Yoram Wurmser, and Jacob Bourne

The Daily: Is the AI train already losing steam, getting rid of 'hallucinations', and where AI is actually taking us

Audio

On today's podcast episode, we discuss the reasons that the AI train might already be slowing down, how to get rid of AI 'hallucinations', and where the AI boom is taking us. Tune in to the discussion with host Marcus Johnson and our analysts Jacob Bourne and Yory Wurmser.

Subscribe to the "Behind the Numbers" podcast on Apple Podcasts*,* Spotify*,* Pandora*, Stitcher,* YouTube*, Podbean or wherever you listen to podcasts.* Follow us on Instagram

Episode Transcript:

Marcus Johnson (00:00):

Partner with eMarketer on data-driven marketing materials. Our custom reports give eMarketer Media Solutions clients the opportunity to generate new category insights through original surveys and analysis. Visit emarketer.com/advertise to learn more.

Yory Wurmser (00:16):

And those should be cheaper to run as well and have more direct impact. I think the problem is that it's hard to see how to apply these large models either in a cost-effective way or in a way that significantly increases productivity yet, but I think we're moving towards that and we're not super far away from that.

Marcus Johnson (00:39):

Hey gang, it's Monday, July the 50th. Yory, Jabob, and listeners, welcome to the Behind the Numbers Daily, an eMarketer podcast. I'm Marcus. Today, I'm joined by two folks. Let's start with Yory Wurmser. He is one of our principal analysts covering everything, advertising, media and technology based in New Jersey.

Yory Wurmser (00:58):

Hey Marcus, how are you?

Marcus Johnson (01:00):

Hey fella. Very good. How are we doing?

Yory Wurmser (01:01):

I'm doing great.

Marcus Johnson (01:02):

Very nice. We're also joined by one of our technology analysts. He is based in California and we call him Jacob Bourne.

Jacob Bourne (01:11):

Hey Marcus, thanks for having me this morning.

Marcus Johnson (01:13):

Yes sir. Thank you for being here. Today's fact is where we start. The least common job in America according to the Bureau of Labor statistics, you'd think podcast host, you probably wouldn't actually. Everyone's doing it. Everyone's doing it. Any guesses on the least common job? They've got a top 10 here. These are jobs with basically a thousand people doing them or

less in the whole country. Average salaries, most of them about 52,000. I wouldn't have guessed any of these.

Jacob Bourne (01:47):

A full-service gas station attendant?

Marcus Johnson (01:49):

Full-service?

Jacob Bourne (01:50):

Yeah.

Marcus Johnson (01:51):

Oh, is this where you pull up and they come out?

Jacob Bourne (01:53):

Yeah. And there's a couple states where it's still, I guess mandatory or something.

Marcus Johnson (01:59):

Jersey right?

Yory Wurmser (02:00):

I think New Jersey is the only one. I think Oregon was also, but I think they're getting rid of it.

Marcus Johnson (02:04):

That's a fun game to try to get out before they get there and do it yourself. Do they fight you for it? Am I like, "I'm going to pump it myself," and they're trying to snatch the nozzle from my hand? What happens if you just do it yourself?

Yory Wurmser (02:17):

You know, I've never tried it.

Marcus Johnson (02:20):

Yory, report back. No, don't. You'll probably go to jail. So the top one, wood pattern makers, people making patterns out of wood, 260 in the whole country.

Jacob Bourne (02:36):

I think it would be more than that.

Marcus Johnson (02:37):

Yeah. Then you've got clock and time precision technicians, 400. Farm labor contractors, 460. Anyway, today's real topic. Where is the AI train taking us? So for this episode, there are a couple of articles we were looking at, looking at where AI is taking us and how it's taking us there. The first one we wanted to talk about was a piece by Christopher Mims questioning whether the AI revolution is already losing steam.

(03:11):

He thinks so. Mr. Mims of the journal writing in a recent article that things look good in the headlines, in video reporting, eye-popping revenue. Elon Musk announcing human level AI is coming next year and big tech buying AI-powered chips like there's no tomorrow. It seems like the AI hype train is just leaving the station and we should all hop aboard. However, he argues that significant disappointment may be on the horizon, both in terms of what AI can do and the returns it will generate for investors because AI improvement is slowing and there are fewer applications for AI than originally imagined, but Jacob, we'll start with you. What do you make of this theory that the AI train is already losing steam?

