EMARKETER

Marcus Johnson, Jacob Bourne, and Grace Harmon

The Daily: Why the California AI bill matters so much, what AI rules are missing, and more

Audio

On today's podcast episode, we discuss whether AI is more like cars or Google search, what's missing from the new AI bill, and what these new rules mean for the rest of the country. Tune into the discussion with host Marcus Johnson, and analysts Jacob Bourne and Grace Harmon.

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:

Grace Harmon:

I think that some of the concern about how much trust you want to put in AI depends on what can go wrong. If you think about what can go wrong with a Waymo versus what can go wrong with the robot umpires that the MLB has been testing, what can go wrong there in terms of harm is a grave difference.

Marcus Johnson:

Hey, gang. It's Monday, September 9th. Grace, Jacob and listeners, welcome to Behind the Numbers Daily: An EMARKETER Podcast. I'm Marcus. I'm joined by one of our analysts who writes for our Connectivity & Tech Briefing based in California. It's Grace Harmon.

Grace Harmon:

Hi, Marcus. Nice to be here.

Marcus Johnson:

Hello there. Thank you for joining us. And someone else in that very state on the West Coast. He is our technology analyst and we call him Jacob Vaughn.

Jacob Bourne:

Hey, Marcus. Thanks for having me today.

Marcus Johnson:

Yes, sir. Yes, indeed. Today's fact, so folks might know, might have heard the stat of we spend about a third of our lives sleeping. However, some other numbers on what we do with our time throughout our lives. We spend one year of our entire lives socializing. Just one-

Jacob Bourne:

That's it?

Marcus Johnson:

... compared to 11 years of our lives staring at screens, according to the Huffington Post article by Leigh Campbell.

Grace Harmon:

11 years is kind of terrible to think about.

Jacob Bourne:

And I guess the idea there is that amount of time on screens is probably going to increase as we socialize more with chatbots.

Marcus Johnson:

Oh, good. Thanks for that, Jacob. We also spend one year of our lives exercising Just one. And get this less than one year in school up to the end of high school. So because we only go to school as kids for a couple of hours a day when we're younger and then we go a little bit further into education. We go for a few months then we have a break then you get the summer off. When you total up all the actual hours that you spend at school before you graduate high school, just one year. So stop complaining children!

Jacob Bourne:

A couple of hours per day at school? I don't know what schools you went to Marcus but that sounds like a pretty... That's sweet, yeah.

Marcus Johnson:

The best kind. They weren't great.

Grace Harmon:

I wonder how much crazier that gets if you go to medical school or become a lawyer.

Marcus Johnson:

That's fair.

Grace Harmon:

Right.

Marcus Johnson:

Be about 50% of your life you spent at school and then one year practicing before you retire.

Jacob Bourne:

And then the other 50% is spent paying off the debt [inaudible 00:02:23].

Grace Harmon:

There you go.

Marcus Johnson:

That is true. Anyway, today's real topic, why the California AI bill matters so much.

All right, folks. Let's get to it. So what are we talking about? Well, the California legislature passed a sweeping AI safety bill. This is a story by a lot of people, but Wes Davis of the Verge, she was explaining that the California State Assembly and Senate have passed the Safe and Secure Innovation for Frontier AI Models Act, or SB 1047 as we might refer to it in this episode. It's one of the first significant regulations of AI in the US.

The bill obligates AI companies operating in California to implement some precautions before they train a sophisticated foundation model, but just the most expensive and powerful ones. In short, AI firms must test these large models for catastrophic safety risk before releasing them to the public. They also have to include a kill switch to be able to shut the model down quickly if it goes rogue. Greg Bensinger of Reuters also noting that the bill requires developers to hire third-party auditors to assess their safety practices and provide extra protections to whistleblowers speaking out against AI abuses.

The AI safety bill will now make its way to California Governor Gavin Newsom's desk for a signature to make it official. He has until the end of September to sign.

Jacob, I'll start with you. What to you is the most interesting part of this bill?

