

[L]LMs and Applications

Where the Magic Meets Reality

Vasudeva Varma

What Are LLMs?

AI systems **trained** on vast datasets to predict, generate, and understand **human-like language**.

Capabilities

Answer questions **Summarize text** Translate languages
Create content Handle complex linguistic tasks

Why LLMs are Important?

Powerful Performance: Handle tasks like reasoning, summarization, and creative generation with **remarkable accuracy**.

Emergent Behaviors: Perform tasks (summarizing novels or generating code) **without explicit training**

Adaptability: Easily **fine-tuned** for specific needs with minimal effort.

Everyday Impact: Enhances tools like chatbots, translators, and content creators for **widespread use**.

Fundamentals:

Pretraining

Fine Tuning

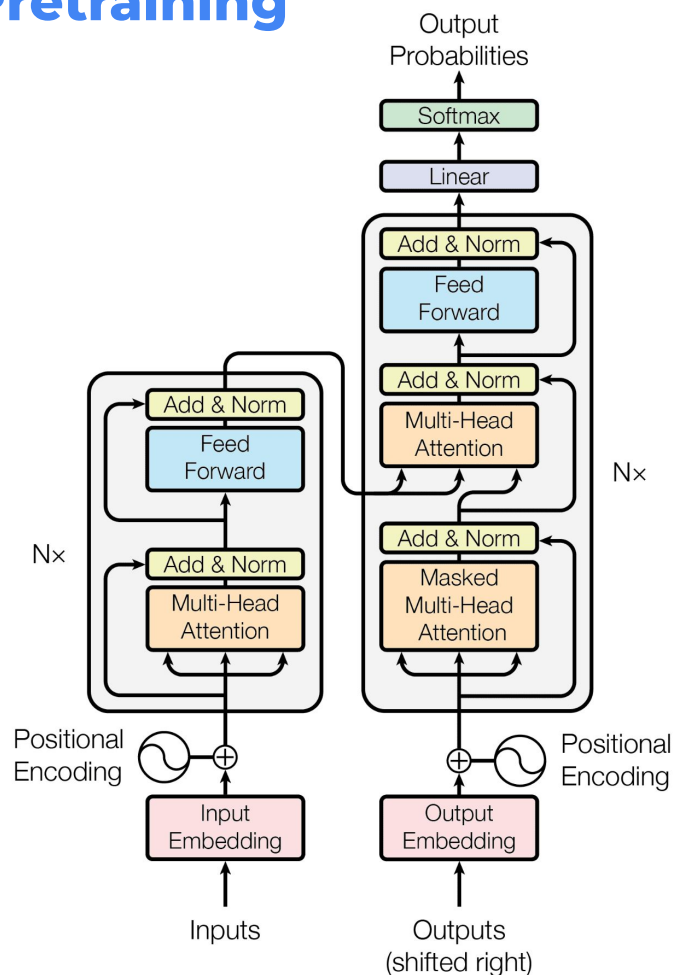
Prompt Engineering

Optimizing

Augmenting

and... agents

Pretraining



Transformer Architecture

Processes **sequences of text** in parallel, understanding **word relationships** effectively.

Key Components:

- **Self (Multi-headed) attention:** Determines the importance of each word in the input (parallelly on different parts)
- **Feedforward Networks:** Adds depth and enhances meaning.
- **Positional Encoding:** Retains the order of words for better comprehension.

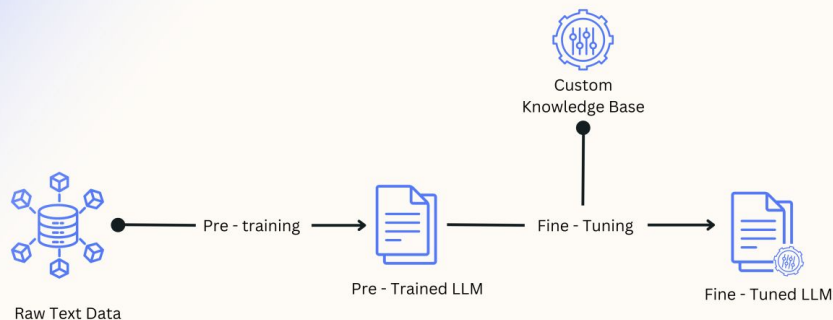
Result: **Pretrained model** enabling fast, scalable processing of large language tasks.

Fine Tuning

Fine-Tuning Techniques

Create **task-optimized**, efficient LLMs to **adapt for specific tasks** like customer service or scientific analysis.

