



引言

CS2916 大语言模型

飲水思源 愛國榮校

<https://plms.ai/teaching/index.html>



内容安排

- 生成式人工智能 & 课程介绍
- 自然语言处理基础
- 机器学习基础 & 相关工具简介



Generative AI is Eating the World



杰夫·贝索斯

目前形式的大语言模型并不是**发明**，而是**发现**。望远镜是一项发明，但通过它观察木星，知道它有卫星，是一项发现。而大语言模型更像是发现，它们的能力不断让我们感到惊讶

人生中让我印象深刻的两次**技术革命**演示，一次是现在操作系统的先驱“图形用户界面”，另一个就是以ChatGPT为代表的**生成式人工智能**技术



比尔盖茨



黄仁勋

ChatGPT相当于**AI界的iPhone**问世，它使**每一个人**都可以成为程序员

马斯克悄悄成立大模型公司xAI





世界在发生着快速的变化

从文本生成文本



You

请帮我写一个面向本科生的《大语言模型》的课程大纲



ChatGPT

《大语言模型》课程旨在向本科生介绍人工智能和自然语言处理领域的一项重要技术——大型语言模型（如GPT系列）。本课程将通过理论学习与实践操作相结合的方式，帮助学生深入理解大语言模型的工作原理、应用场景、优点与局限，以及模型训练和部署的基本方法。以下是课程的大纲：

课程大纲

第1周：引言

- 课程介绍与目标
- 人工智能与自然语言处理概述
- 大语言模型的历史与发展

第2周：基础知识回顾

- 机器学习基础
- 深度学习入门
- 文本处理基本技术

第3-4周：大语言模型原理

- 什么是大语言模型
- 模型架构（如Transformer）
- 模型训练方法（包括预训练和微调）

第5-6周：GPT系列模型深度解析

- GPT模型结构详解
- 从GPT-1到GPT-4：演进与创新
- 案例研究：特定应用中的GPT模型



世界在发生着快速的变化

从文本生成图片



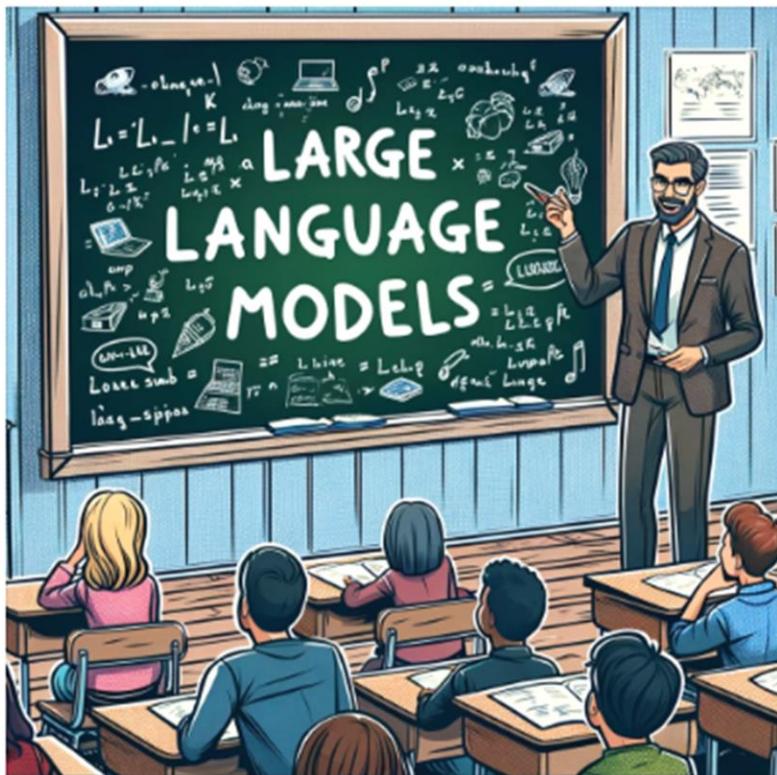
TF

You

帮我画一张卡通风格的图片，里面一位老师正在教室里授课，黑板上写着课程名《Large Language Models》



ChatGPT





世界在发生着快速的变化

从文本生成代码



You
生成一段python代码，运行后可以画出一个心型图

ChatGPT

Heart Shape



这是使用Python绘制的心形图。通过运行上述代码，我们得到了一个美丽的心形图像，没有坐标轴，突出了心形的纯粹和简洁。 [↩]



