



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

COMP 4901B

Large Language Models

Instructor: Junxian He

<https://jxhe.github.io/teaching/comp4901bf25>

Teaching Team & Office Hours

Instructor: Junxian He

Office Hour: Wed 11am-12am, CYT3004



TA1: Yuzhen Huang

Office Hour: Wed 10am-11am



TA2: Wei Liu

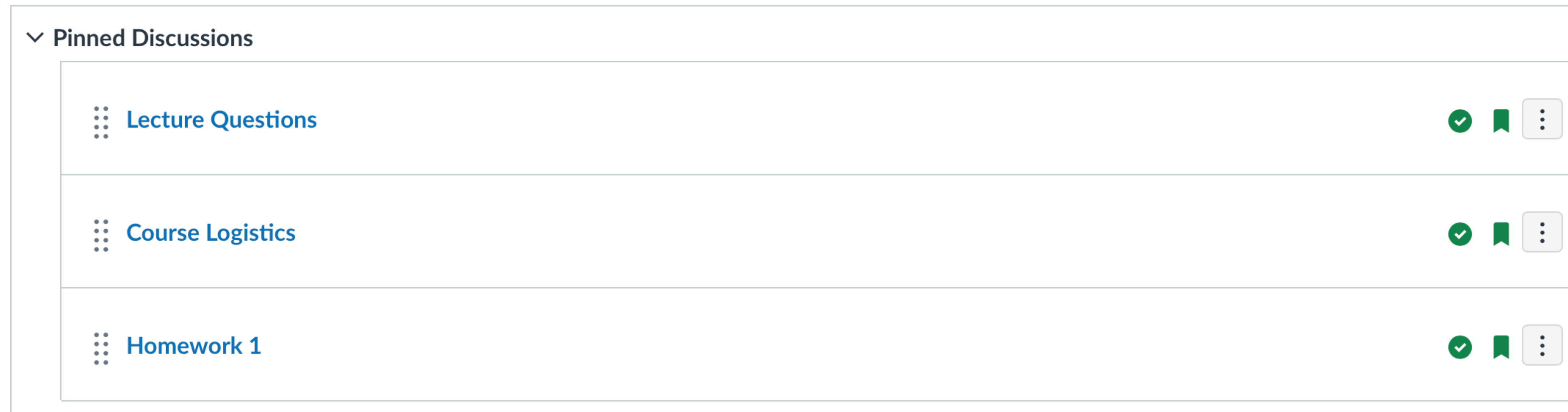
Office Hour: Tue 2pm-3pm

TA Office Hours are through Zoom

This is an in-person Class

- Lectures are recorded and will be released within the week of the lecture
- Lecture slides will be available before each lecture

Communication and Discussion



Please direct all technical questions on Canvas, and do not directly contact the instructor or TAs for technical questions. We will answer your questions on Canvas promptly

Canvas allows anonymous communication

Pre-requisite

- Probability
 - Distribution, random variable, expectation, conditional distribution, variance
- Machine Learning
 - Training/validation/test, Objective Optimization, Gradient Descent
 - Neural Networks, Deep Learning
- Python programming

Grading

- Attendance (10%)
- 4 assignments (40%)
 - All are programming-oriented, may need a simple writing report
 - 3 free late days in total, for additional late days, 20% penalization applied for each day late
 - No assignment will be accepted more than 3 days late
- Mini Group Project (2-3 people) (25%)
- Final exam, close-book (25%)

Attendance

- Occasional quiz questions
- 80% of attendance will give you full grade
- Correctness of quiz answers does not influence attendance grading

Honor Code

Do's

- Work on the homework independently unless otherwise specified

Don'ts

- Copy, refer to, or look at solutions from previous years, online, or others
- Longer versions on the course website

We have zero tolerance — in the case of honor code violation for a single time, you will fail this course directly

Use of Generative AI Tools

You Can:

- Use AI tools to help with your code
- Use AI tools to help polish your writing for report

You Can't:

- Directly ask the AI tools to write your report and copy

More Info on Course Website

- Canvas is the main platform for announcement, discussion, homework submission
- Recorded videos on canvas
- Syllabus, slides and relevant reading materials
- Detailed course logistics