FINETUNING PROCESS



Please see:

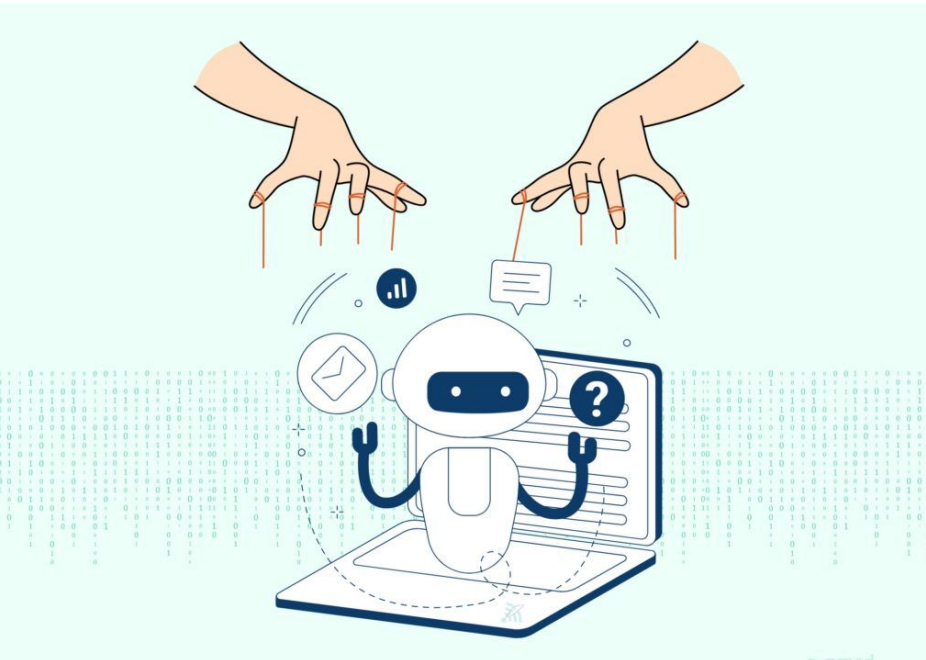
[TechForward Research Seminar Series - JANUARY EDITION](#) (imp: 44:05 to 53:20 of Manohar's talk)

Key Methods:

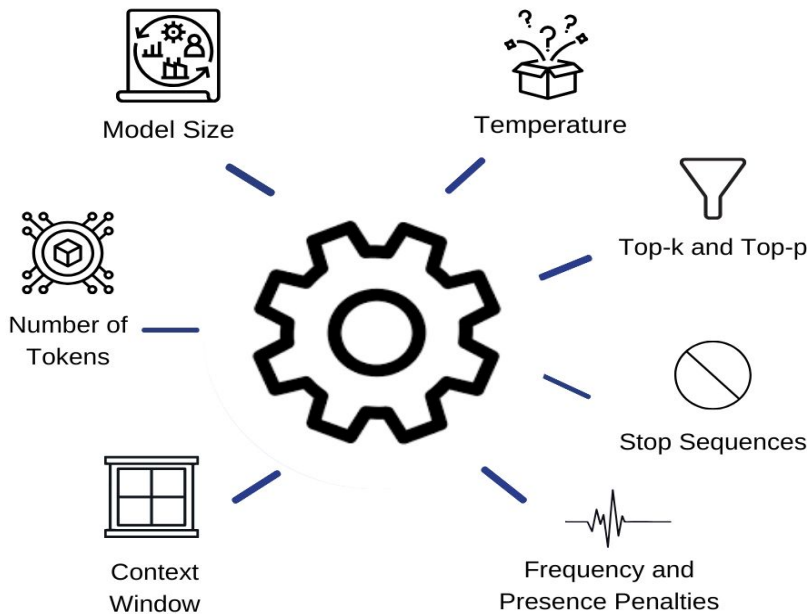
- **Supervised Fine-Tuning (SFT):** Uses labeled data to train for task-specific outputs.
- **Reinforcement Learning from Human Feedback (RLHF):** Refines outputs to align with user preferences.
- **Parameter Efficient Fine-Tuning (PEFT):** Uses techniques like **adapters** for cost-effective customization.