世界在发生着快速的变化

查看详细代码



Analysis

```
python Copy code  
  
import matplotlib.pyplot as plt  
import numpy as np  
  
# Define the parameters for the heart shape  
t = np.linspace(0, 2 * np.pi, 1000)  
  
# Define the heart shape in polar coordinates  
x = 16 * np.sin(t) ** 3  
y = 13 * np.cos(t) - 5 * np.cos(2*t) - 2 * np.cos(3*t) - np  
  
# Plot the heart shape  
plt.figure(figsize=(8, 7))  
plt.plot(x, y, color='red')  
plt.title('Heart Shape')  
plt.axis('equal')  
plt.axis('off')  
plt.show()
```

这是使用Python绘制的心形图。通过运行上述代码，我们得到了一个美丽的心形图像，没有坐标轴，突出了心形的纯粹和简洁。



世界在发生着快速的变化

两艘海盗船在一杯咖啡中航行时相互争斗的逼真特写视频

从文本生成视频





大语言模型驱动的生成式人工智能(GAI)

□ GAI是一种特定类型的AI

- 直接应用：可以生成不同模态的内容
- 间接应用：利用**生成力**解决复杂任务

基于大语言模型的“generative power”被看成是一种新的算力，可以用来解决许多复杂的任务，比如数学、化学等。

□ 核心技术基础

- **生成式预训** (Generative Pretraining)
 - 数据存储 (数值向量)
- **提示工程** (Prompting Engineering)
 - 数据读取 (自然语言描述)

是一种基于Decoder-only 的 Transformer 架构，我们将会在第4周课程介绍。

提示学习是一个伟大的技术，它使得人类第一次可以和AI模型直接对话，且用的是人类的语言，第五周课程介绍。



生成式人工智能技术的独特优势

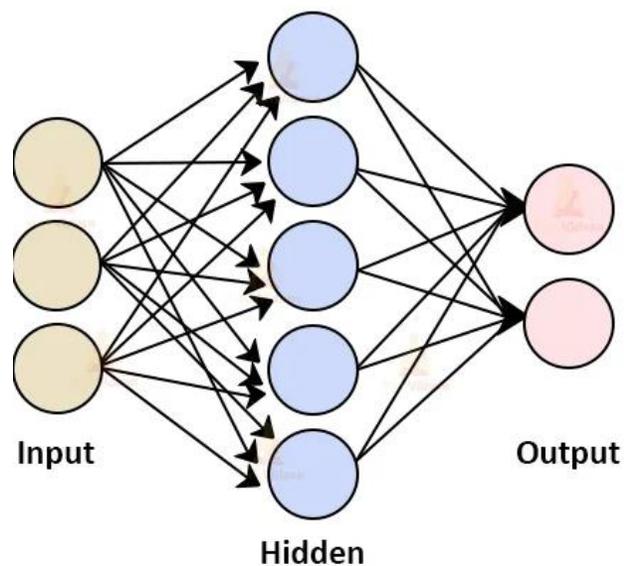
- 数据存储
- 数据检索
- 数据标注
- 建模范式统一

数据存储

- 过去：存储在磁盘，数据库里
 - 无需学习过程
 - 可理解
 - 需要特殊的语言进行交互



- 现在：存储在参数里
 - 泛化性能好
 - 需要数学运算, GPU
 - 可以用自然语言来进行交互





数据检索

□ 过去：搜索引擎

- 无需学习过程
- 信息检索技术支持
- 字符的匹配

□ 现在：提示词

- 自然语言描述
- 可引导、可学习

中国的首都是

Q 网页 视频 知道 贴吧 图片 资讯 文库

百度为您找到相关结果约79,400,000个

 中国 北京
BEIJING CHINA

中华人民共和国首都：
北京市

北京(Beijing), 简称“京”, 古称燕京、北平, 是中华人民共和国直辖市、国家中心城市、超大城市, 国务院批复确定的中国化中心、国际交往中心、科技创新中心, 截至202... [详情](#)



You

中国的首都是



ChatGPT

中国的首都是北京。



数据标注

□ 过去：雇佣标注工人进行标注

- 纯人工
- 标注的质量难以控制
- 标注方式与具体任务高度相关

□ 现在：大模型辅助标注&直接合成

- 标注变成QA问题
- 不需要人工标注太多数据

人物 1 组织 2 产品 3 位置 4

经过多学科专家集中会诊后，在 **广州医科大学附属第一医院** 组织集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果。

Playground

Load a preset...