Jacob Bourne:

Yeah. I think the most interesting thing is really at the heart of it which is that it seeks to regulate essentially what are computer applications versus targeting people who might use those applications for their various purposes. So the application here is that there might be an inherent danger lurking within the most advanced AI systems themselves that's not just contingent upon how they're used. And I think the kill switch is the greatest point there, that it shows that there's concern about a future when AI gets so advanced that it could just run amok on its own.

Marcus Johnson:

I want to throw this question out there because Kelsey Piper of Vox was saying, to talk about your point of the danger lurking within these models, is AI more like cars or Google... or Google Search, basically?

So here's what she's saying, "So if I build a car that is far more dangerous than other cars, don't do any safety testing, release it, and it ultimately leads to people getting killed, I would probably be held liable and have to pay damages if not criminal penalties. However, if I build a search engine that, unlike Google, has as its first result, 'How can I commit a mass murder?' Detailed instructions on how to best carry out a killing spree and someone uses my search engine, follows the instructions, I likely won't be held liable thanks largely to Section 230 of the Communications Decency Act of '96."

So here's the question. She says, "Is an AI assistant more like a car where we can expect manufacturers to do safety testing or be liable if they get people killed? Or is it more like a search engine?"

Grace Harmon:

I think there's one part of it that's a little bit vague in terms of the wording, which is one thing the bill's addressing is the really low-probability, high-risk situations.

Marcus Johnson:

Yes.

Grace Harmon:

So looking at the bill, one of the things it mentions is the consequences for the creation of... I think it's like a chemical, biological or nuclear weapon that causes mass casualties. That wouldn't be physically manufactured by the AI, but to your point of, "Could you go into Google and search how to make a weapon?" There is the degree there of you just say creation. There is the implied creation would be done by a human being.

So I think there is still points where they're a little bit vague about who would be liable. And with the case of the Telegram CEO being arrested last month, one of the points they were making is that a platform shouldn't be liable for misuse, the user should be liable for their own misuse. I think that the bill is teetering on either edge. You need to be able to create a model that will not allow those things to happen.

Marcus Johnson:

What do you make of the fact that they're going after that, some people have said, kind of sci-fi-esque AI taking-over-the-world issues, and maybe not addressing some of the other AI issues as a priority.

Jacob Bourne:

Well, I think there's maybe a misconception there because there's a slew of AI bills on Gavin Newsom's desk-

Marcus Johnson:

True.

Jacob Bourne:

... right now that address those other issues, or at least some of them. So this particular SB 1047 is... Yeah, it's looking at these advanced frontier models themselves and it's a bit forward-looking in terms of addressing some concerns among many prominent AI researchers that, as they become more powerful, the risk for things really going awry increase. Now one interesting thing around this is that one of the opponents of the bill, Meta's Chief AI Scientist, Yann LeCun, says that this mass casualty event is science fiction. But it's interesting then... Well, if it's science fiction then I guess there's no need for AI companies to worry about liability from the bill.

Marcus Johnson:

That's true. So that was one of the points that jumped out to me, is how divided folks are on this. Because one of the three godfathers of AI or machine learning, Yann LeCun, who you mentioned is skeptical, as you pointed out, of AI's catastrophic potential. Whilst the other two, Yoshua Bengio and Geoffrey Hinton are more concerned and welcome the new bill. AI company, Anthropic, is cautiously supportive. OpenAI says, "No thank you." So this is a rather... Controversial is probably a bit of a strong word, but you do have people on both sides of this bill for and against it.

But to push back a bit on the idea that this bill does prevent or will significantly prevent a cataclysmic event from happening, I'm wondering if the rules are a too-little, too-late measure, because the bill says that you have to test them rigorously, these models. But Greg Bensinger of Reuters, again, was saying, "The new bill would also give the state attorney general the power to sue if developers are not compliant, particularly in the event of an ongoing threat such as the AI taking over government systems like the power grid." If AI has taken over the power grid I think you have bigger problems on your hands than making sure to sue the company that designed rogue AI.

Jacob Bourne:

Yeah. Well, it's supposed to be a deterrent, right?

Grace Harmon:

Yeah.

Jacob Bourne:

We don't want AI taking over the power grid at any point.

