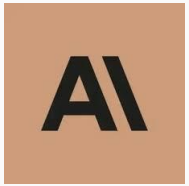
Cost of Building an LLMs

Presentation by:
Sankalpa Pokharel

What Is a Frontier AI Lab?

Frontier AI Labs: Research institutions focused on developing cutting-edge artificial intelligence models.

Examples of Frontier AI Labs:



TRAINING A MODEL LIKE CHATGPT IS LIKE LAUNCHING A SPACE PROGRAM

Reality of creating a State of the Art LLM

Training a model like ChatGPT is like launching a space program:

It requires multi-billion-dollar investment for

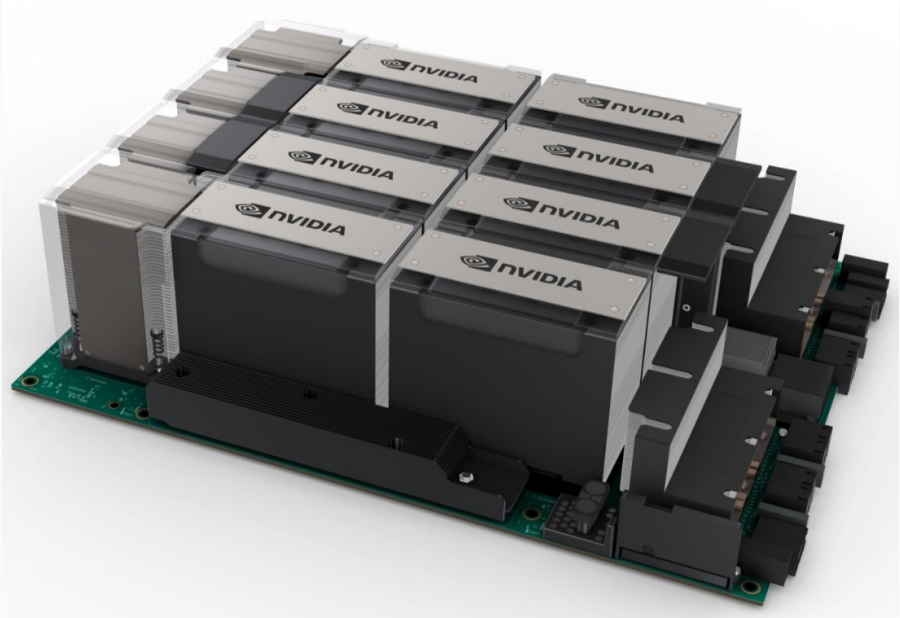
- Compute (the supercomputer).
- Data (the fuel)
- Talent (the astronauts)
- Data centers
- Networking
- Storage
- Training
- Other techniques

The Raw Ingredients

The Supercomputer

- OpenAI : Over 1 million GPUs
- Gemini : Over 100,000 TPUs
- Llama 3.1 : Over 16,000 GPUs

OpenAI has GPUs worth over
25 billion.