Save

经过多学科专家集中会诊后，在广州医科大学附属第一医院组织集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果
地点：广州市
经过多学科专家集中会诊后，在广州医科大学附属第一医院组织集中收治的广州最后一位新冠肺炎危重症患者核酸检测和影像学均达到出院标准，意味着广州市在救治新冠肺炎危重症患者工作中取得了阶段性成果
机构：广州医科大学附属第一医院

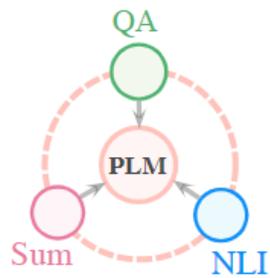


建模范式统一

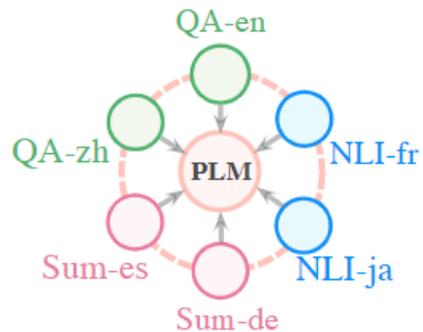
□ 不同任务, 语言, 模态, 一个框架



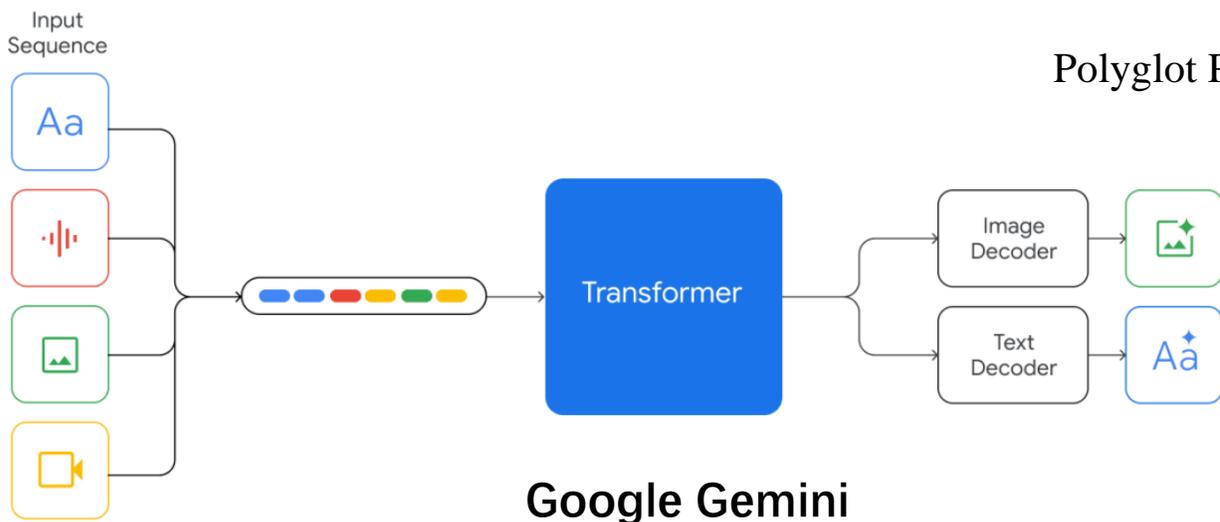
(1)



(2)