Marcus Johnson:

No.

Jacob Bourne:

And so if there's liability for the companies developing a model that might do that, then they might be careful not to do that.

Marcus Johnson:

Yeah.

Grace Harmon:

Yeah. I think a lot of the bill is about liability. I mean, you can't necessarily stop all these things from happening but you can put enough of a burden of fear, of consequence, of financial consequence that those companies are a lot more careful.

Marcus Johnson:

Yeah.

Jacob Bourne:

Yeah.

Grace Harmon:

And to your point about Anthropic getting on board, part of that was that the bill directly took on suggestions from Anthropic. So they aren't necessarily being so stringent in the rules and policies they set out before that they aren't willing to go back and change anything.

And one change I know they made was about prioritizing open source development. And that liability if a smaller startup takes an AI model developed by X, Y or Z AI giant, and that AI giant makes some sort of error and makes it so that model can do something dangerous, if the developer, the smaller developer has spent, I think something under... Jacob, is it maybe... I think it's $100 million or $10 million.

Jacob Bourne:

It's $100 million-

Grace Harmon:

$100 million.

Jacob Bourne:

... which currently is a pretty high threshold, but-

Grace Harmon:

It is, yeah. If they've spent less than that fine-tuning it and the AI giant who created the original model makes a mistake, then that smaller developer isn't as liable. So I do think there's points of deterrence but there's also points where they're not trying to punish absolutely everyone.

Jacob Bourne:

Right. And I think Grace made a good point bringing up Anthropic because the bill in its original form was a lot stricter and it got watered down to these amendments, which made Anthropic happy. But I would say that there are people in the AI doomer community, so-called doomer community, who think that this bill is not going to stop a catastrophe because if AI companies successfully build some type of super intelligence, then the safety testing is not going to detect it. It's going go undetected.

Grace Harmon:

Well, there's also some points where they can walk away. I think it was related to a bill in Europe. It was not related to Senate Bill 1047, but Sam Altman had said something about... something that was being proposed. They would try to comply but if they couldn't comply they would just stop operating in Europe. And I do think that that's a risk here is California has 35-ish of the world's top AI companies, but if these companies don't figure out how to comply here or too scared of not complying here, they could stop operating here. And that's a pretty big endeavor. That isn't just moving an office.

Jacob Bourne:

Yeah, the bill... Jacob, you were saying... got watered down. A few things here. The bill no longer states that AI companies can be sued before any harm has occurred. They can't be sued for criminal charges, just civil. And they also took out this new AI agency. So Mr. Scott Wiener, co-author of the bill, scrapped a proposal for new agency dedicated to AI safety to allegedly appease some of these big tech folks. So there is definitely some things that have been taken out of the original bill.

I want to circle back to this idea of whether AI can even be regulated given how amorphous it is. Because Alison Snyder of Axios is writing, "Unlike computer programs, they use a set of rules to produce the same output each time they are given one input, gen AI models find patterns in vast amounts of data and produce multiple possible answers from a single output. The internal mechanics of how an AI model arrives at those answers aren't visible, leading many researchers to describe them as black box systems."

And she cites Peter Lee, President of Microsoft Research. She says, "The unknowns about what happens in a large language model between input and output echo observations in other areas of science where there is an inexplicable middle, as it's called."

What do we make of this idea that because AI is so amorphous and we can't get our arms around it as a concept or in terms of trying to rationally think through all the different computations or possibilities of what the output might be. And what happens in that inexplicable middle between input and output, that the idea of really regulating it is... What's the right word? Not fiction.

Jacob Bourne:

Impossible.

Marcus Johnson:

A bit of a fantasy. Yeah, exactly.

Jacob Bourne:

Yeah, I mean, two things. One, I think it's because of this amorphous aspect that that's where the risk is, right? Because you don't really fully know what it's going to do. But I think it's really more, again, that gap between what goes in and then what comes out of the model that's amorphous. But the technology that builds AI is not amorphous. I mean it's really three components. It's AI chips, it's algorithms, and then it's data. And when you look at those ingredients, then you can see that there is probably a pathway to sensible regulation by regulating those three things that make them up.

