LARGE LANGUAGE MODELS

Presented by >

Abdul Rahman
Anirudh

INTRODUCTION

> What is an LLM

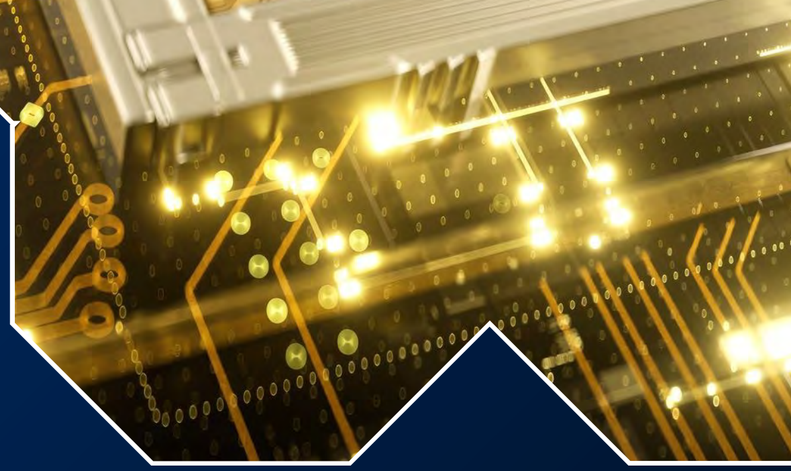
- LLM stands for Large Language Model.
- It’s a type of AI trained on massive text data.
- Designed to understand, generate, and manipulate language.

> Why is is called large?

It’s called "large" because it has:

- Billions of parameters (ChatGPT-3 had 175B, GPT-4 has more).
- Trained on massive datasets from the internet.

HOW IS IT RELATED TO ML?

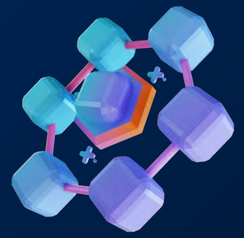


SUPERVISED LEARNING

- 🔍 Trained on labeled data
- ✅ Input: Email, Label: Spam/Not Spam
- 📊 Used in: Classification, Regression

UNSUPERVISED LEARNING

- 🧩 Trained on unlabeled data
- 📌 Model finds patterns on its own
- 📊 Used in: Clustering (e.g., Customer Segmentation)



SEMI-SUPERVISED LEARNING

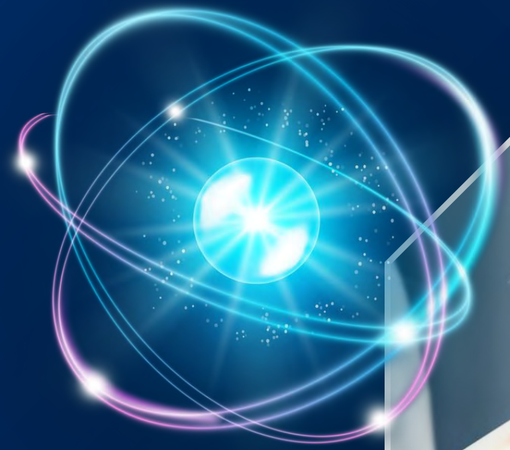
- 🌗 Some data labeled, most not
- 📌 Combines the best of both worlds
- 💡 Used when labeling is expensive

SELF-SUPERVISED LEARNING

- 🤖 No human-labeled data
- 📌 Model creates its own labels from data
- 💬 Used in LLMs like ChatGPT

HOW DOES AN LLM WORK?

> An LLM is a combination of:



1

DATA

"The fuel of the model"

- Text from books, websites, Wikipedia, forums, and code
- Billions of words (tokens)
- Unlabeled and diverse sources
- Helps the model learn language patterns, facts, and context

2

TRAINING

"The learning process"

- Self-supervised learning: predicts next word in a sentence
- Involves fine-tuning and RLHF (human feedback)
- Converts raw data + architecture into a powerful language model

3

ARCHITECHTURE

"The brain of the model"

- Based on Transformer architecture
- Uses self-attention to understand word relationships
- Handles long-range dependencies in text
- Millions or billions of parameters

ARCHITECHTURE OF LLM

TRANSFORMER ARCHITECTURE

- *Foundation of all modern LLMs (like GPT, BERT, LLaMA)*
- *Uses multiple layers of attention and feed-forward neural networks*
- *Processes all words in parallel (faster than old RNNs)*

SELF-ATTENTION MECHANISM

- *Helps the model focus on important words in a sentence*
- *Understands relationships between words — even far apart*
- *Core reason LLMs can write with context and coherence*

POSITIONAL ENCODING

- Tells the model the order of words, since Transformers don't read sequentially
- Adds position-based numbers to each word's embedding
- Keeps sentence structure and grammar intact

APPLICATIONS

> Some real world applications of llms are:

Chatbots & Virtual Assistants

- Examples: ChatGPT, Google Bard, Bing AI
- Use: Answering questions, customer support, personal productivity
- Why it matters: LLMs make conversations feel natural and human-like

Code Generation & Assistance

- Examples: GitHub Copilot, Amazon CodeWhisperer
- Use: Writing, completing, and explaining code for developers
- Why it matters: Saves time and helps beginners learn to code faster

Content Creation & Summarization

- Examples: Jasper, Notion AI, GrammarlyGO
- Use: Writing blogs, summarizing reports, drafting emails
- Why it matters: Automates writing tasks, boosts productivity, and supports creativity

CHALLENGES

Biases	• Trained on human-generated internet data (includes stereotypes) • May reflect or amplify gender, racial, or cultural biases • Can lead to unethical or offensive outputs
Hallucinations	• Generates false or made-up facts with confidence • No built-in fact-checking mechanism • Problematic in domains like medicine, law, education
Computational Cost	• Requires powerful GPUs or TPUs for training • Expensive to run and maintain at scale • High environmental impact due to energy use

FUTURE OF LLMS

> How We Do It

SMALLER & MORE EFFICIENT MODELS

- Companies are developing compact LLMs (like Mistral, LLaMA) that run on personal devices
- Goal: reduce cost, improve speed, and make AI more accessible
- May support offline AI or on-device AI

PERSONALIZED AI ASSISTANTS

- LLMs will become more personalized based on your preferences, goals, and habits
- May assist in everyday tasks, learning, health, and productivity
- Think: A smarter, more intuitive digital co-pilot

BETTER ALIGNMENT & SAFETY

- Focus on making LLMs safer, more factually accurate, and aligned with human values
- Advances in AI alignment, guardrails, and explainability
- Important to prevent misuse, bias, or harm

DEMOCRATIZATION & OPEN SOURCE GROWTH

- Open-source models like LLaMA, Falcon, and Mistral are growing
- More people and organizations can build, customize, and use LLMs
- Encourages innovation, transparency, and fairness

THANK YOU

Questions

?