

# Why Power and Energy? And Why Now?

Jae-Won Chung  
December 2<sup>nd</sup>, 2025



ML.ENERGY



# Unprecedented Scale

Large

- Frontier models
- Kimi-K2 IT params
- Closed models could be larger

Numerous

- Specialized models
- Thousands of application use cases

Everywhere

- Mixture of models
- Deployed worldwide
- Serving millions

# Demand for Compute

More accelerators

## Zuckerberg's Meta Is Spending Billions to Buy 350,000 Nvidia H100 GPUs

In total, Meta will have the compute power equivalent to 600,000 Nvidia H100 GPUs to help it develop next-generation AI, says CEO Mark Zuckerberg.



By [Michael Kan](#) January 18, 2024

f X [Share](#) ...



(David Paul Morris/Bloomberg via Getty Images)

# Demand for Compute

More accelerators

## Zuckerberg's Meta Is Spending Billions to Buy 350,000 refrigerators

In total, Meta will have the compute power equivalent to 600,000 Nvidia H100 GPUs to help it develop next-generation AI, says CEO Mark Zuckerberg.



By [Michael Kan](#) January 18, 2024

f X [Share](#) ...



(David Paul Morris/Bloomberg via Getty Images)

# Demand for Compute

More accelerators  
Gigawatt datacenters

Zuckerberg says Meta will build data center the size of Manhattan in latest AI push

CEO says company plans to spend hundreds of billions on developing artificial intelligence products



Meta Founder and CEO Mark Zuckerberg speaks at LlamaCon 2025, an AI developer conference, in Menlo Park, California, on 29 April 2025. Photograph: Jeff Chiu/AP

Billions to

a H100 GPUs to help it

# Demand for Compute

More accelerators  
Gigawatt datacenters

Zuckerberg says Meta will build data center the size of Manhattan in latest AI push

Billions to

a H100 GPUs to help it

xAI could build 1.56GW natural gas power plant for new data center, campaigners claim

Report outlines plan for up to 90 turbines in Memphis

May 09, 2025 By: Matthew Gooding  Have your say



 Meta Founder and CEO Mark Zuckerberg speaks at LlamaCon 2025, an AI developer conference, in Menlo Park, California, on 29 April 2025. Photograph: Jeff Chiu/AP

# Demand for Compute

More accelerators  
Gigawatt datacenters

Zuckerberg...  
cent...  
push...  
  
xAI c...  
plan...  
claim...  
  
Report c...  
  
May 09, 2025

September 23, 2025 Company Global Affairs

## OpenAI, Oracle, and SoftBank expand Stargate with five new AI data center sites

New data centers put Stargate ahead of schedule to secure full \$500 billion, 10-gigawatt commitment by end of 2025.



Meta F...  
in Menlo Park, California, on 29 April 2025. Photograph: Jeff Chiu/AP

Context | 3

# Getting Power is the Challenge

We need the power now

- Next generation large frontier models were due yesterday

It takes *years* to procure large scale power

- Siting, planning, approval, construction, ...

# Who Would Care?

## AI Builders & Providers

Power capacity heavily influences  
research and service capacity

## Users & General Public

Concerns for grid stability  
and electricity price

## Hardware Vendors

No one can buy new hardware  
without the energy to run them

# Energy as a Critical Computing Resource

Session 1: How to accurately measure power & energy for ML

- Jae-Won Chung (University of Michigan)

Session 2: To maximize performance, we need to think about power

- Ahmet Inci (NVIDIA)

Session 3: Slashing ML energy without making ML useless

- Ruofan Wu (University of Michigan)

<https://ml.energy/tutorials/neurips25>

# Energy as a Critical Computing Resource

## Panel Discussion: A Deep Industry Perspective

- Bhargava Gopireddy (NVIDIA)
- Cooper Elsworth (Google)
- Arun Tejas Raghunath Rajan (Meta/MLCommons)

*<https://ml.energy/tutorials/neurips25>*