Prompt Engineering

Prompt Engineering:



Sampling Techniques & Parameters:

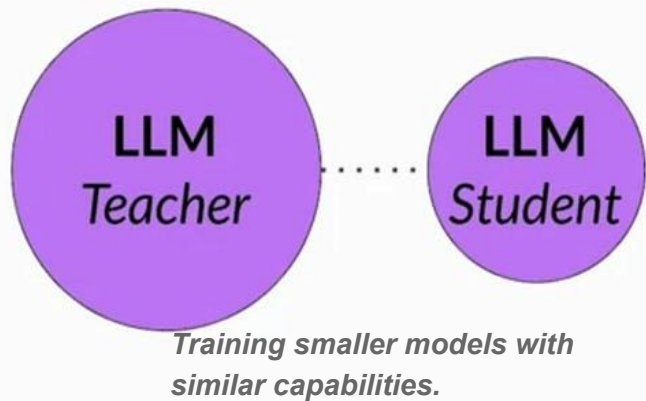


Optimizing

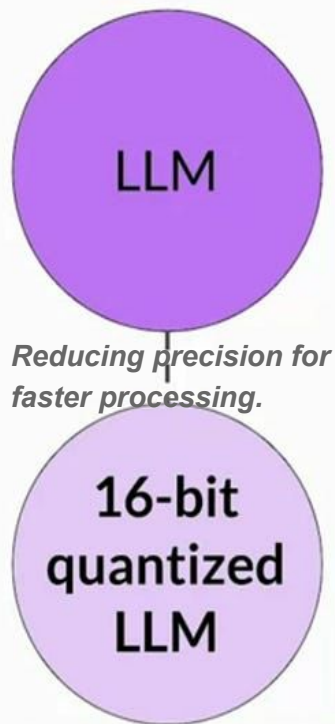
LLMs require significant computation and memory

