

AI Dirty Dozen 2020 Part II (<https://mindmatters.ai/podcast/ep114>)

Robert J. Marks:

We continue our countdown of the dirty dozen hyped AI stories of 2020 today on Mind Matters News.

Announcer:

Welcome to Mind Matters News where artificial and natural intelligence meet head on. Here's your host Robert J. Marks.

Robert J. Marks:

Greetings. For our countdown, we are joined by two members of the Bradley Center Brain Trust. Jonathan Bartlett is the director of the Blyth Institute, and he's a senior fellow with the Bradley Center. Welcome, Jonathan.

Jonathan Bartlett:

Thanks for having me, Bob.

Robert J. Marks:

And, Dr. Holloway works for the NIH and is a current captain in the United States Air Force. Dr. Holloway is also a senior fellow of the Bradley Center. Welcome.

Eric Holloway:

Thank you. And, I like to go by Dr. Captain.

Robert J. Marks:

Dr. Captain. In Germany, I think they call them Dr. Professor or Professor Dr., don't they? If you have a professorship in Germany, you are referred to as Professor Dr. or Dr. Professor. I forget which one it is. So, they enjoy their titles in the front. So, you enjoy... What are you, Captain Dr. or Dr. Captain?

Eric Holloway:

Yeah, just Dr. Captain. No need for the rest of my name. I'm the Dr. Captain.

Robert J. Marks:

Okay, doc. I'm going to call you Eric. Okay. We have already gone through number twelve through number nine. We're on number eight. And, the question here is AI really better than physicians at diagnosis. We're told AI is going to replace lawyers and doctors and accountants and all sorts of people. So, let's look at a case of the physicians. This was a piece written on Mind Matters News. Eric, what do you think? Do you think that AI will ever be better than physicians at diagnosis?

Eric Holloway:

Well, I don't know if they ultimately will or will not, but right now they definitely are not. And, this gets back to something that Jon brought up last session about just how unscientific AI sites is. This particular

author, he took a look at 10 years worth of studies for deep learning algorithms on medical problems, and only two of them actually relied on randomized trials while 81 were not randomized. This means basically people can just pick and choose the type of data that makes their algorithm work well. So, really their results don't really tell us anything about how well their stuff works in the real world.

Robert J. Marks:

Yeah. That's fascinating. Gary Smith, who was also one of the fellows of the Bradley Center, he talks about the idea that when you publish a paper based on statistics, you got a problem, and that 90% of those papers that are published on statistics are wrong. They're not wrong, but faulty, I think is the word he used. And, indeed, that's the case when you have incomplete or unstructured data that you're trying to train with. Is this your point?

Eric Holloway:

Yeah. Basically, they didn't follow very strong protocols, and so they can just make a thing that works really well in the lab. And, we have no idea if it's going to work in real life. And, they tend to not work very well in real life.

Robert J. Marks:

Yeah. Gary points to a case of, what did he call it, he called it the Texas Sharpshooter Fallacy. And, the idea is is that if you have a barn door and you paint a bunch of targets on it and you shoot at the barn door with an arrow, you're going to get close to a bullseye if you have a thousand targets up there. So, there was this one case about pancreatic cancer, and they began to look at correlations with pancreatic cancer. And, well, they thought it was caused maybe by smoking. No, it wasn't caused by smoking. What about, I don't know, what about chewing tobacco? No. Chewing tobacco. Drinking tea? No. How about smoking cigars or pipes? No, it didn't correlate. What about drinking coffee? Oh my gosh, there was an incredible correlation there.

Eric Holloway:

Right.

Robert J. Marks:

So, they publish this in the New England Journal of Medicine and coffee futures fell and people stopped drinking coffee. And, in fact, in the end it turned out that it was totally just a coincidental correlation. And, subsequent studies showed that the correlation was just coincidental. In fact, another study said if you drank a lot of coffee, your chances of contracting pancreatic cancer were improved. So, it's just crazy. And, I think that that's one of the problems that we have. But, you hold out promise for the future maybe, huh?

Eric Holloway:

Yeah. I would say probably if you restrict the domain enough, you're going to be able to pull out some stuff. But, the other problem, too, is how they tend to build these systems. They get a data set from some doctors, and then they just go off for a bunch of years and try to make some algorithm that scores highly. What they really need to be doing is working much more closely hand-in-hand with the doctors and trying to optimize particular parts of their workflow with these algorithms instead of just trying to replace them.

Robert J. Marks:

Okay. So, any claims that AI is better than physicians is probably incorrect.

Eric Holloway:

Right.

Robert J. Marks:

Okay. Number seven. AI can implement video games just by watching. This was from an article called learning to simulate dynamic environments with GameGAN.

Eric Holloway:

Yeah. This one is fun to look at, but you won't really be selling these video games that makes for millions of dollars. So, it's able to learn some kind of feedback matrix based on looking at the game screen and the players' input. And, so you get something that looks a little bit like PAC-MAN or a little bit like that game Doom. But, it doesn't stay coherent for very long. Walls will appear and disappear, and ghosts will pop up and disappear. Yeah. So, it's not super coherent, but because you already know what's going on with PAC-MAN, you can squint your eyes and say, "Yeah, that's a PAC-MAN game."