Jacob Bourne (03:51):

Yeah, I don't really think it is actually. I mean, I think what we're seeing is the consequences of overhyping the technology. The AI revolution was never going to happen overnight or in a couple of years. It's going to take longer to play out. And if you consider that Google invented the transformer, which underpins current generative AI models in 2017, well, ChatGPT didn't launch until 2022. That's a five-year gap and we had a pandemic during that time.

(04:20):

So I think that's more the pace that we're going to see. I think progress takes time and we're going to see other AI breakthroughs, but it's not necessarily going to happen this year. And if

it doesn't happen this year, I don't think it means it's losing steam. It's just not a linear process. And I think in terms of the tech industry spending 50 billion on NVIDIA's chips while in 2023 while only getting 3 billion in revenue, I don't think they expected to make up those costs that year. I mean, they're playing the long game. They want to beat each other to developing artificial intelligence and it's just capital expenditures that they need to spend in order to get there eventually. But yeah, this is not an overnight revolution.

Marcus Johnson (05:01):

Yeah, that's quite a capital estimate that you just mentioned, is an interesting one, 50 billion into NVIDIA chips and gen AI startups only making 3 billion. That's a kind of similarly disproportionate ratio as OpenAI's revenue versus valuation because it doesn't disclose revenue, but the Financial Times in December was saying it's probably making about 2 billion. It wants to try to double that in a few years, but that's a far cry from its valuation, which is closer to 90 billion. So that is how things typically start. Mr. Mims was saying though, what really matters for the long-term health of the industry is how much it costs to run AIs saying, "Costs of running popular services that rely on gen AI far exceed the already eye watering cost of training it." So is that more of a concern here is that we haven't even got to the big expenditures because the one thing to train it is another thing to pay money to run these systems?

Yory Wurmser (05:53):

So I think there are a couple of things there. I mean, first of all, I think I agree overall with Jacob that it was overhyped and I think it's not meeting that extraordinary hype, but the promise is still really strong. I think the costs, some of the innovations that we're seeing in the past half year or so are going to address that cost issue. The on device models, so models that work on your phone or on smaller servers because they're smaller but still pretty powerful, I think that's going to make a big difference in terms of cost. And a lot of the hype was around these giant models, but I think we might get some more specialized models for industries to have very specific purposes and those should be cheaper to run as well and have more direct impact. I think the problem is that it's hard to see how to apply these large models either in a cost-effective way or in a way that significantly increases productivity yet. But I think we're moving towards that and we're not super far away from that.

Jacob Bourne (06:49):

Yeah. And I think along those same lines, in addition to the smaller models, I think we're going to see more efficient chips be developed too, so powerful efficient chips that can run bigger models but for more cheaply. So I think that's the other thing that... other kind of innovation that's going to make a big difference.

Marcus Johnson (07:08):

One of the points in this piece from Mr. Mims was that others are catching up with OpenAI and that that's a sign of the industry slowing down. He was saying, "Further evidence of the slowdown in improvement of AIs can be found in research showing that the gap between the performance of various AI models is closing Meta, Mistral," which is a French company in AI, logging similar scores on tests of their abilities saying a mature technology is one where everyone knows how to build it. My first question is are they? Do you feel like these companies are catching up with the leader, with OpenAI? And then second part to that is what will the development trajectory look like? Is it going to be kind of step changes every time the leader releases a new model? Are we going to see breakthrough spikes? What is the kind of growth of AI look like in terms of companies each trying... like in a horse race each trying to pull ahead?

Jacob Bourne (08:03):

Yeah, I don't know if that argument is really airtight in terms of showing that things are slowing. I mean, there's not really that many companies that are rivaling OpenAI first of all. I mean, Meta is a long time AI developer, so it's not like when OpenAI released ChatGPT that Meta started building AI models. They had been doing this for many years. Same thing with Google. Anthropic was founded by someone who worked for OpenAI. So I think that these companies are just... they have a similar set of resources and knowledge base as OpenAI and they have models that are competitive and it hasn't taken that long for them to catch up with OpenAI. But really there's not like that many companies I think that are really close rivals at this point.

