



# **Arm Holdings:** **The rising challenger to Nvidia**

Presenter: Chris Sze



## Bloomberg's estimate of Generative AI

- Compound growth of 42% over the next 10 years.
- Generative AI has \$137 Billion USD Market in 2024.
- Generative AI will become a \$1.3 trillion USD Market by 2032, ten times increase in 8 years.

<https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds/>



## What's the big deal about AI?

- 99% of investors underestimate the market value of AI.
- Market research companies estimated that AI market size will grow 10 times within 8 years in the first phase, then grow another 10 times in the second phase in next 8 to 10 years.
- So, AI will grow 100x within the next 20 years.

## Altman's ambition



- OpenAI's Sam Altman is proposing a \$7 trillion USD investment for the semiconductor industry, to address the substantial shortage of AI hardware.
- The big question of any serious and ambitious investor will be - how to invest ahead of the AI curve?



## Case Study 1: Microsoft Maia AI Accelerator

The Azure Maia 100 and Cobalt are **ARM-based chips**. Maia 100 is a custom-designed chip by Microsoft that is optimized for artificial intelligence tasks and generative AI. It is based on the **ARM architecture**, Maia 100 is built on TSMC's 5nm process and features 105 billion transistors.

It is designed to run large language models, such as GPT-3.5 and GPT-4, that power some of Microsoft's AI services, such as Copilot.

Microsoft developed the Maia 100 chip in collaboration with OpenAI, and plans to integrate it into its Azure cloud platform in early 2024.



## Case Study 2: Amazon's Graviton processor

AWS Graviton is a family of processors designed by Amazon Web Services (AWS) to deliver the best price performance for cloud workloads running on Amazon EC2.

AWS Graviton processors are based on the **ARM architecture** and use less energy than comparable x86-64 processors.



## Case Study 3: Huawei AI chip and Kirin processors

Huawei AI chip using **ARM** is the Ascend series of processors, which are designed to provide high-performance and energy-efficient computing power for various artificial intelligence applications.

Huawei also produces other types of **ARM-based processors**, such as the Kirin series for smartphones and the Kunpeng series for servers.



## Case Study 4: Google Tensor Processing Unit (TPU)

Google Tensor processors feature a custom-designed Tensor Processing Unit (TPU), which is optimized for AI and machine learning tasks, such as image processing, language translation, speech synthesis.

Google's TPU used **ARM-based architecture**, which are designed for Google Cloud AI services.





## Case Study 5: META's two AI chips - MTIA & Artemis

Meta Facebook's custom **ARM-based** AI chip is called MTIA (Meta Training and Inference Accelerator), which is part of a family of chips that Meta is developing in-house with Arm to target AI training and inference workloads.

Meta is also working on another custom AI chip by **ARM**, called Artemis, which is expected to be deployed in its data centers in 2024. Artemis is designed to run large language models, such as GPT-4, that power some of Meta's AI services.

Lastly, META is the creator of the the popular PyTorch library which powered OpenAI.



## Case Study 6: Apple Silicon

Apple's M1 M2 M3 series of processors, which are designed by Apple and used **Arm's RISC instruction sets** to built integrated chipc for macs.

Apple Silicon processors are based on the **ARM architecture**, the latest M3 model uses 3-nanometer process technology. M3 processors feature a next-generation GPU and a faster CPU and Neural Engine, which are optimized for AI and deep learning tasks.



## Case Study 7: Samsung's AI mobile phone chip

Samsung's custom AI by [ARM](#) is the Exynos 2200, a system-on-chip (SoC) that powers Samsung's Galaxy S24 series smartphones. The Exynos 2200 features an octa-core CPU, a [Mali-G78](#) GPU, and a dual-core neural processing unit (NPU) that are based on the [ARM architecture](#).

The NPU is designed to accelerate various AI and machine learning tasks, such as image processing, face recognition, natural language processing, and more. The Exynos 2200 also supports 5G connectivity, 8K video recording, and 200MP camera resolution and more.



## So, what is Arm?

Arm is a company nearly wholly bought by Nvidia in 2022 for USD 40 billion, but blocked by UK government due to "National Security". It is now listed in Nasdaq.

Arm licensed its CPU / GPU instruction sets to technology companies to make their own chips. Arm has pricing powers but restrains it for now.

Arm Holding is the biggest challenger to Nvidia's dominance in AI. Now Arm's market cap is less than 1/10 of Nvidia.



## Real AI Companies

Nvidia

Arm Holdings

Taiwan Semiconductor Manufacturing Company (TSMC) 台積電

ASML Holding N.V. 艾司摩爾

AMD

Samsung

Lam Research

Applied Materials

Micron

Super Micro

Google

Microsoft

Huawei



## How large is the addressable market size of AI?

- 2006 World GDP: 51.79 Trillion USD
- 2024 World GDP: 109.73 Trillion USD (World Bank)
- We can roughly estimate that in 20 years time, world GDP will hit 200 Trillion USD, and more than half will be transformed by the AI revolution, including the oldest industry - agriculture.
- AI will go through a 100 times increase in dollar terms within 20 years.



## How to profit from Arm and AI?

The best game plan to profit from AI is setting up an AI fund to profit from OpenAI's Sam Altman USD 7 Trillion Initiative.

Placing bets before Middle East princes, private equities whales and sovereign wealth funds.

## About the presenter



Thank you very much for spending the time to read this presentation.

I am the founder and lead programmer of Sand Madison Group Limited, a company specialized in computational equity research.

I am looking for a lead investor to launch an AI fund, and to partner with prominent investors in Hong Kong to create a large capital base.

I would be glad to discuss the details in person. Please call me at +852 9649-5943 / 2866-2800 or email me at 3chris@gmail.com, chris@sandmadison.com.

<https://www.linkedin.com/in/chris-sze-61957723b/>