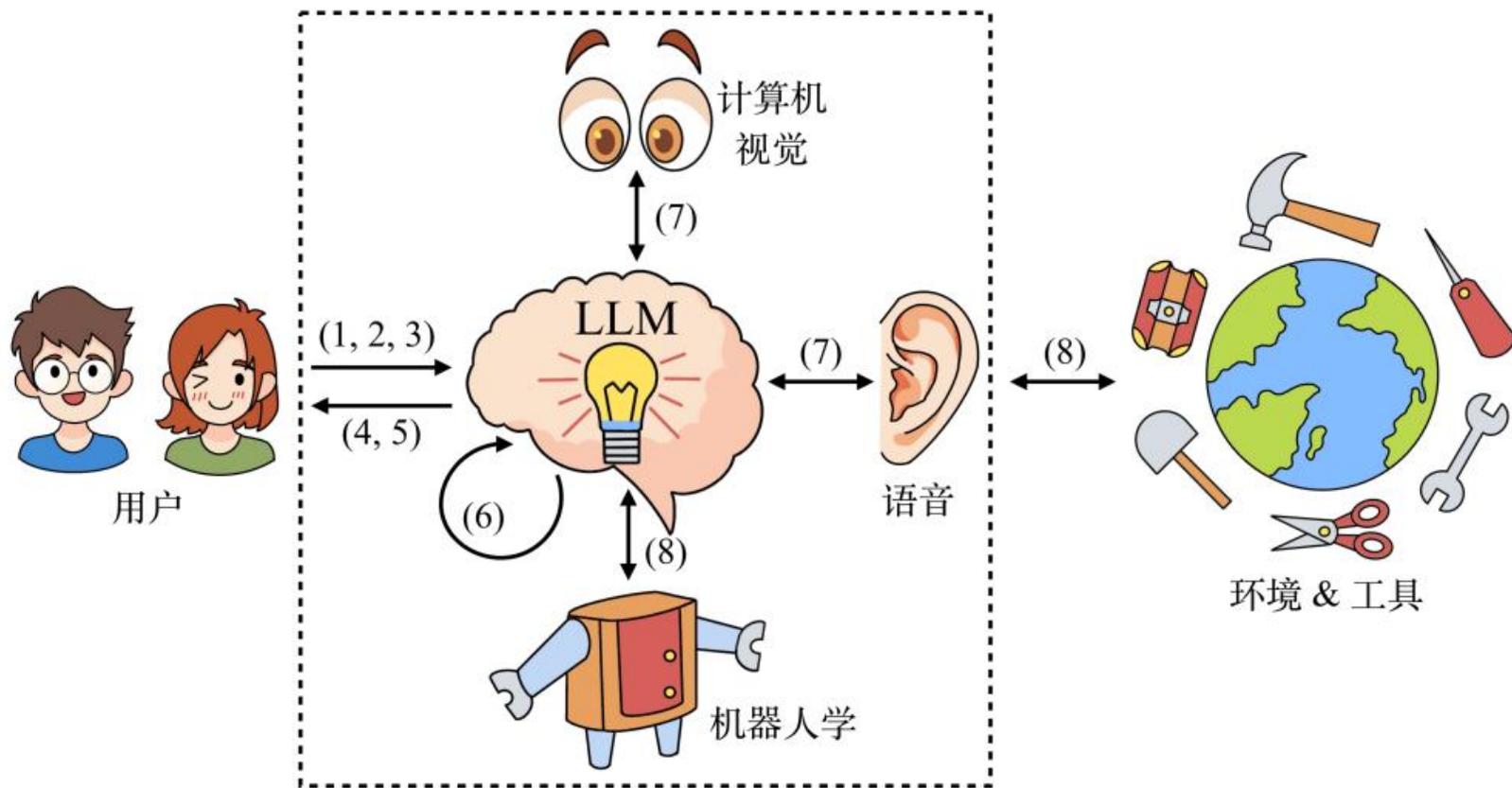
(3)



Polyglot Prompt: Multilingual Multitask Prompt Training Fu et al.2022



大语言模型是人类延伸的大脑



生成式AI

1. 需求挖掘
2. 提示工程
3. 系统安全
4. 用户接口
5. AI安全 & 对齐
6. LLMops
7. 多模态功能
8. 环境交互性

传统的AI是“降劳”（取代重复的劳动力），生成式AI是“增智”



生成式AI (GPT) 将如何影响我们的工作?

- GPT 的影响横跨各类薪资层级：尽管存在部分特殊情况，但整体来看，**工资越高，受GPT 冲击的程度越大**；
- 职业技能与 GPT 的冲击程度有关：**科学和批判性思维技能最不容易受GPT冲击**，而编程和写作技能受影响的程度最高；
- **高学历更容易受到 GPT 的冲击**：持有学士、硕士和更高学位的人比没有正规教育学历的人更容易受到 GPT 的冲击；

GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models

Tyna Eloundou¹, Sam Manning^{1,2}, Pamela Mishkin*¹, and Daniel Rock³

¹OpenAI

²OpenResearch

³University of Pennsylvania



技术发展过快带来的问题

- 安全防控与评估技术**滞后于**大模型的发展
 - 如何评估大模型生成结果的可靠性
- AI相关政策制度**滞后于**技术发展
 - 辅助医疗诊断
- 缺乏科学原理解释
 - 例如：多模态大模型幻觉产生的原因？
- 人才缺口大
 - 新的知识迭代太快，待解决问题太多，

Gemma: Introducing new state-of-the-art open models

Gemma is built for responsible AI development from the same research and technology used to create Gemini models.

