AWS chief Adam Selipsky talks generative AI, Amazon's investment in Anthropic and cloud cost cutting

December 11 2023, by Haleluya Hadero

Credit: Tara Winstead from Pexels

Adam Selipsky is shepherding Amazon's cloud unit at one of the most important moments in tech history.

Selipsky, the CEO of the company's cloud computing unit AWS, has been behind the various generative AI offerings Amazon has rolled out over the past few months as it aims to compete with Microsoft and others in the growing AI arms race.

AWS is a leader in the cloud market and a deeply profitable business for Amazon. However, some of its growth has been slowing down the past few quarters, a problem attributed to customers cutting back on their spending due to challenges with the wider economy.

Simultaneously, the business has been at the forefront of Amazon's push into generative AI, which flooded the public consciousness last year with the release of OpenAI's popular chatbot ChatGPT. During a speech held late November at an AWS conference in Las Vegas, Selipsky unveiled the company's response (kind of) to ChatGPT—an [AI assistant for businesses called Q](#).

The Associated Press recently spoke with Selipsky about how companies are spending on [cloud services](#), Amazon's investment in the artificial intelligence startup Anthropic and the future of generative AI. The conversation has been edited for length and clarity.

Q: Companies have been cutting cloud spending this year. Is that still happening?

A: A lot of our customers over the past several quarters have been pursuing cost optimization. Since day one, we've said that AWS and the cloud are the place to do that. We've seen that a lot of customers have gotten far through that cost optimization. And we have other customers who are still in the middle of it. We're further through it, but it's not over yet.

We're also still seeing a lot of customers investing. The companies who are going to win are the ones who are investing now—in uncertain economic times—when some others are hesitating in their overall investments. And we're working with a lot of customers who are doing just that. We're also seeing tremendous interest in our generative AI offerings.

Q: What's your vision for generative AI?

A: We really think about three different layers of the generative AI stack.

At the bottom layer of the stack is the infrastructure required to to do generative AI. We have a very large Nvidia GPU-based business and have designed and delivered our own custom-designed chips, including our Trainium and Inferentia chips.

Most of our enterprise customers are not going to build models. Most of them want to use models that other people have built. And so that's the middle layer of the stack. We offer customers multiple foundation models with a service called Amazon Bedrock, including from Anthropic, Meta and Amazon itself. The idea that one company is going to be supplying all the models in the world, I think, is just not realistic. We've discovered that customers need to experiment and we are providing that service.

At the top layer of the stack is consuming applications that have been built using generative AI. And for that, we have a coding companion for developers.

Q: Speaking of models, there were reports that Amazon is building a large language model called

Olympus. Is that something that we should expect to see soon?

A: You should definitely expect to see multiple iterations of Amazon's first-party models, which are already out there today under the Titan brand. It goes back to the idea that there's no one model to rule them all. We want multiple models with different use cases. And I expect that they will collectively be very capable and very powerful.

Q: Can you chat with me about Amazon's investment in the artificial intelligence startup Anthropic? There are reports that Google, which is also backing Anthropic, is upping its investments. Some say this is becoming some sort of proxy war between Amazon and Google. Do you see it that way?

A: No, I don't. We have a very close, very tight relationship with Anthropic that's very beneficial to both companies. Anthropic has chosen Amazon as its primary cloud provider for its mission critical workloads. The majority of Anthropic workloads will run on AWS. Period.

Now, these aren't exclusive relationships. They have very large computing needs. and Anthropic has obviously used other cloud providers over time.

Anthropic has used AWS since it was founded in 2021. We've worked together and the relationship has deepened. Anthropic is training larger and more capable models. And they saw the opportunity to get from AWS a very large amount of compute capacity needed to train their models. It's also very important for Anthropic to be inside Amazon

Bedrock, which is our trusted service for accessing foundation models.

We realize that the folks in Anthropic are very smart and world experts in what they're doing. And by cooperating and collaborating together on training and running Anthropic's models on our chips, they're going to help us improve that technology. So it's really a mutually beneficial relationship where I think both companies and most importantly, our mutual customers, will really benefit for years to come.

Q: How does Amazon think about safeguards as its building this technology?

A: Responsible AI is incredibly important and something that Amazon has been taking very seriously. We have a number of principles for responsible AI that we've been public about. We've done things like create these cards for our services, which talk about the uses of the model, the intended use of the [model](#), about how they were trained. We try to provide more transparency into how some of these AI services are constructed and what they're used for.

We think that a lot of the solutions around responsible AI are going to need to be multilateral solutions. We need a collaboration between cloud industry leaders, folks like AWS, and those producing models, like Anthropic, as well as governments, academia and others. That's why we've been so active at participating in responsible AI forums at the White House and in the U.K.

Q: Where do you see the AI race going next year?

A: I think you're going to see a very rapid evolution and change. And that's partially reflective of the fact that we are still so early in the evolution of generative AI. That's why I think adaptability and flexibility

are actually incredibly important advantages for customers. In order for them to succeed with their business objectives and to delight their customers, they're going to need to be very flexible, agile and adaptable in how they evolve their use of generative AI.

Citation: AWS chief Adam Selipsky talks generative AI, Amazon's investment in Anthropic and cloud cost cutting (2023, December 11) retrieved 27 April 2024 from https://techxplore.com/news/2023-12-aws-chief-adam-selipsky-generative.html