yahoo!

DATONG CHEN

Senior Director - Machine Learning Engineering & Data Science | Yahoo! Inc.

WHAT WERE YOUR MAIN OBJECTIVES BEHIND ATTENDING?

I wanted to gain insights into the future perspectives of AI as seen through the eyes of our hardware experts.

Furthermore, I am interested in understanding the anticipated hardware infrastructure that will be capable of accommodating the next generation of AI applications, especially "GenAI," over the next five years.

WHAT WERE YOUR KEY TAKEAWAYS FROM THE EVENT?

- Hardware experts at the Summit displayed a keen interest in AI and demonstrated a collective belief that AI represents the next significant opportunity in the technology landscape.
- Despite various emerging technologies, the GPU (Graphics Processing Unit) continues to stand out as the most promising hardware infrastructure within data centers for effectively accommodating GenAI applications in the foreseeable future.
- In the realm of Edge AI, the predominant focus remains on computer vision applications. However, a noticeable shift towards addressing recent breakthroughs in Large Language Models (LLMs) suggests that Edge AI is beginning to diversify its capabilities and adapt to evolving AI paradigms.
- Specialized, highly efficient hardware solutions designed specifically for tasks such as fine-tuning and inferencing transformers, including encoders and decoders, are still in development and hopefully expected to make significant strides in the near future. Innovations are highly expected to optimize AI processes even further than GPUs, enhancing their overall cost efficiency, edge deployment flexibility, and timely performance.

AI HARDWARE & EDGE AI SUMMIT

SEPTEMBER 2023 | SANTA CLARA, CA

WERE YOU ABLE TO CONNECT WITH LIKEMINDED COLLEAGUES AND NEW INDUSTRIES?

I'm delighted to have had the opportunity to engage with numerous esteemed hardware experts actively contributing to the advancement of AI chips and boards. One particular solution that has left a profound impression on me is GraphCore's innovative IPU technology, which undoubtedly holds significant promise in the AI hardware landscape.

“I'm delighted to have had the opportunity to engage with numerous esteemed hardware experts...”