Feb 21, 2024 · 3 min read



应对措施（高校）





课程特点

- 内容非常新
- 很难一个讲师覆盖
- 与业界合作紧密
- 讲者年轻化
- 需要更多资源支持

Week 1

Jan 9:	LECTURE Introduction	Percy Liang
Jan 11:	SPEAKER FlashAttention	Tri Dao (Stanford)
<i>Reading materials</i>		
<ul style="list-style-type: none"> • The Illustrated Transformer • Attention is All You Need (Optional, but recommended!) • Making Deep Learning Go Brrrr From First Principles. • NVIDIA deep learning performance guide. (Optional) 		

Week 2

Jan 16:	HOLIDAY MLK Day	
Jan 18:	SPEAKER Model Training	Ce Zhang

Week 3

Jan 23:	LECTURE Scaling Laws and Emergent Behavior	Tatsu Hashimoto
Jan 25:	SPEAKER Pathways Language Model (PaLM) and Model Scaling	Aakanksha Chowdhery (Google Brain)

Week 4

Jan 30:	SPEAKER Mechanistic Interpretability – Reverse Engineering Learned Algorithms from Transformers	Stella Biderman (EleutherAI), Booz Allen Hamilton
Feb 1:	SPEAKER A data-centric view on reliable generalization	Ludwig Schmidt

Week 5

Feb 6:	LECTURE Security + Privacy	Tatsu Hashimoto
Feb 8:	SPEAKER Building ML Models like Open-Source Software	Colin Raffel (UNC, Hugging Face)

Thursday, Oct 12: LLMs for Music Generation (**Guest Lecture from Chris Donahue**)

Readings: [AudioLM](#), [Oore+ 18](#)

Tuesday, Oct 17: **Fall break (no class)**

Readings: Enjoy Fall Break

Thursday, Oct 19: **Fall break (no class)**

Readings: Enjoy Fall Break

Tuesday, Oct 24: **Anonymous Feedback for the Class. LLM for Search Engines Continued** **HW2 DUE** [Slides](#)

Readings: [Scaling Laws](#), [Chinchilla](#)

Thursday, Oct 26: **Paradigms of Visual Representation Learning** [Slides](#) (**Guest Lecture from Xinlei Chen**)

Readings: [SimCLR](#), [MAE](#)

Tuesday, Oct 31: **Midpoint Project Presentations** [Project Midpoint Slides](#) **PROJECT MIDPOINT DUE**

Thursday, Nov 2: **Scaling Up: Scaling Laws** [Slides](#)

Readings: [Adam](#), [8-bit Adam](#)



课程学习目标

- 掌握以大语言模型为核心的生成式人工智能**基础概念和技术**
 - Transformer、提示学习、预训练、续训练、指令精调、安全与对齐、奖励函数、基于人类反馈的强化学习、评估、智能体等
- 具备**实现**大模型网络结构和基于大模型的具体应用的能力
- 培养学术创新思维和学会“**写故事**”与“**讲故事**”
- 理解由生成式人工智能引发的**技术革命的本质**、历史和未来发展，为在相关领域的进一步工作做好准备



课程基本介绍

- 主页: <https://gair-nlp.github.io/cs2916/>
- 课程形式
 - 每个章节都会推荐一些阅读材料, 建议提前阅读
 - 课后问题讨论
 - Canvas论坛或
 - 课程讨论微信群
 - 课程一共有四个小作业, 最终课程成绩将会根据权重确定得分



课程团队: 讲师 & TAs

角色	姓名	邮箱	OH时间	OH地点
刘鹏飞	讲师	pengfei@sjtu.edu.cn	周五15:30-16:30	剑川路930 A 栋 3楼313B
陈奕群	助教	ethanicchern@gmail.com	周一13:00-14:00	Zoom
周凡	助教	koala99.zf@gmail.com	周二13:00-14:00	Zoom
刘一秀	助教	1770418133@qq.com	周四13:00-14:00	Zoom
夏世杰	助教	xiasj20@fudan.edu.cn	周三13:00-14:00	Zoom
李学峰	助教	xuefengli0301@gmail.com	周五13:00-14:00	Zoom



