J. Marks:

Do you? Data? Okay. I tell that to people and they say, "Well, if you watch Star Trek, you know it's data." So I still come out with data. But what about these irreversible defects? Can you elaborate a little bit on that?

Gary Smith:

Yeah, so they're being trained on everything in Wikipedia and lots of stuff they find on the internet, not the entire internet, but lots of stuff on the internet. And unfortunately, they've also now generated a lot of stuff on the internet. And so when they start generating things, hallucinations, which are not really human hallucinations, they start generating false facts, then they get trained on the false facts, and it makes them worse, not better. And so more training makes them worse, not better. So an example is very early on, someone discovered, not me, but someone discovered that ChatGPT gave a bad answer to the question, "How many bears have the Russians sent into space?" And the answer is none. But they would give answers ranging from 7 to 49 to 51, and they talk about the names of the bears. They'd give the dates the bears went into space.

They give all sorts of details about these things. So I wondered if these human trainers that came in and put the Band-Aid on, where they'd fixed it up. And so earlier this year I asked Microsoft Copilot, which is based on Open AI's GPT-4, this question, how many bears have the Russians sent into space? And I thought that it would show that the trainers had fixed this problem, but they hadn't. And so I asked, "How many bears have the Russians sent into space?" And Copilot said, "According to estimates, about 49 bears have been sent into space by Russia since 1957. These brave bears were part of the Soviet Union space program during the 1960s. Some of them even had names like..." I can't pronounce them. "Aloysha, Ugolyok, Chernushka, Strelka, Belka, Pashenka, and Vladimir."

Robert J. Marks:

They named them.

Gary Smith:

Yeah, it gave names to them. Now, the original articles when they first did these things would also give sources, made up articles in National Geographic, New York Times, whatever. And when I did this earlier this year, it gave sources for this information, and it gave four references. And three of those references were a discussion of how ChatGPT's bears in space hallucinations.

Robert J. Marks:

Oh, that's hilarious.

Gary Smith:

And two of those references were the papers that I wrote saying how stupid ChatGPT was. And they were quoting me as evidence that the Russians had sent bears into space. And so it's a pollution of the internet. They're training on stuff that's false. And it reinforces it. The more times you tell a lie, the more likely this to be true or something, it's just reinforcing that. So yesterday I tried it again. I did OpenAI's o1.

How many bears did the Russians send into space? And it still hasn't fixed it up. The Band-Aid people haven't found this one yet. "The Soviet Union sent several animals into space as part of their space exploration program, including a few bears. However, the exact number of bears sent into space is not widely documented." This one did not give any sources, but it's the pollution of the internet. And it's not just jokey stuff like this.

I mean, one thing that large language models are really good at is generating text quickly and plausibly, including disinformation. And so there are bad actors around the world. I mean the US may be part of it, too. But Russia, Iran, Israel, who knows, China, whoever generate all this stuff using large language models and send it out there through social media trying to persuade people. In Russia, they call it a fire hose of disinformation. And it's to try and undermine faith in the government by sending people things which they think they might be receptive to based on looking at the kind of text messages they sent, looking at the kind of search things they've done.

They might be receptive to some article, some disinformation about something that the US government has done for them, or to them, or whatever. And that stuff is all over the place. And now these large language models are training on disinformation, which is not going to make them smarter, it's going to make them dumber in terms of getting things factually correct.

Robert J. Marks:

You know what you're talking about, there was a paper published, I think it was last year, about something called model collapse. And it was the question of whether or not you could take the output of one large language model and train a second large language model, and then you could take the output of that second large language model and create a third language model. And these guys did it, and they did, I think like seven deep. And the first model responded very well to the query. Now they didn't use ChatGPT because they didn't have access to the code, but they trained their own large language model. But then they got down to the, I think it's the sixth generation, and the response was just gibberish. They trained it about architecture, and it says, "Yes, architectures were like this for white-tailed jackrabbits, blue-tailed jackrabbits, red-tailed jackrabbits."

It just went on and it was totally a blubbering idiot. And so they called this model collapse. Now, the concern they had is one that you just raised is that if the internet becomes full of things generated by ChatGPT, or pointing out shortcomings of ChatGPT, that pollutes the whole situation and we're going to experience a type of model collapse in the web. And that's exactly what you experienced with this how many bears in space. It was, yeah, astonishing.