How to Accelerate Inference to achieve Optimize speed, cost, and efficiency

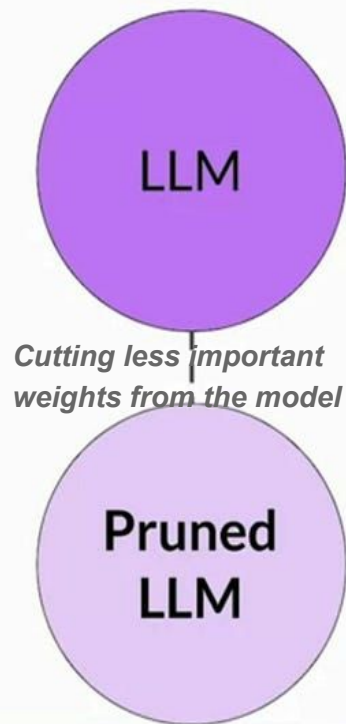
Distillation



Quantization



Pruning



Augmenting

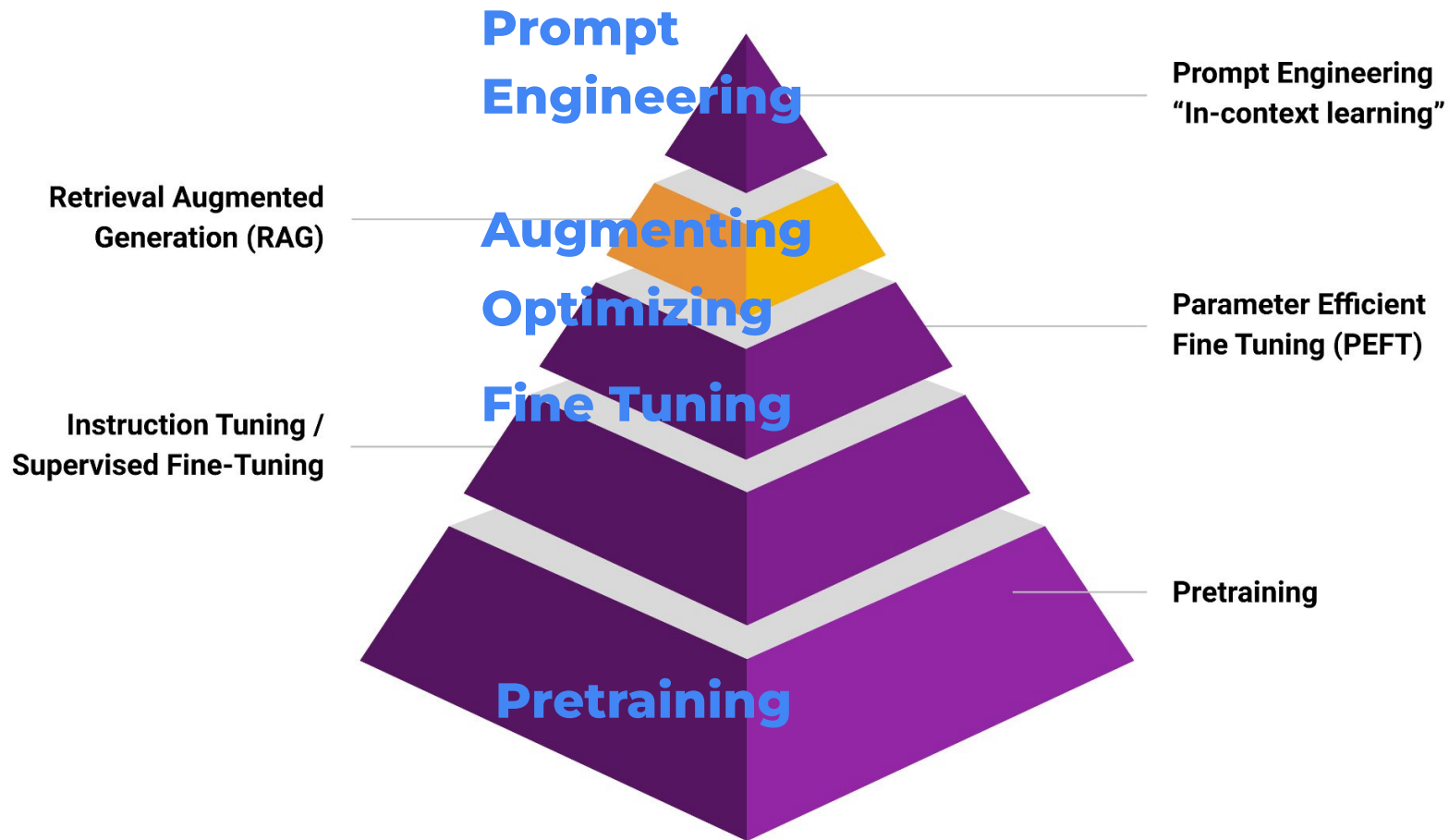
Retrieval-Augmented Generation (RAG) *combines retrieval systems with LLMs*