课程团队: 特邀讲师

Guest Lecturers (授课顺序)

- **邱锡鹏** (复旦大学教授, 上海高校青年教师教学竞赛优等奖, 著作《神经网络与深度学习》, MOSS大模型领导者)
- **闫宏秀** (上海交通大学教授, 主要研究方向: 技术哲学、数据伦理、大模型安全对齐)
- **魏忠钰** (复旦大学数据智能与社会计算实验室负责人, 上海市青年教师教学比赛优秀奖, 主要研究方向: 多模态智能交互)
- **颜航** (浦江国家实验室青年研究员, 浦语大模型InternLM核心完成人)
- **刘知远** (清华大学计算机系副教授, 清华大学教学成果一等奖, 大模型技术相关领域取得多项创新成果)



课程作业

- 该课程一共有四个作业，其中作业1必须是单人完成，作业2，3，4必须由2-3人的团队完成。如果有组队上的困难，请联系老师或者TAs。

作业	合作要求	预期产出	得分占比
作业1	个人	打包代码，系统输出结果文件，Github链接	10%
作业2	团队 (2-3)	打包代码，系统输出结果文件，PDF英文报告(~5页)，Github链接	10%
作业3	团队 (2-3)	打包代码，PDF英文报告(~7页)，Github链接	20%
作业4	团队 (2-3)	打包代码，PDF英文报告(~7页)，Github链接，PPT，海报	60%



提交细节

- 通过Canvas提交作业，打包成一个包含如下信息的“zip”包
 - 代码
 - 都放在名称为“code”的文件夹里面
 - 系统输出结果
 - 针对作业1和作业2，格式具体任务中会指定
 - PDF英文报告
 - 针对作业2, 3, 4，命名为“report.pdf”
 - pdf需要为英文，推荐使用Overleaf (latex) 进行编辑 ([模板](#))
 - 允许使用chatgpt等AI工具进行辅助润色
 - 一个存放github链接的txt文件（比如命名为github.txt）
 - PPT和海报（只有作业4需要）
- 作业2,3,4迟交天数总计不超过5天



作业细节

- 作业1: Build Your Own LLaMa (from CMU CS11-747)
 - 设计动机: 通过实践作业深入理解基础网络结构

- 作业2: Build Your Own Lima
 - 设计动机: 通过自己构建数据、精调模型和评估实现大模型的对齐

- 作业3: Project Proposal State-of-the-art Reimplementation
 - 设计动机: 接触感兴趣话题的最新方法

- 作业4: Final Project
 - 设计动机: 培养学术创新能力、演讲和表达



课程资源

□ GPU

- [阿里ModelScope](#): 免费用100(36 + 64)小时的32GB GPU资源

- [Google colab](#)

- 更多资源需求可以联系老师

□ OpenAI API

□ ChatGPT Account



课程大纲

Week	Date	Content	Homework	Lecturer
1	2024/02/22	Why LLMs?	-	Pengfei Liu
2	2024/02/29	Neural Networks and Deep Learning Basics	-	Pengfei Liu
3	2024/03/07	Language Models and Representation Learning	👉 hw1 out	Pengfei Liu
4	2024/03/14	Transformers and Pretrained Language Models	-	Xipeng Qiu
5	2024/03/21	Prompting Engineering	-	Pengfei Liu
6	2024/03/28	Evaluation	👉 hw2 out, hw1 due	Pengfei Liu
7	2024/04/04	Qingming Festival	-	-
8	2024/04/11	Supervised Finetuning and Instruction Tuning	-	Hang Yan, Pengfei Liu
9	2024/04/18	Reward Model and RLHF	-	Pengfei Liu
10	2024/04/25	Alignment and AI Safety	👉 hw2 due, hw3,4 out	Hongxiu Yan, Pengfei Liu
11	2024/05/02	Multi-modal LLM	-	Zhongyu Wei, Pengfei Liu
12	2024/05/09	Agents	-	Zhiyuan Liu
13	2024/05/16	Long-context LLM	-	Pengfei Liu
14	2024/05/23	Retrieval Augmented Generation	👉 hw3 due	Pengfei Liu
15	2024/05/30	Presentation	-	-
16	2024/06/06	Presentation	-	-
17	2024/06/13	Poster	👉 hw4 due	-