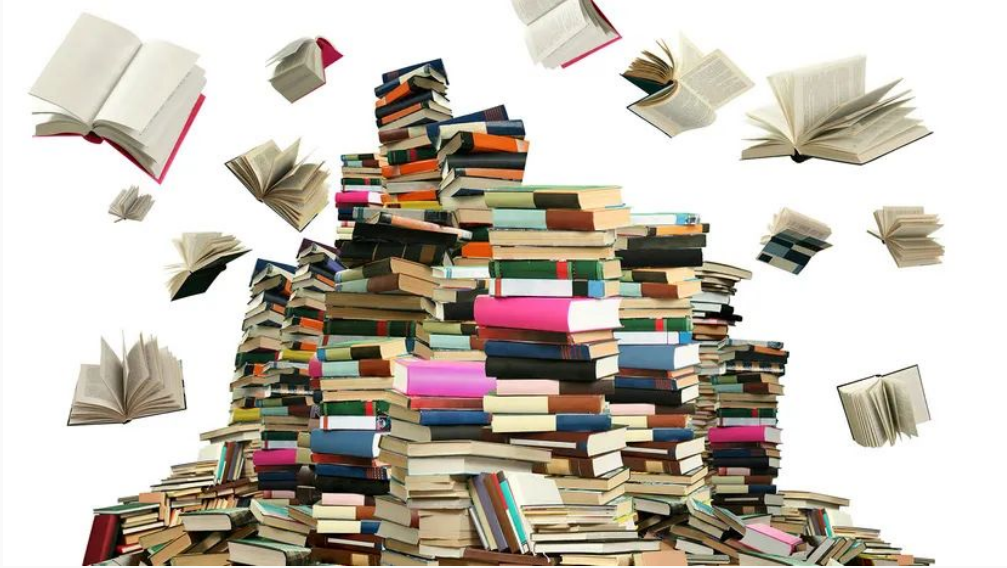


NVIDIA H100 cost $25,000

Data

Trillions of Words from:

- Books
- Web text
- Code



After data acquisition, these words go through rigorous steps of cleaning and licensing.

Data Acquisition and Processing can take more than $100 million.

Talent

Researchers, engineers and safety experts.
Top talent packages: $500K - $1M
Team Cost: $50M - $200M / year

Raw Ingredients Cost

Space Program:
$100 Million per rocket family

Infrastructure cost building an LLM:
$26 Billion

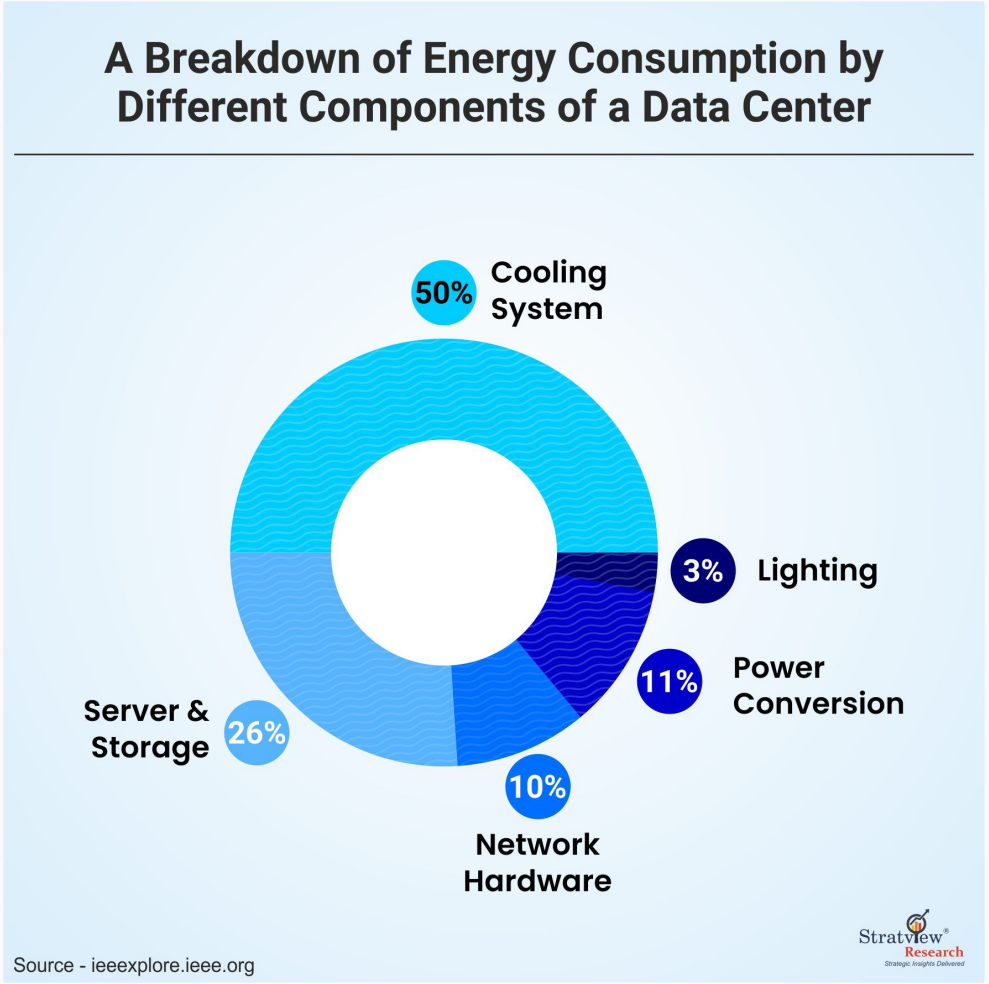
Infrastructure Demands

Data Centers

Power - Hundreds of megawatts.