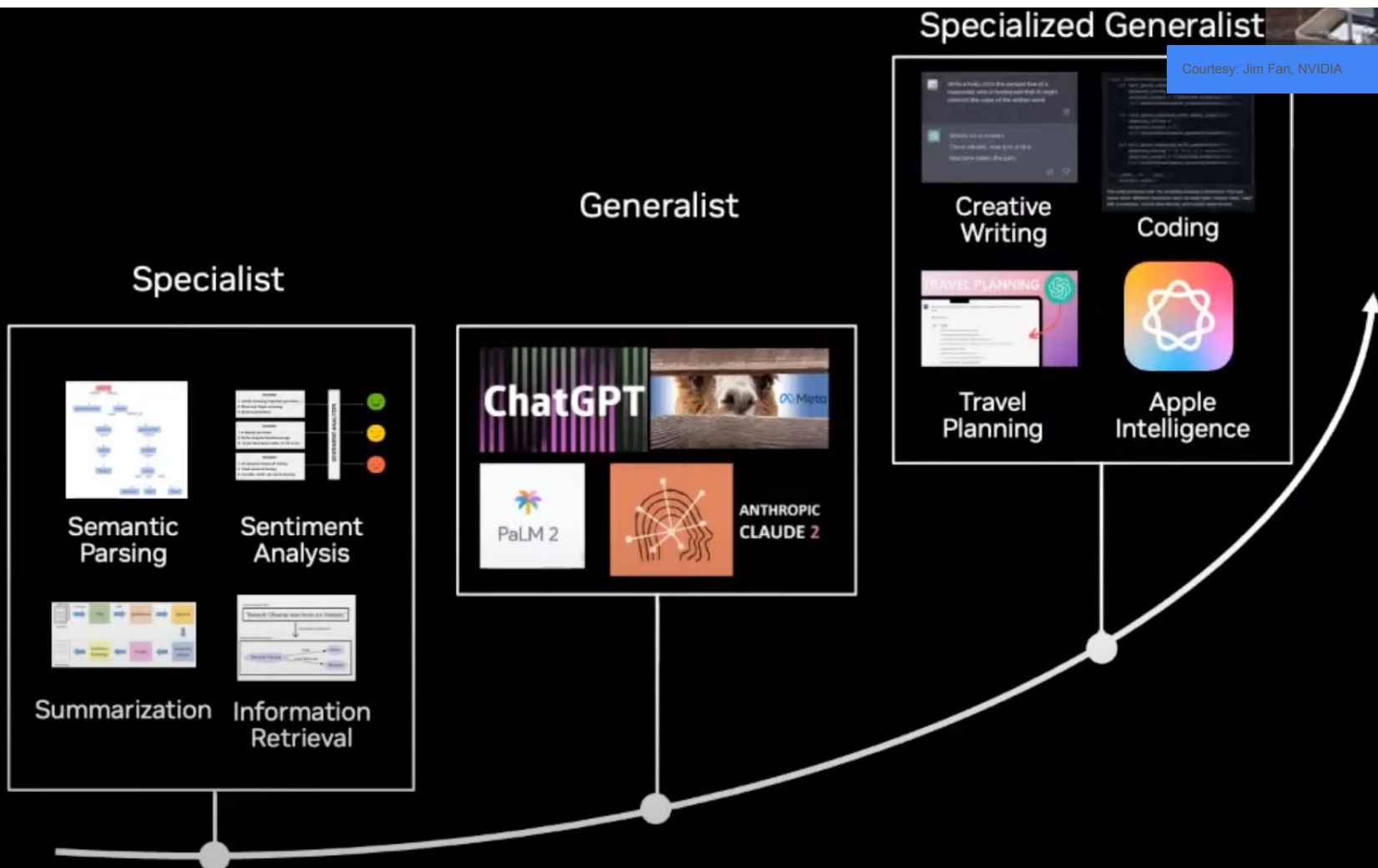
How It Works

1. **Query Understanding:** The LLM *interprets* the user's question.
2. **Information Retrieval:** The query is matched against a *Document Repository*
3. **Contextual Generation:** The retrieved data is fed into the LLM, which *generates a grounded, factual response*.

Key Advantages

- **Enhanced Accuracy:** Incorporates real-world, dynamic data, generating **up-to-date, accurate information**
- **Domain-Specific Adaptability:** Tailored to specialized datasets (e.g., healthcare, legal).
- **Reduced Hallucinations:** Limits reliance on outdated or inferred knowledge.



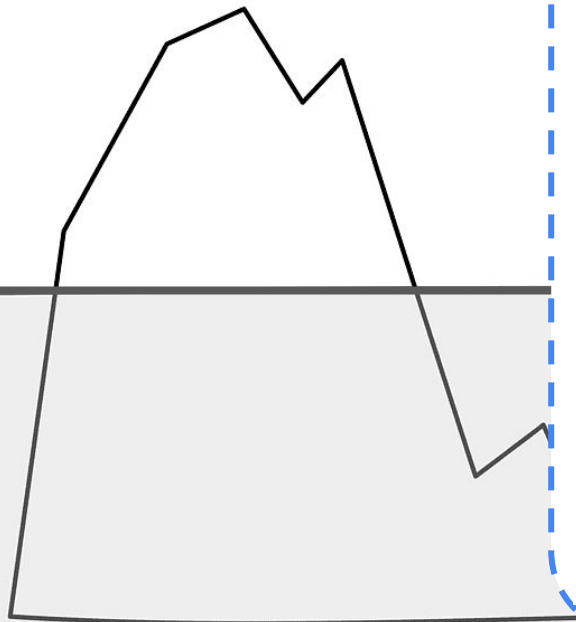


I Don't Work on LLMs...

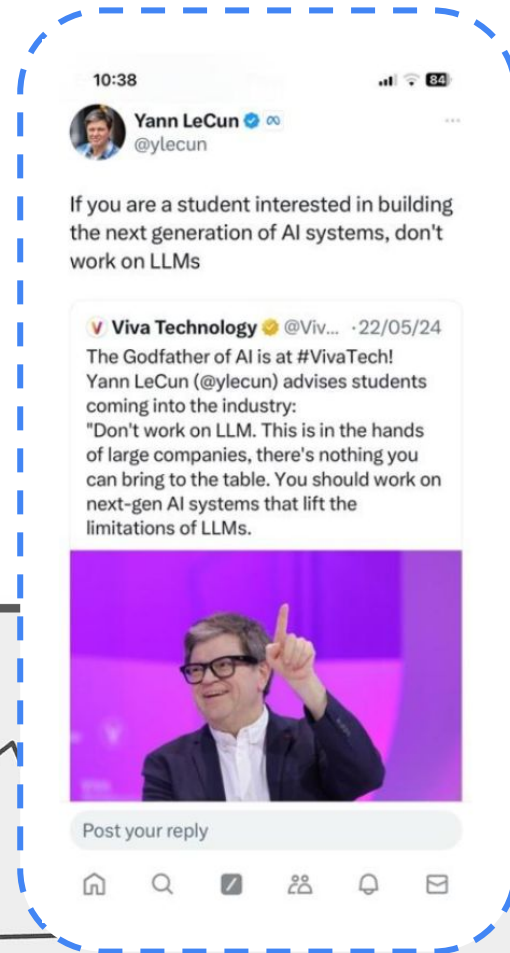
My focus is on addressing the limitations of LLMs

"Most of the progress in AI has been on the part of the iceberg that is above the water."