Topics to be Covered

	Chain-of-Thought Reasoning
Introduction to Large Language Models	Deep Reasoning Models
Machine Learning Basics	Chung Yeung Festival
Language Model	Synthetic Data Training: Data Synthesis and Distillation
Neural Networks and Transformers	Synthetic Data Training: Self-Improving through RLVR
Transformers	LLM Agents
Pretraining: Objectives and Data Curation	LLM Agents
In-Context Learning, Prompt Engineering	LLM Safety: Attacks and Defense
Evaluation of Large Language Models	Hallucination and RAG
The National Day Holiday	Bias and Ethics
Instruction Tuning and Alignment	Sparse Models
Instruction Tuning and Alignment	Review Representative LLMs
Reinforcement Learning Basics	
Reinforcement Learning from Human Feedback (RLHF)	
Scaling Up LLMs: Scaling Laws	

Introduction to Large Language Models

ChatGPT

Can you introduce large language models for me

3:00 PM ✓

 GPT-5 ...

Sure! Here's a concise introduction to large language models (LLMs):

What they are

- Large language models are AI systems trained on vast text datasets to predict the next token (word/character). Through this, they learn grammar, facts, reasoning patterns, and styles.
- Examples include GPT-4/5, Claude, Llama, and PaLM.

What they can do

- Natural language: drafting, editing, summarizing, translation, question answering.
- Reasoning and problem solving: planning, explaining code/math, multi-step instructions.
- Structured tasks: extracting entities, classification, generating JSON, SQL.
- Multimodal variants: can analyze images, audio, or produce code that generates visuals.
- Tool use: when connected to external tools, can browse, call APIs, run code, or control apps.

How they work (high level)

- Architecture: typically transformer neural networks with self-attention layers.
- Training:
 1. Pretraining on internet-scale corpora via next-token prediction.

Coding Tools



Cursor

Auto-complete



GitHub
Copilot

```
class transformer_encoder_block(nn.Module):
    def __init__(self, d_model, n_heads, d_ff, dropout):
        super().__init__()
        self.self_attention = MultiHeadAttention(d_model, n_heads, dropout)
        self.feed_forward = FeedForward(d_model, d_ff, dropout)
        self.layer_norm = LayerNorm(d_model)
        self.dropout = nn.Dropout(dropout)

    def forward(self, x):
        x = x + self.dropout(self.self_attention(x))
        x = x + self.dropout(self.feed_forward(x))
        return self.layer_norm(x)
```

Agentic Coding



With Search Abilities

Can you list holidays in the 2025 fall semester for HKUST
--web_search true

3:37 PM ✓

 GPT-5

Here are the public/university holidays that fall within HKUST's Fall Term in 2025 (Fall Term runs Sep 1–Dec 19, 2025):

- Wed, Oct 1 — National Day. (registry.hkust.edu.hk, gov.hk)
- Tue, Oct 7 — The day following the Chinese Mid-Autumn Festival. (scribd.com, gov.hk)
- Wed, Oct 29 — Chung Yeung Festival. (scribd.com, gov.hk)
- Thu, Dec 25 — Christmas Day. (scribd.com, gov.hk)
- Fri, Dec 26 — The first weekday after Christmas Day (Boxing Day). (scribd.com, gov.hk)

Notes:

- HKUST's Fall Term dates are Sep 1–Dec 19, 2025. (registry.hkust.edu.hk)
- HKUST follows Hong Kong's gazetted public holidays; the above dates are those that occur during the Fall Term window. (gov.hk, info.gov.hk)

If you need these in an .ics calendar file, I can generate one for import to your calendar app.

Learn more:

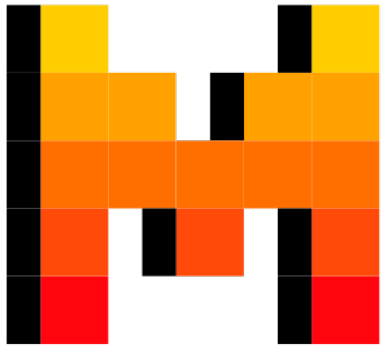
1. [Upcoming Key Dates | HKUST - Academic Registry](#)
2. [GovHK: General holidays for 2025](#)
3. [dates25-26confirmed | PDF | Thesis | Observances](#)
4. [General holidays for 2025 published](#)