Cooling - Liquid Systems for GPUs

Cost - $500M–$2B for facilities



OpenAI and Oracle partner for $500 billion StarGate datacenter project

Networking

High-Speed fiber and InfiniBand interconnects.

Cost: $100M–$300M

Storage

Peta bytes of high-throughput storage

Cost: $50M–$100M

Total Infrastructure Cost

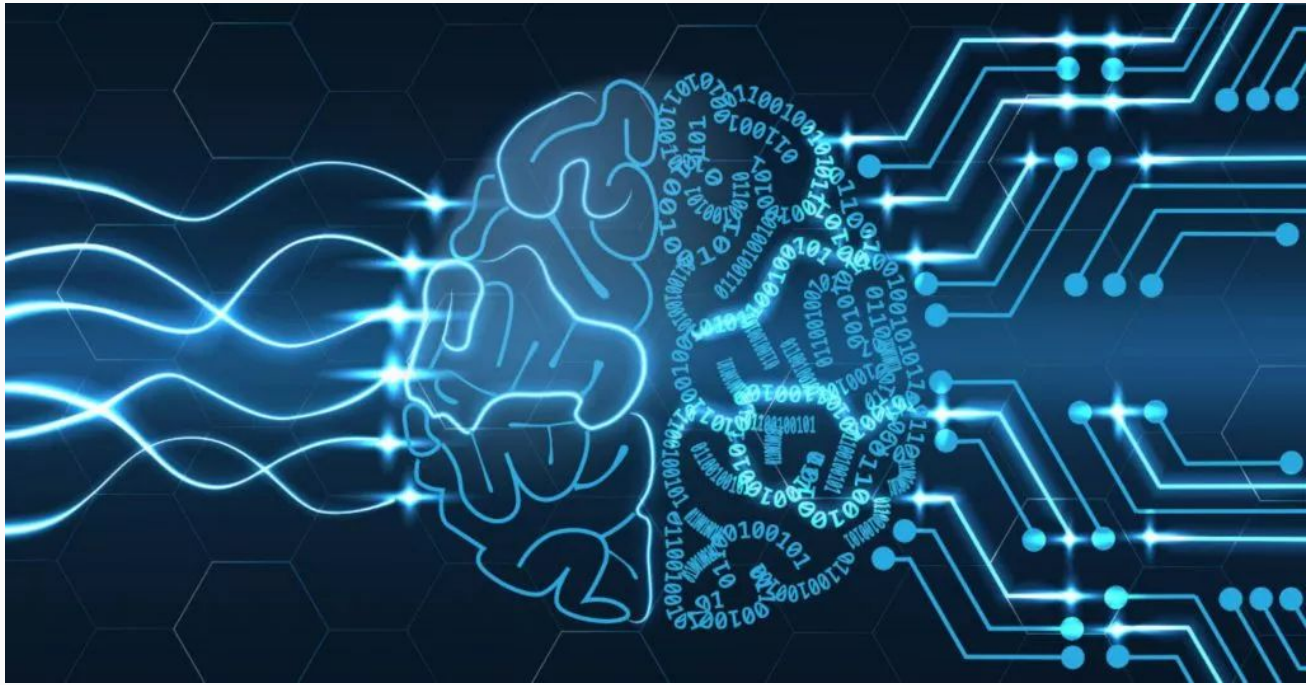
Launching Space Program:
$2Billion - $8 Billion

For Building an LLM:
$1Billion - $2.5 Billion

Training and Modern Techniques

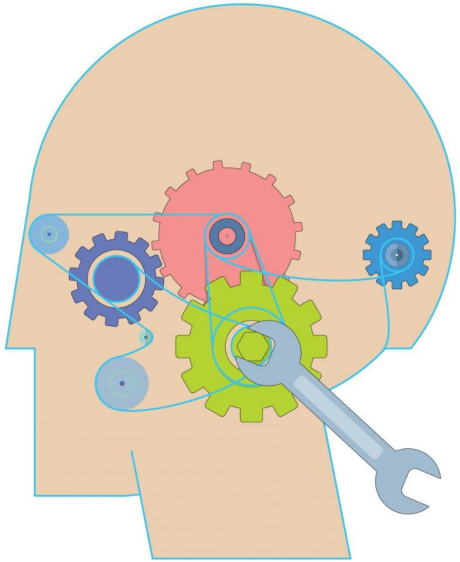
Pre-Training

The initial “Massive reading” of text data.