There's a lot of progress yet to be made on the part of the iceberg that is below the water."



-Yann LeCun



Who will work on local issues?

THE ECONOMIC TIMES | tech

English Edition | Today's ePaper

Subscribe

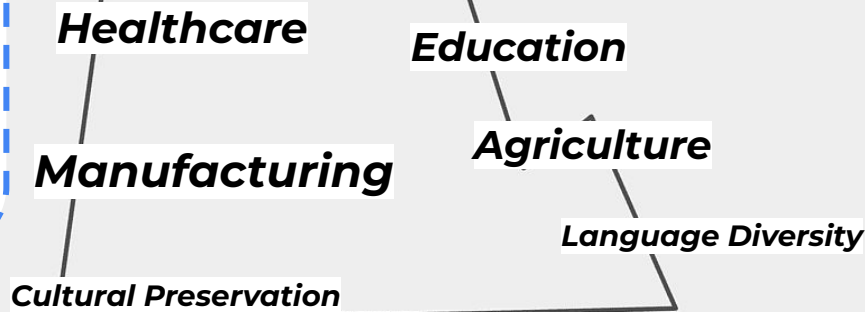
Freedom

ETPrime Markets Market Data News Industry Rise Politics Wealth MF Tech Careers Opinion NRI Panache ETV Spotlight

telligence Web Stories Information Tech (IT) Tech & Internet Funding Startups Tech Bytes Newsletters Blogs & Opinion

ws Tech Artificial Intelligence India's big AI opening is at application level on top of LLMs: Andrew Ng

India's big AI opening is at application level on top of LLMs: Andrew Ng



... and many unaddressed local issues

The biggest opportunity for India is to look at the sectors where the Indian economy is already strong and figure out the applications of AI to those sectors that enable it to maintain its advantages

Andrew Ng (Co-founder of Coursera and founder of DeepLearning.AI)

3 Use cases/Applications

Healthcare

Science Research

Education/Learning

Role of Gen AI and LLM in **HealthCare**

Experiences with healthcare Apps

- Personalised Diagnosis
- Doctor Patient Communication
- Drug Discovery
- Insurance

The other side of the coin

- Risk / accuracy
- Fairness/ Bias
- Accountability
- Transparency
- Ethics
- Safety



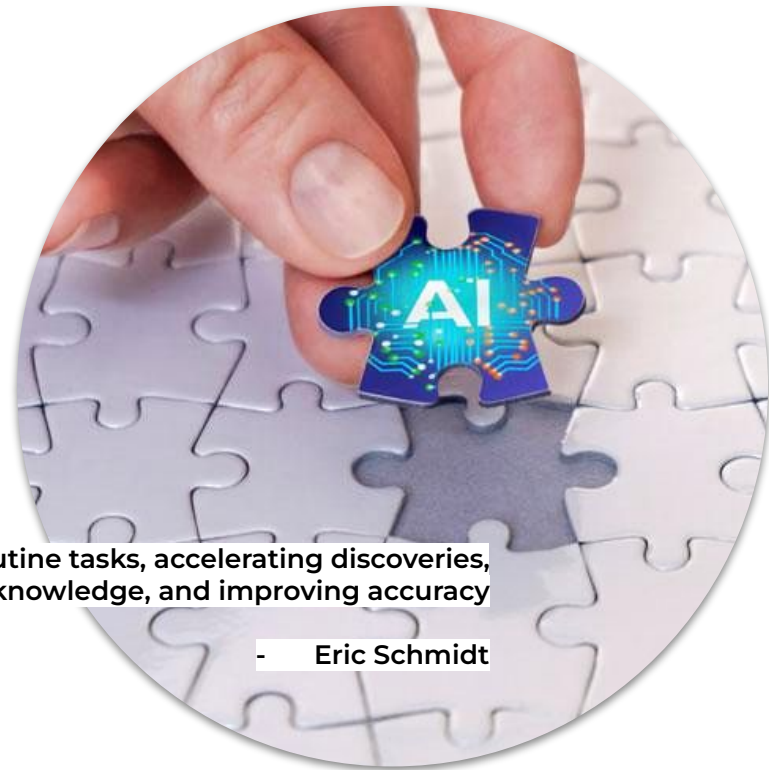
Role of Gen AI and LLM in Scientific Research



How will AI revolutionize science and science research?