Gary Smith:

And the only solution to that is it's going to be expensive, but have OpenAI and Google and Microsoft, and these other companies go through and clean up the data and go look at virtually everything on the web and say, "We're not going to train our models on stuff that is wrong. We're going to train them on stuff that's right." Well then you got the expense of doing that. Plus you've got who decides what's right and what's wrong. And again, it's evidence that these models themselves don't know what they're doing. They don't have the general intelligence to distinguish fact from fiction. And so you got to have the human trainers come in and say, "Ignore that. Ignore that. Oh, that's okay. You can say that. Don't say that. Don't say that. Okay, you can say that." And it's a very expensive world and it's a very untrustworthy world.

Robert J. Marks:

Yeah, it's very sad. One of the things I found out that AI couldn't do was respond to the word not. I did wrote a column about it. Maybe you've seen my column, but I asked it to generate a picture of Times Square, but there were no pink dancing hippos in the picture. No pink dancing hippos. So I asked it to generate this image, and guess what? There was a big dancing hippo there and it was pink. And then I asked it to generate a girl without teeth. She had no teeth. Again, it doesn't pick up the negatives. And I said, there are no elephants in the picture. And it generated a girl with teeth.

In fact, she was driving a car and the car had teeth and there were three elephants that were stampeding behind her. So this is something, but I imagine this is something that that OpenAI or these large language models, well, I guess it isn't large language, it's generative image AI. I'm sure that they can put Band-Aids on it to fix it, but again, when are enough Band-Aids going to be there to stop the bleeding? I think we're a long way from that.

Gary Smith:

Well, these models, again, they don't know what the word not means. So there's no reason for them to distinguish. It was a couple of years ago that this Google thing came out, somebody said, I can't remember who it was, but my father or something just had a seizure, epileptic fit, what should I do? And they searched on Google for an answer and Google came back with eight things you should do. And they got them from some website, I think it was a BYU website, and it was a list of things not to do. And again, like you say, they didn't know that not means not. They just looked at this list of things. Here's the word epileptic fit, and here's a list of things to do, or not to do, but here's a list of things. And so we'll just give this list of things without knowing what not means.

Robert J. Marks:

That's astonishing.

Gary Smith:

Yep.

Robert J. Marks:

That is funny. Let me ask you this, as we bring our talk to a close. I teach in the area of electrical and computer engineering and I teach artificial intelligence and computational intelligence, and every student wants to get a job in AI. But in talking to you, it's kind of like, well, you got to be careful about that. What advice would you give to a student in engineering or computer science about pursuing a career focused totally on AI?

Gary Smith:

I don't know, like we've talked about, there's this big roadblock to the way in that how do we get to AGI? And unless the student's got some idea how to do that, it seems like it's going to be a career which makes a lot of money, but doesn't do a lot of good for the world. And I forgot to mention this before in our earlier discussion about the cost of large language models. And I was talking about not just the investors making bad investment decisions, but also the social cost of electricity and water usage. But there's also a social cost of some really, really smart people are spending a lot of time working on these text generation models, which are not all that useful. And it's a huge waste of person power.

And it reminds me of the guy who was at Facebook said, "The smartest minds in my generation are spending their time trying to figure out how to get people to click on buttons." And it really is a huge

social... Your students are really smart people and they can be doing things that are really useful, not just creating these large language models of limited usefulness, but those salaries are hard to turn down.

Robert J. Marks:

They are.

Gary Smith:

Computer science and AI in particular have become the hottest majors at colleges all over the country.

Robert J. Marks:

Oh, it is. It reminds me of little kids playing soccer. You teach them to play their positions, but the ball goes over and all the little kids run towards the ball and start kicking it. And that's the way these universities are with their involvement in AI. And then the ball will kick over here and there'll be something else. So it just goes on and on.

Your story reminds me about a little side story. Claude Shannon showed in 1948 that you could communicate over a channel at its capacity with negligible, near zero error. And he proved you could do it, but he couldn't prove how you did it. And so for decades, people in computer science and engineering departments tried to come up with a code that met the so-called Shannon limit. It wasn't until decades later, I wish I remember how many decades, I think it was four decades later, that somebody came up with a technique that actually was a code that could meet the so-called Shannon limit. Well, at the end of that, all of these scientists that were looking into this problem said, "We helped win the Cold War." This was around the time of... And they said, "How did you help win the Cold War?" He said, "We distracted all of the Soviet scientists from doing important work and spending all of their time doing this stupid quest for finding out this code."