Fine-tuning & safety filters

Additional passes to adapt to domains, reduce bias.



Cost for Training and Modern Techniques

Launching Space Program:
$1Billion

For Building an LLM:
$40 Million - $50 Million

Hidden Multipliers

Retries

Failed experiments

Tens of millions per redo

Safety Teams

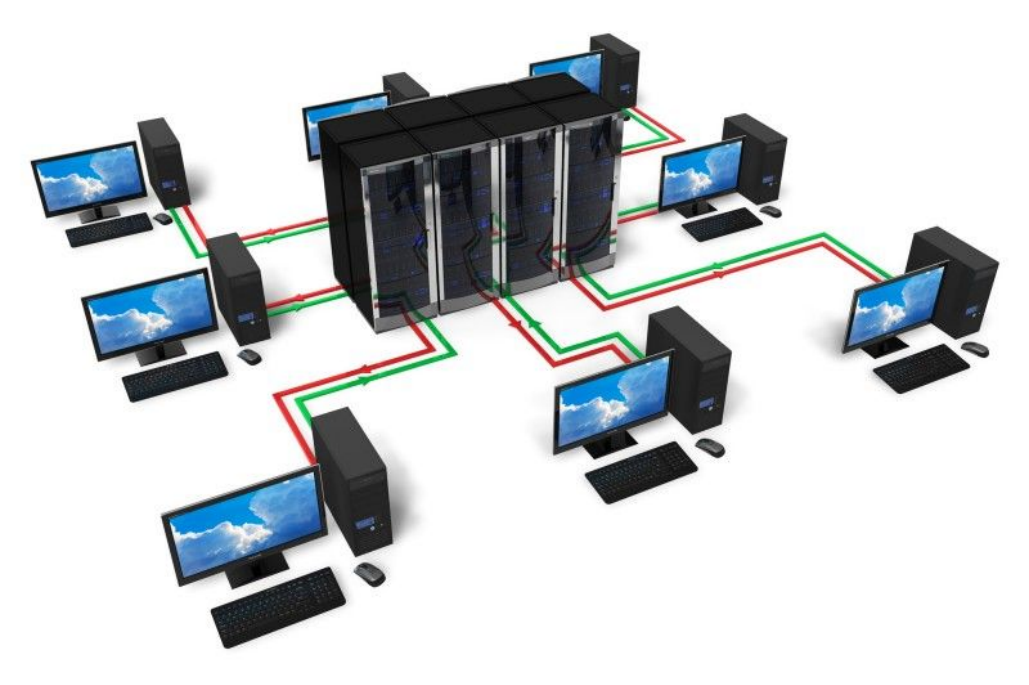
Continuous red-teaming & audits
Cost - $10M–$30M/year

Users

Serving millions of users often costs as much as training.

Cost: $1Billion

-$2 Billion per year



Hidden Multiplier Costs

Launching Space Program:
Upto 2-3 times the initial cost.

For Building an LLM:
$100+ Million

Cost to Build LLM	Launching Space Program
Raw Ingredients Cost (Supercomputer, Data, Talent): $ 25 Billion	Raw Ingredients Cost (Material & Propellants): $100 Million per rocket family
Infrastructure Costs (Data centers, Networking, Storage): $ 1 - 2.5 Billion	Infrastructure Costs (Launch Pads, Test facilities, storage): $2-8 Billion
Training & Techniques (Pre-training, Fine Tuning): $ 40 - 50 Million	Training & techniques (Astronauts, Engineers, Simulations): $1 Billion
Hidden Multipliers (Serving users, Retries, Safety teams): $ 100+ Million	Hidden Multipliers (Failures, Regulations, Redundancy): 2-3 times the total cost. (~ $25 Billion)

“Building a frontier AI lab today is less like a startup and more like building a nuclear reactor—multi-billion scale, complex risks, and global impact.”

Thank You!