General Tool Use in the Wild

Examples: <https://agent.minimaxi.com/>

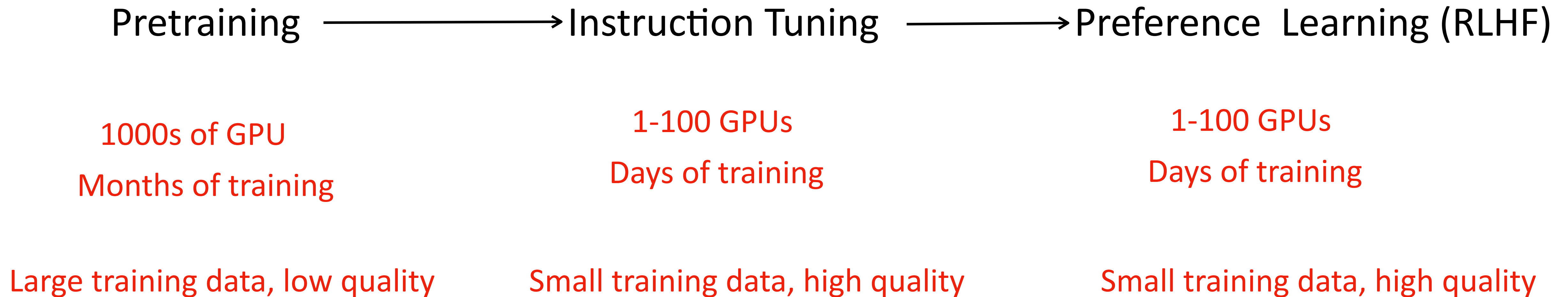
The Major LLM Industry

US:  OpenAI  Gemini  Grok

Europe:  MISTRAL
AI_

China:  deepseek  Qwen  KIMI  ZHIPU · AI
 MINIMAX  StepFun

The LLM Development Stages (ChatBot)



Language Model Reasoning

Problem

Every morning Aya goes for a 9-kilometer-long walk and stops at a coffee shop afterwards. When she walks at a constant speed of s kilometers per hour, the walk takes her 4 hours, including t minutes spent in the coffee shop. When she walks $s + 2$ kilometers per hour, the walk takes her 2 hours and 24 minutes, including t minutes spent in the coffee shop. Suppose Aya walks at $s + \frac{1}{2}$ kilometers per hour. Find the number of minutes the walk takes her, including the t minutes spent in the coffee shop.

Solution 1

$\frac{9}{s} + t = 4$ in hours and $\frac{9}{s+2} + t = 2.4$ in hours.

Subtracting the second equation from the first, we get,

$$\frac{9}{s} - \frac{9}{s+2} = 1.6$$

Multiplying by $(s)(s+2)$, we get

$$9s + 18 - 9s = 18 = 1.6s^2 + 3.2s$$

Multiplying by 5/2 on both sides, we get

$$0 = 4s^2 + 8s - 45$$

Factoring gives us

$$(2s - 5)(2s + 9) = 0, \text{ of which the solution we want is } s = 2.5.$$

Substituting this back to the first equation, we can find that $t = 0.4$ hours.

Lastly, $s + \frac{1}{2} = 3$ kilometers per hour, so

$$\frac{9}{3} + 0.4 = 3.4 \text{ hours, or } \boxed{204} \text{ minutes}$$

Self-Reflection

Question: If $a > 1$, then the sum of the real solutions of $\sqrt{a - \sqrt{a + x}} = x$ is equal to

Response: <think>

To solve the equation $\sqrt{a - \sqrt{a + x}} = x$, let's start by squaring both \dots

$$\left(\sqrt{a - \sqrt{a + x}}\right)^2 = x^2 \implies a - \sqrt{a + x} = x^2.$$

Rearrange to isolate the inner square root term:

$$(a - x^2)^2 = a + x \implies a^2 - 2ax^2 + (x^2)^2 = a + x \implies x^4 - 2ax^2 - x + (a^2 - a) = 0$$

\dots

Wait, wait. Wait. That's an aha moment I can flag here.

Let's reevaluate this step-by-step to identify if the correct sum can be \dots

We started with the equation:

$$\sqrt{a - \sqrt{a + x}} = x$$

First, let's square both sides:

$$a - \sqrt{a + x} = x^2 \implies \sqrt{a + x} = a - x^2$$

Next, I could square both sides again, treating the equation: \dots

\dots

From DeepSeek R1

Objectives of This Course

1. Understand the principles and large language models
2. Roughly understand how models like DeepSeek-R1 and Qwen are developed
3. Know how to better tame LLMs through prompting engineering
4. Learn to train language models (pretraining and instruction tuning) in minimal settings
5. Understand the harms, risks, and safety concern of LLMs

Thank You!