Grace Harmon:

There's also the human element that this is trying to address as well. And that is actually the human element of what people can do to make the models better because this is including some whistleblower protections. So it isn't always going to be the executives at the company that can notice when something's going wrong with the model when there's a vulnerability, when there's something that needs to be tightened up. I mean, it's especially relevant considering everything that's going on with outcries from former OpenAI employees. I think they only got rid of some of their policies against whistleblowers earlier this year. So I do think that that's an important part of it. You know, like Jacob said, "It's not entirely vague how these are being created. The prompts are written."

Jacob Bourne:

Grace, I think that's a great point to bring up about the whistleblower faction because actually, right now, there's a report that OpenAI is looking to hire people to investigate its own employees because it's worried about this internal threat. It could be things getting leaked, but it also potentially could be a whistleblower or things of that nature showing that there's... I mean, there's a lot of risk in what they're doing, and I think there is some acknowledgement, implicit acknowledgement, of that in trying to keep things under wraps at the organization.

Grace Harmon:

Bound to have those spheres of leaks too. Because, I mean, Anthropic put out its prompts and now you can go back and see over time how they've changed what the Claude models are told that they can and can't do. And some of the things are kind of benign. It's not allowed to apologize, and some of them are a little scarier in wording. It's supposed to act as if it's face blind. It's supposed to act as if it doesn't have political views. So there's the fear of leaks but, to a degree, I think those leaks help to create more brand and consumer trust.

Marcus Johnson:

What do you both make of this hurdle for AI? Because there's another line from Mr. Lee of Microsoft Research, and he's basically talking about, what I interpreted as, "We are okay with humans being flawed, but not computers." And he says that with GenAI we are now for the first time allowing ourselves to fantasize about the possibility of computers doing highly-skilled knowledge work. "That sort of work," he adds, "isn't about achieving perfection, but being effective and trustworthy."

However, I don't know if people are comfortable with computers just being effective and trustworthy, and not necessarily being perfect. I liken it to self-driving cars. It seems like it's very hard for people to trust self-driving cars. If you look at the numbers, 40,000 Americans killed every year by climbing into a car with human drivers. 94% of motor vehicle crashes are caused by driver error, according to National Highway Traffic Safety Administration. And about a third of serious car crashes could be prevented by autonomous vehicles, according to a study published in the Journal of Safety Research.

Despite this people trust driverless cars less, I think, because they want them to be perfect, not just effective and trustworthy. Two quick numbers on that, according to Forbes Advisor survey, just 12% of people are trusting of driverless cars versus 46% of Americans who are either very or somewhat untrusting. And according to a AAA survey, 9% of people, so similar share of people, would trust a car that drove itself, with 66% saying they're afraid of self-driving cars. So what I'm saying here is I think people want their AI to be perfect. I don't know if that's possible. I don't think they're going to accept just effective and trustworthy as a compromise.

Jacob Bourne:

Yeah. A couple of things. One is in terms of this SB 1047, the focus really is on models that don't even really exist yet. AI companies want to make really powerful AI, like agentic AI, that can do things and operate in the background on its own. And that's where the concern comes in, that it's potentially... If they're successful, it could do things that are beyond the capabilities of people like shutting down the power grid in a country, for example, or something like that. And so that's really part of what the bill is concerned about.

The other thing is... I mean, AI is not an individual. When we're thinking about people driving cars, it's like, "Well, okay, do you trust a drunk driver more than AI driving a car?" Well, probably not. But we do trust people who are competent and sober driving cars, right? So I think we trust individual people who are in the right frame of mind to drive a car. But AI is not an individual, it's a system. And the system might have an inherent flaw that maybe a person might not.

Grace Harmon:

I think that some of the concern about how much trust you want to put in AI depends on what can go wrong. If you think about what can go wrong with a Waymo versus what can go wrong with the robot umpires that the MLB has been testing, what can go wrong there in terms of harm is a grave difference. And I think Jacob had a great point that we're not necessarily thinking right now about what can Claude do so wrong that it would cause massive harm? It's what can these models do in three, four, five, 10 years?