AI will transform the scientific process by automating routine tasks, accelerating discoveries, enhancing experimentation, integrating cross-domain knowledge, and improving accuracy

- Eric Schmidt



Use case

Education/Learning

Experiments in using GenAI in Learning



**Behavioural and Social
Learning Course
RGUKT & ISB**



**Building a Negotiation
Course with GenAI
ISB**

00:00 00:38

using plastic bags because it pollutes **it pollutes** the environment as it cannot be recycled easily it is also important to maintain personal hygiene by taking a shower every day and washing our hands [before] [every] [meal] [it] is our responsibility to ensure cleanliness because an unclean environment becomes a prominent reason for various diseases like dengue malaria etc

Errors

- 0 Mispronunciations ① On
- 4 Omissions ① On
- 2 Insertions ① On

Pronunciation score

Score breakdown

Accuracy score ① 94 / 100

Completeness score ① 96 / 100

Fluency score ① 84 / 100

88

0 ~ 59 60 ~ 79 80 ~ 100

Developer resources

Issues with scalability and adaptability

Some are natural fits and some tough ones.

- STEM subjects?

Overall Reading Comprehension

Reading Report

73

Current Score

Reporting Indicators

Overall Proficient ★★★★★

Reading Transitional ★★

Comprehension Proficient ★★★★★

Parameters Influencing Current RCA

Pronunciation Correct pronounced Words - 107 Mispronounced words - 7 Missed Words - 4

Rhythm & Intonation ★

Fluency 120 words/min Range - 180 - 190 words/min

helping others in need is very **those** **who** **need** **my** **help** **helping** **others** **in** **need** **to** **very** important i often help those who need my **mother** **my** help recently i helped my mother clean the **house** **it** **house** **it** was a few days ago that my mother was busy cleaning my mother seemed to be **tired** since it was my holiday that day i decided to help my mother out i helped her clean all the rooms one by one i **swept** **swept** the floor **in** **and** every **room** mopped the **floor** **windows** and wiped the windows i also dusted off the dust from photos that **photos** **the** hung **hung** on the **walls** **walls** i also gave my **mother** **some** paracetamol **paracetamol** and cut fruits for her to eat i was exhausted after **she** the day's work but i really enjoyed helping my **mother**