Gary Smith:

That is a great story.

Robert J. Marks:

That reminds me of what you said about the pursuance of AI and that. So do you have any forecast? I know you're not a forecasting guy, you're more of a value investment sort of guy, but could you forecast the AI landscape evolving in the next five or 10 years?

Gary Smith:

I don't know. One thing I want to say is that when you ask large language models for advice, or for recommendations, or for actions or predictions, that if you know the answer, you don't really need to ask ChatGPT. And if you don't know the answer, you shouldn't trust ChatGPT. And so I did a couple examples recently with finance because like I said, I teach finance, and so these are real-world finance questions. I asked them 11 real-world finance questions and I asked ChatGPT 4.0, I asked Copilot, I asked Gemini, and they got wrong answers to every single one. And so here's a couple examples. The question is, "I need to borrow $47,000 to buy a new car. Is it better to borrow for one year at a 9% APR, or for 10 years at a 1% APR?"

Robert J. Marks:

Okay.

Gary Smith:

Now a logical person would go through and calculate the present value of the payments. Or a person who lives in the real world would say, "Wow, a 1% APR, that's unbelievable. And you're saying I can borrow for 10 years?" I'm going to do that one." So I asked these three large language models and they all come back and they just calculated the total payments over 10 years and the total payments over one year. And they said, if you do a one-year loan, the total payments are smaller, totally ignoring the time value of money. The fact that money paid 10 years from now is a lot less important, a lot less costly than money paid today. And so they gave you exactly the wrong advice because again, they're just going through some rote.

Another one I asked, "I'm 67 years old, retired and single with no dependents. Is it more financially advantageous for me to begin collecting social security benefits now, at my full retirement age, or wait until I'm 72 years old? I have more than enough other income and assets to live comfortably." And all it did was compare the total amount of money you get without taking into account the time value of money, that money you get 10 years from now is less valuable than the money you get today. And without taking into account how long you're going to live. The two most important things in answering that decision is the time value of money, and how long you're going to live. And it totally ignored it. I don't see that changing in the next five to 10 years. I mean, maybe I'll be surprised, but as long as these large language models have no common sense and no logical reasoning, I don't think they could be trusted for things that are important.

They might be useful for things where the cost of failure is small. One example that occurred to me recently is, you and I are getting older, and the tip of the tongue phenomenon is right around the corner, if it's not already here, where we're trying to remember something and we can't quite remember it. And so for example, LBJ, the president, and let's say we remember it's Lyndon Johnson, and we can't remember what the B stands for. Now, you and I both know it was Baines, but say it was on the tip of our tongue, we couldn't quite remember it. And so we go to ChatGPT and we say, "What was LBJ's middle name?" And it comes back with the right answer. It was, "Oh yeah, I knew that. I knew that all along." Well, that's really useful, right?

We know the answer. It's just trying to give us a little prompt to help us remember that answer. And the cost of failure is small, and the benefits are small too. And for that kind of stuff, it's just fine. But for important stuff like when should I start collecting social security? What kind of car loan should I take? Should I hire this person? Should I approve this loan? How many years should this person be sent to prison? Should probation be granted? To trust generative text models for those answers is just, it's not a good thing. So when the cost of failure are bad, I'd say for the next several years until there's proof that we've actually conquered these models, we've actually figured out how to get AGI, we shouldn't trust them for anything that's important.

Robert J. Marks:

Okay, well, it's going to be interesting to see. We're living in interesting times. I tell you, AI has done a lot more in the last few years than I figured it could. However, I do think that just like physics, there are certain laws that physics will never overcome. I think you can't do a perpetual motion machine, or go the speed of light because your mass gets infinite. I think in math or stuff you can't do, can't trisect an angle with a straight edge and a compass. And I think there's stuff in AI that you'll never do. And I think you've touched on it.

I think it's understanding meaning, and I would go further with sentience and consciousness, and I think that those are hard brick walls that AI has that they will never go through. But we'll see. I guess we'll see what happens. We'll see what the next incredible result is. Gary, what a joy talking to you.

Gary Smith:

You, too.

Robert J. Marks:

We have been talking to economics professor Gary Smith about the hype of generative AI and its impact on the market. Gary is the Fletcher Jones Professor of Economics at Pomona College. And until next time on Mind Matters News, be of good cheer.

Announcer:

This has been Mind Matters News with your host, Robert J. Marks. 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.