So I think it's good that we're starting to think about now and actually write out what could really go wrong. Again, like a creation of a chemical or biological weapon that can't really be done right now but if we don't start talking about it now, then I think it's going to be a lot harder to implement rules for liability later on.

Marcus Johnson:

It seems like California residents would like AI to be regulated, found one survey. 70%, seven zero percent, of residents were in favor of the bill, with even higher approval ratings among Californians working in tech. That's according to an AI Policy Institute survey. I think the institution was in favor of this bill. Let's end quickly with what Cecilia Kang of the New York Times is writing, which was that the new bill might set the national standard for regulating the new tech. Grace, how likely is it that this new AI bill spreads across the country from California?

Grace Harmon:

I think it's unlikely that it happens fast because it's just not that big of a priority for that many other states, again, with the concentration of AI companies that are here. I mean, it just wouldn't be as relevant in a lot of other states. Maybe Washington, New York, maybe Maryland or Michigan... because they have a lot of tech companies. But I just don't think there's going to be that many states that feel the need to hop on this as quickly. And I guess that would change depending on if we're talking about companies that can be held liable for incidents in other states. But in terms of where the developers actually are and where the tech is coming out of, I think California... It makes sense that we're... I'm not going to say "we're" because not everyone lives here. It makes sense that California is hopping on this so quick. I could see it spreading. I just don't see it being a priority for senates and assemblies when there are other things on the table.

Marcus Johnson:

Jacob, where do you land?

Jacob Bourne:

Yeah. First of all, we have to see if it passes. It's not a done deal. We might see Newsom approve the other AI bills, sign the other AI bills, but not this one. I think the more specific AI bills that deal with things like bias and autonomous vehicles and things like that, I can see other states certainly crafting similar bills to those. If California does pass SB 1047, I think there will be some states that do something similar. Other countries might actually look to craft similar legislation using it as a framework. I think what's going to be slow is the US federal government because it is concerned about being less competitive compared to China and other countries on AI specifically.

Marcus Johnson:

I mean to your point that you both were making, it can be found in other parts of the country. Belle Lin of Wall Street Journal was noting that hundreds of other AI bills are in state legislatures with some 30 bills in California alone. So it's around the country. How far along those bills are and how likely they are to get signed is another matter.

Jacob Bourne:

On that note, a law in Illinois is actually going to take effect in January-

Marcus Johnson:

Oh, interesting.

Jacob Bourne:

... that protects employees and job applicants against being discriminated against by AI systems. So that's already a done deal and I bet we're going to see more. But again, that has nothing to do with frontier models and the development. That has to do with a specific end use of an AI model.

Marcus Johnson:

Yeah, lots of different parts of this.

Grace Harmon:

Some of the states that you would think would be a lot more laissez-faire about tech deployment and tech development are actually being a lot more strict. I mean, Florida's rule about no one under 14 can use social media, it's starting next year. That is a state that you would maybe in concept think would not be hopping as quickly on creating these regulatory rules. They are.

Marcus Johnson:

Yeah. If you zoom right out, it seems like Americans wouldn't mind a bill that holds AI companies accountable. Circling back to the original purpose of this SB 1047 bill in California, 73% of voters believing AI companies should be held liable for harms from technology they create, compared to just 11% that believe that they should not. It's a 2023 survey from, again, the AI Policy Institute.

That's all we've got time for this episode. I want to thank my guests for hanging out with me today. We thank Jacob.

Jacob Bourne:

Thanks, Marcus.

Marcus Johnson:

Yes, indeed. Thank you to Grace.

Grace Harmon:

Thank you for having me.

Marcus Johnson:

And thank you to Victoria who edits the show. Stuart runs the team. Sophie does our social media. Thanks to everyone for listening in to the Behind The Numbers Daily: An EMARKETER Podcast.

You can hang out with host, Rob Rubin, tomorrow on The Banking & Payments Show: An EMARKETER Podcast where he'll be speaking with Maria Elm and David Morris all about whether financial media networks can succeed.