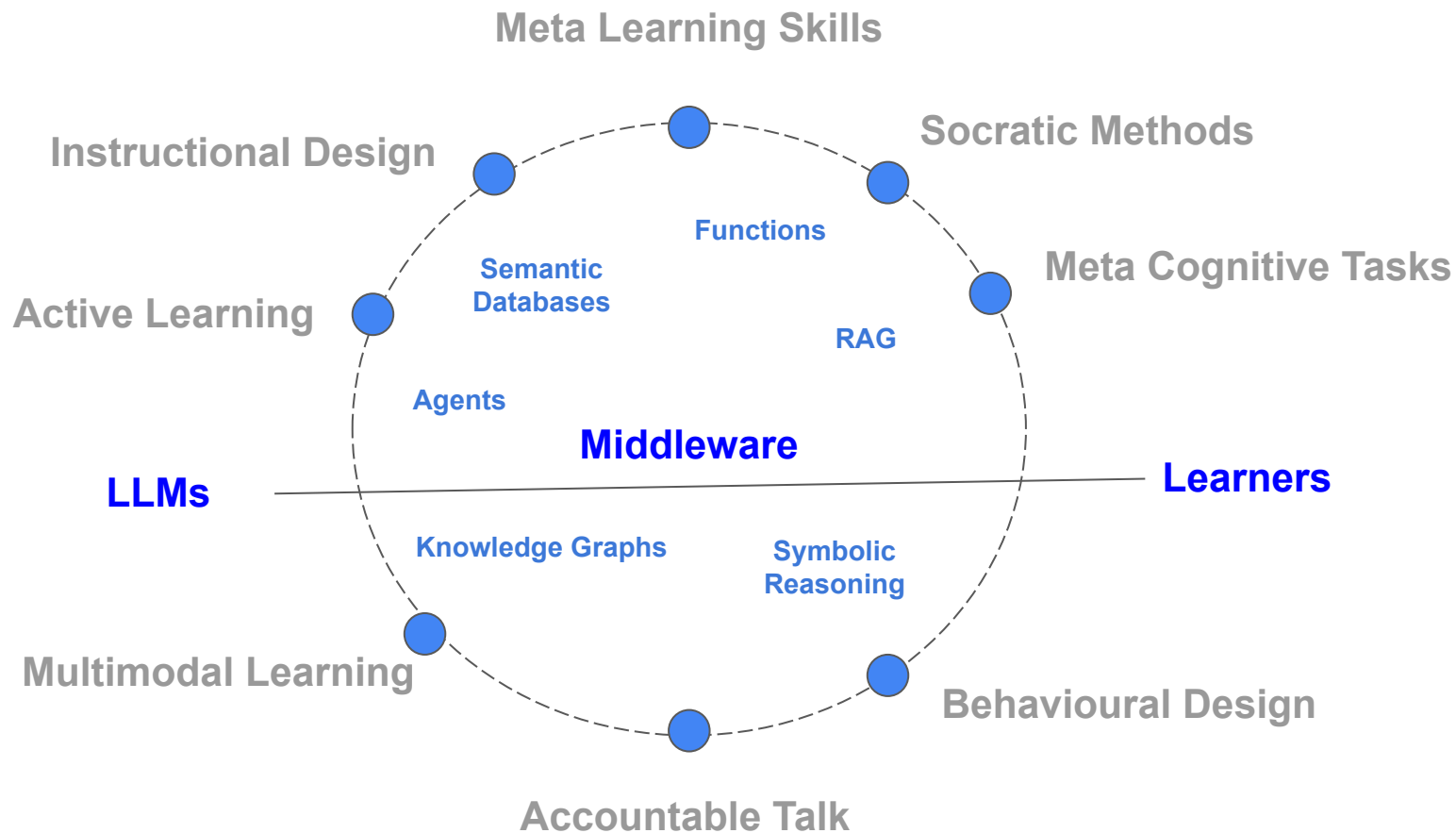
Feedback

Excellent pronunciation standards.Very close to the benchmark.

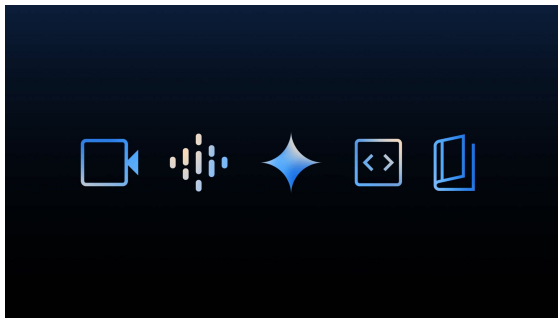
Work harder to improve your stress and intonation patterns

You are quite fluent.Some more practice will take you closer to the benchmark.

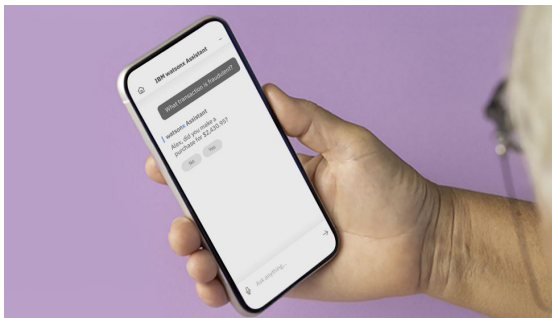
7 Mispronunciation
4 Omission
26 Insertion



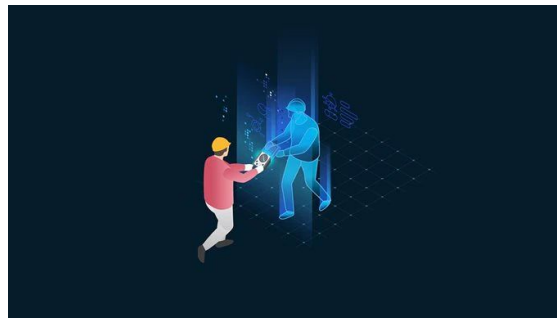
3 key advancements that is rapidly evolving Gen AI's capabilities (courtesy: Eric Schmidt)



Really Large Context Windows



Agents



Text-to-Action Capabilities

Key messages:

The last mile is the hardest; if one doesn't do well, magic may fall flat on its face.

While technology tremendously enhances productivity, *domain knowledge is the cornerstone to success.*

Need to build **middleware**, **responsibility layer**, and **domain specific foundation models**

Thank you