Robert J. Marks:

Oh, so in other words, they train some artificial intelligence with a number of games and this artificial intelligence creates a game. Is that the idea?

Eric Holloway:

Right. Yeah. And, it's not creating a new game. It's basically just reproducing what it already learned.

Robert J. Marks:

Very interesting.

Eric Holloway:

So, they train on a whole bunch of screens of PAC-MAN and player input, and it just learns how to map the input to different screen frames and finds the gradient between those. So, what they can do with that is they can randomize it and come up with random variants of PAC-MAN, but still it remains PAC-MAN in general, just a much weirder kind of PAC-MAN.

Robert J. Marks:

Now things are much more sophisticated today, but I had a friend, Russ Eberhart, who trained a neural network to compose music. And, he used only four or five songs. And, when you listen to the synthesized songs, you can hear the refrains very clearly from the original songs. The AI is much better than that today, but it sounds like something similar was happening then. Now, they call this GAN. GAN stands for generative adversarial network. Tell me what a GAN is.

Eric Holloway:

Yeah. I believe that's what it is. So, what it does is firstly learns a basic model from its input, and then it generates a new variant of that input from what it learned. And, then it learns from that again. And, so

it's a feedback cycle of it learns a little bit and then adds that to its data source and then tries to learn from that again. I'm not quite sure how that works so well, because it seems like you'd end up perpetuating errors you learned all over the place, but...

Robert J. Marks:

Well, you do. I think GAN is the source of the fake faces that we see recorded.

Eric Holloway:

Right. Yeah.

Robert J. Marks:

And, that's really interesting. In fact, our editor and our director of Mind Matters, Austin Egbert, just published a paper on GAN where he applied GAN to radar data for extrapolating data when it was a little bit sparse.

Eric Holloway:

Oh. Very nice.

Robert J. Marks:

Yeah. So, it's still at the formative stages, but yeah, GAN has some interesting things. But, in terms of this particular application where they tried to extrapolate games, it didn't work too hot, huh?

Eric Holloway:

This is what I see with pretty much all the convolutional neural network type results like the GPT result, which I think we'll talk about a little bit later. If it's generating texts, if you look at just a few words or a few sentences or maybe at the paragraph level, and you squint your eyes, you can get something that makes sense out of that. But, once you start stepping out and gain the bigger picture, it falls apart because the neural network is really good at learning these very closely related relationships, but it doesn't really have a concept of the overall structure of anything. And, so that's what you see in these video games, too. PAC-MAN, you move around and within four or five squares, you see pretty much the same maze. But, once you leave an area and come back to an area, it starts misremembering what it came up with before. It's like a bad dream of PAC-MAN.

Robert J. Marks:

Fascinating stuff. We're counting down the dirty dozen hyped AI stories of 2020, and we're on number six. And, that's what Dr. Captain mentioned, Eric mentioned. GPT-3. Those are four alphanumeric letters that rhyme. GPT-3. And, there was a headline that says there's a subreddit populated entirely by AI personifications of other subreddits. First of all, what's a Reddit for those of us who are not socially media savvy?

Jonathan Bartlett:

Reddit is a site where people post links and comment on them. Reddit is grouped into categories and subject matters, and you can go and find what's interest to you and people post links and articles. And, then you make lots of comments. But, it develops a social gathering type of feel. And, so basically there

was some posters who were posting within some of these subreddits, these subcategories, and it took a while before anyone noticed that these were actually bots that were posting.

Robert J. Marks:

Well, it's interesting that people were able to notice. Now, GPT-3 stands for, I looked it up, Generative Pre-trained Transformer 3. And, some of the headlines of GPT-3 were just scary. Wasn't this, when it came out, the developer said this might be too dangerous to release because of all the fake headlines that it would generate?

Jonathan Bartlett:

They've made lots of different claims about GPT-3. And, it's impressive as a demo. It really does do some impressive text generation. In fact, I think someone actually built a code generation system based off of it. So, you could describe in plain words what you wanted the code to do, and it would actually generate a functional code to do what you asked it to do. So, it's actually got quite a bit of wow sizzle to it, but it turns out that it's not... Once you try to get it to do anything serious, it loses it's luster.

Robert J. Marks:

Yeah. GPT-3 was trained with billions and billions of articles, including all of Wikipedia and a bunch more. And, I think one of the big claims from GPT-2 to GPT-3 was this great, massive increase in the amount of training data that it did. And, you could just take a few words and prompt it and, boop, it generates a paragraph corresponding to those words. And, a review in Wired said GPT-3 was provoking chills across Silicon Valley. But, like you said, it was one of these real quick things where you didn't get into too much of depth. And, I think it was you in your article that you wrote for Mind Matters News said, it's very impressive if you don't look too closely. Is that right?