Yory Wurmser (08:50):

I mean, I think it's a valid point that some of these models are getting fairly comparable and it might lead to a commoditization of these AI models or at least some applications. So in terms of the market value of some of these companies, I think that is a little trickier and more debatable whether to hold onto that market value. But in terms of the applications and

whether that means the development trajectory is slowing down, I don't think that those are the same thing. And I think the development trajectory is slowing down, but for different reasons than it's becoming commoditized.

Marcus Johnson (09:24):

Mm-hmm. The one area it might be slowing down is maybe on the consumer side. And by that I mean gen AI user growth according to our estimates is already losing steam. So maybe that's part of this. We estimate 29% of Americans already use generative AI, that will grow to 34% next year and 37% the year after that. Folks using it for work, one-third of gen AI users will use it for work climbing to half by 2026, but that's still 17% of the whole population using gen AI for work in two years time. So those numbers aren't earth-shattering and they are starting to slow. But also 29% of the population this year using gen AI, there's 100 million people. So there's still a lot of folks.

Yory Wurmser (10:08):

No, but those numbers and Jacob knows these numbers better than I do, they exclude things where generative AI is being used in the background, search or your photo application on your phone, things like that, Snapchat with generating 3D images. There's a ton of use of generative AI where consumers aren't going in and saying, "Okay, ChatGPT, give me an answer." So it's becoming more pervasive without it being on the forefront of consumers' recognition that it's becoming pervasive.

Marcus Johnson (10:45):

Mm-hmm.

Jacob Bourne (10:45):

Yeah. And I agree and I think that's the direction we're headed in. Really it's just more deeper integration where you're using it and you're not even thinking about it. So it's not that it's losing steam or significance, it's just that it's not at the forefront of everybody's minds because it's faded in the background.

Marcus Johnson (11:02):

One of the thing that is at the forefront of everyone's minds, oh, it does come up quite a lot, are hallucinations where you ask a question to a generative AI model, it doesn't know the

answer and so it just spits anything back at you, which ends up being an incorrect answer. And Kelsey Piper of Vox wrote a piece titled Where AI Predictions Go Wrong. She was writing that there is one school of thought on large language models, LLMs like ChatGPT, that as we run larger and larger training runs and learn more about how to fine tune and prompt them, their notorious errors which are hallucinations will largely go away. However, Yann LeCun, Facebook's head of AI research, and Gary Marcus, an NYU professor disagree, arguing that some of the flaws in LLMs, the difficulty with logical reasoning tasks are not vanishing with scale. They expect diminishing returns scale in the future and say, we probably won't get to fully AGI, artificial general intelligence, reasoning like a person by just doubling down on our current methods with billions more dollars. Yory, what'd you make of this debate?

Yory Wurmser (12:05):

I'd probably come down more on the side of Yann LeCun and the people who are skeptical about how close we are to AGI. I think that these models are going to get better and better at interpreting and limiting hallucinations. But in terms of full reasoning, I'm not sure there's a linear pathway from larger data sets to the human reasoning. I think there's a lot more going on in human reasoning like planning and things like that which aren't really accounted for in the current type of models we're talking about. So I think that the leap to AGI is probably a little further away than a lot of people think. But at the same time, I think the hallucinations are going to reduce substantially as these models get better, not necessarily bigger but better.

Jacob Bourne (12:46):

Yeah. Yeah, I mean I agree with Yann LeCun as well, and I think his point about that the future of AGI is probably not LLMs period might be accurate. I mean, LLMs get a lot of the focus because that's what ChatGPT is, but there are other types of advanced neural networks being developed and we might see breakthroughs happen there. We also might see breakthroughs happen with just the linking up together of several models that had different strengths working in tandem to try and correct some of these hallucinations or just have more powerful reasoning abilities.

Marcus Johnson (13:22):

Mm-hmm. See, on that note, thinking about the future and what we can expect from this technology, there was a piece from Josh Mitchell of The Wall Street Journal titled, Where is the AI Boom Taking Us, he was citing Aidan Gomez chief executive of AI enterprise platform,

Cohere, who says, "AI will do everything from choose your shampoo to half the hours that doctors spend transcribing notes," saying it will also ultimately boost living standards, but could be used for nefarious purposes like swaying an election. So it's capable of a lot, it seems good and bad, but Jacob, I'll start with you. What do you is something that you are paying close attention to when it comes to where the AI boom is taking us and what it can do for humans?

