

Amazon unveils their very own AI chip, "Inferentia"



Amazon.com launched their very own cloud computing microchip named "Inferentia" and end the dependency of Intel and Nvidia AI chips to perform high-end artificial intelligence computing operations.

Earlier, the AI chip market was ruled by Intel and Nvidia, and even Amazon was using their machine learning microchip for their cloud computation operations, especially after Amazon's cloud computing boomed. Earlier Intel was the major key player in the market of machine learning processors, which supported cloud computing, but after the launch of Nvidia's processors in September this year, Nvidia also has taken the market of machine learning processors for cloud computing. Yet, currently, the AI and cloud computing chip market is dominated by Intel's processors for machine learning inference, and analysts believe that Intel's worth will be around \$11.8 billion dollars by the year 2021.

However, with this announcement of Amazon's own AI microchip "Inferentia", Amazon will no longer be using the chips from Intel and Nvidia. Amazon's main purpose for developing their own custom machine learning chips is to avoid dependency on the chips from other companies and also customize them as per requirement. Also, Amazon is not the only one cloud computing vendors, Google is also run by its own artificial intelligence chip, introduced way back in 2016. Now, Google is running on its own AI chip, rather than using the chips from other companies.

About Amazon Inferentia, it will help researchers with inference, which means that the chip will use artificial intelligence algorithms to convert incoming audio to text, vice versa, and other such high-end operations. Besides this, Inferentia will also support hundreds of TOPS (Tera Operations per Second) which allow complex models to predict faster, and make the operations even faster. Using multiple Inferentia chips, Amazon will be able to deliver thousands of TOPS of throughput.