Jonathan Bartlett:

Exactly. It's one of those things where when people see some of these results, I think people start expecting things that they really shouldn't be expecting from these sorts of systems. For example, one thing that was really impressive is that this is a text processing engine, but it turns out that it can do math. It can do basic arithmetic. But, it turns out that once you get past three digits, it doesn't do basic arithmetic at all.

Robert J. Marks:

Oh, really?

Jonathan Bartlett:

Yeah. So, if you asked what's the number before a hundred, it would tell you it's 99. If you ask it what the number before 100,000 is, it would say 99,009, which is not the number before 100,000.

Robert J. Marks:

I see. Okay.

Jonathan Bartlett:

Anyway, so it's one of those things where, because it... I can just imagine some mid-level manager playing with this and giving it lots of simple arithmetic things and just assuming that since it did all of the

examples he threw at it actually could do arithmetic. And, if he then said, "Hey, use this as our engine for this," and we expect people to throw arithmetic at it, then as soon as they get into four digit numbers, it starts breaking. This is the sort of thing that if you take these systems too seriously, then they can wind up causing damage in the end when you expect them to be more than they really are.

Robert J. Marks:

Yeah. That's what I understand also. GPT-3 was able to write short paragraphs that were just astonishing in their coherence. But, if you ask them to write a chapter, all of the sudden that coherence was lost. Is that a fair statement?

Jonathan Bartlett:

Yeah. Somebody did a, I forgot if it was a... I think they did a series of blogs with GPT-3, and they actually were really good. But, it turns out that they said that they did it unedited, but really what they did is that it's unedited in the sense that they didn't actually modify words, but it is edited in the sense that if it said something nonsensical, they would try something else.

Robert J. Marks:

Now, you talked about there's a subreddit populated entirely by AI personification of other subreddits. That's the title of the article. But, you mentioned that somebody noticed that this was a product of GPT. Unless somebody volunteered it, how did they know that this was generated by GPT? Is it something that can be recognized by people? Do you know?

Jonathan Bartlett:

I don't remember exactly how they wound up figuring it out. But, at the end of the day, usually AIs wind up saying something that's completely nonsensical. One of the things that GPT-3 does, somebody was poking at it a bit, and if you ask it basic questions about the United States, it could tell you who is the President of the United States in different times. But, you could also ask it, "Who is the President of the United States in 1600," and it would give you an answer not recognizing that the United States didn't exist in 1600. And, you could ask it, "How many eyes does a blade of grass have?" And, it would give you an answer of a blade of grass has one or two eyes.

Robert J. Marks:

I saw that. Yes. I saw that article. Yes. So, yes.

Jonathan Bartlett:

And, so that's usually how you wind up sussing these out is they'll start up talking something that sounds logical, but winds up being more or less nonsense.

Robert J. Marks:

Well, this is one of the limitations of AI, isn't it, is it can only think inside the box, it can only interpolate on its training data. And, extrapolation outside of the box has to require creativity. And, artificial intelligence doesn't have that creativity. So, that's the reason it could be fooled so easily, or at least that's one of the indicators that it's not as wise as it seems.

Robert J. Marks:

Okay. Number five. Lack of sleep could be a problem for artificial intelligence as we continue our countdown. Now this is from Scientific American. Lack of sleep could be a problem for artificial intelligence. Does artificial intelligence need to sleep, Eric?

Eric Holloway:

Yeah. I looked into this a bit. It's a little bit hard to figure out what they mean exactly by sleep. And, it seems to be one of those cases where they're trying really hard to make an analogy behind some obscure mathematical thing they do in everyday life just to make AI sound more humanlike. My best guess is... Well, what they say they do is they train these networks and then they have to subject the networks to waves of noise that, in their opinion, resemble something about the brainwaves during sleep. And, then apparently the networks are able to learn more effectively. What I suspect might actually be what they're doing is they're just adding random perturbation to the weights after some training which is a standard technique. And, they just happen to like one particular way of adding noise to the network.

Robert J. Marks:

You know, that's what struck me, too. There's a method in training neural networks called simulated annealing wherein you do basically add noise into the training process to make it much more effective. And, there's other things such as weight saturation avoidance, where all of the weights, all of the interconnects, are so big that they saturate each of the neurons. And, so you have to back them off a little bit. So, you have to halt your training in order to back these things off.

Robert J. Marks:

But. These are problems which have been known for 30 or 40 years. These are techniques which people have practiced for a heck of a long time. And, this is an example of what I refer to as seductive semantics. It's like you said, Eric. They are trying to make this thing sound more human, and they do that by trying to relate it to human attributes when the relationship really isn't there, is it?

Eric Holloway:

Right, right.

Robert J. Marks:

So, it's very, yeah, it's very frustrating. Look, we have been working our way through the dirty dozen AI hype list with Bradley Center Brain Trust members, Eric Holloway and Jonathan Bartlett. We will continue our final countdown next time on Mind Matters News. Until then, be of good cheer.

Announcer:

This has been Mind Matters News with your host Robert J. Marks. Explore more @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.