Jacob Bourne (14:09):

Well, kind of reiterate what I had said before, I think where we're going mostly is deeper integration. We're going to see it more in social media, smartphones, search engines, AR/VR devices, e-commerce, everything. New Apple Intelligence I think is a great recent example of integration. So many people use Apple devices, now they're going to have generative AI operating in the background without thinking about it.

(14:32):

But I think in terms of where we're headed, it really depends mostly on decisions that people make around AI. And of course, like you just said, people make different decisions. Some people want to use it for nefarious purposes, others want to really get this higher level business productivity from it, others want to use it to fight disease and climate change. So I think it's really... I think we're headed to all those places and I think that's partly why we need some sensible regulation to kind of steer the ship, not so much that stifles innovation, but so that we don't have some of these very negative unintended consequences. And some of those consequences could be mass job losses, but so other people say that no, actually people are going to get pay raises due to being more productive at work using generative AI.

(15:16):

But I don't think anyone really has definitive answers for some of these questions. And that's why I think if we are seeing a little bit of a slowdown, it might not be a bad thing so that we can kind of get prepared as a society for some of these outcomes.

Marcus Johnson (15:29):

It does... I mean, it's such a hard question to answer and Ms. Piper of Vox, Yory, was saying it is hard to know the limits on what they'll be able to do being AI, LLMs before we've seen them. Equally, it's hard to confidently declare what capabilities they'll have. So part of the problem is

in two years time, we just don't know what's going to be possible. And so it's hard to predict what's going to happen in two years time because you don't know what you're going to be working with in two years time to build the future in two years after that.

Yory Wurmser (16:01):

Yeah, I mean someone in one of those articles that you mentioned add the analogy that we're trying to regulate spaceflight before we've built a rocket, and I think that's true. That's another way of saying what you just said. We don't know exactly the shape they'll be, but I think it's what Aidan Gomez says is completely accurate that first of all, that agents, AI agents, agentic AI is coming. But there's huge dangers also in terms of bots on social media platforms and in media, bots to mimic politicians, celebrities, individuals. I mean, there are all types of nightmarish scenarios that can come out of AI, all types of criminal security type of dangers. So the dangers are real. Regulation is going to be important. The exact shape of it is still to be determined though.

Marcus Johnson (16:50):

Mm-hmm. That's where we'll leave it for today. Thank you so much to my guests for hanging out with me today. As always, thank you to Jacob.

Jacob Bourne (16:56):

Been a pleasure, Marcus.

Marcus Johnson (16:57):

Yes, sir. Thank you to Yory.

Yory Wurmser (16:58):

Yeah, been great. Thanks.

Marcus Johnson (17:00):

Yes, indeed. Thanks to Victoria who edits the show, Stuart who runs the team, Sophie who does our social media, Lance for helping to produce this episode. And thanks to everyone for listening into the Behind the Numbers Daily, an eMarketer podcast. You can hang out with Rob Rubin tomorrow. He's the host of the Banking and Payment Show where he'll be speaking with Lauren Ashcraft and Jasmine Emberg, all about social media and banking. And then all of

these ones have less than a thousand, furnace and kiln repair technicians. I can't say this word, prosthodontists. If your tooth falls out of your mouth, they'll drill it back in. Wood model makers, private cooks-

Jacob Bourne (17:39):

I'm not sure about this data here, Marcus. A furnace repair technician really?

Marcus Johnson (17:43):

Yeah, just 500 of them. This is the Bureau of Labor statistics.

Jacob Bourne (17:47):

I don't know about this.

Marcus Johnson (17:49):

Me neither.

Yory Wurmser (17:49):

I had a great uncle who is a harp mover. That was his profession. He moved harps.

Jacob Bourne (17:54):

And apparently there's more than a thousand people [inaudible 00:17:58].

Marcus Johnson (17:58):

I know. Yeah. That does sound fun.

Yory Wurmser (17:59):

Doesn't make the top 10, I guess, right?

Marcus Johnson (18:01):

No, no. Industrial psychologists and pediatric surgeons is the last one. 1200 of them in the whole country. That's not enough. But there are more harp movers. You'll be glad to know.