

主導課程七：自然語言處理 (Natural Language Processing)

課程基本資料

開設學校：清華大學

開授教師：高宏宇

班級人數：_ 2360 _ 人 (保留 _ 200 _ 人給開課學校，條件式授權聯盟校不限，封閉式授權每校40人)

開課級別：研究所

授課語言：中文

學分數：3

授權方式：混成式

若選擇混成式，請附註混成式條件：

協同教師須經主導教師同意，條件式授權盟校：成績考核方式由協同教師自訂。

上課時間：星期二13:20-15:10、星期四13:20-14:10

是否接受非同步授課：是

遠距上課位置：<https://cool.ntu.edu.tw/courses/41436/modules>

課程網頁：https://github.com/IKMLab/NTHU_Natural_Language_Processing

協同教師學經歷建議：

聯盟學校修課人數與助教比例：每_40_名學生需_1_名助教

課程概述

本課程旨在介紹自然語言處理 (NLP) 和大型語言模型 (LLM) 的基礎知識和前瞻技術，適合對自然語言技術感興趣的學生。隨著生成式人工智慧技術的快速發展，NLP在各個領域中的應用日益廣泛。

本課程將提供學生NLP理論基礎，並結合實際應用，幫助學生掌握最新的NLP與LLM技術。課程內容主要分為以下幾個部分：

1. 文字處理基礎：介紹NLP的基本概念和常用技術。教學基本的文字處理技術，如分詞、詞性標註、命名實體識別等。
2. 機器學習模型：機器學習基本概念和算法，如線性回歸、決策樹、隨機森林等。介紹如何將機器學習應用於NLP，包括文本分類、情感分析等。
3. 語言模型：語言模型的基本概念與原理，如N-gram模型、Word2Vec等。深度學習在語言模型中的應用，如RNN、LSTM、Transformer等架構。詳細講解BERT、GPT等先進語言模型，並探討其在不同NLP任務中的應用。
4. 前瞻大語言模型技術：介紹大型語言模型的發展歷程與最新研究進展，如GPT-3等。探討這些模型的訓練方法、大規模資料集的使用，以及在不同領域中的能力。介紹輕量化微調技術 (PEFT)，如LoRA，並說明其在提高訓練效率和效果方面的優勢。

參考書目

無 (自編講義)

課程內容大綱

週次	日期	課程內容	備註
1	9/2、9/4	課程簡介 Syllabus / Introduction to NLP	

2	9/9、11	自然語言處理簡介 (1/2) Introduction to NLP (vector space, indexing, parts of speech, phrase structure)	
3	9/16、9/18	自然語言處理簡介 (2/2) Introduction to NLP (Language model)	
4	9/23、9/25	基礎文字資料機器學習 (1/2) Basic machine learning for text (Text Classification, NB, NN)	
5	9/30、10/2	基礎文字資料機器學習 (2/2) Basic machine learning for text (word embedding, text representation)	
6	10/7、10/9	Python for text tutorial (1/2)	本週不直播上課，將有課程錄影與學習進度，請學生自行學習。
7	10/14、10/16	Python for text tutorial (2/2)	
8	10/21、10/23	文字生成式AI簡介(1/3) Introduction to GAI (text): Word embeddings, Language Modeling (RNN), Sequence-to-sequence Models, and Attention Mechanisms, Sub-word Tokenization; Transformers	
9	10/28、10/30	文字生成式AI簡介(2/3) Introduction to GAI (text): ELMo, BERT, GPT, and T5 (BERT and its Family)	
10	11/4、11/6	文字生成式AI簡介(3/3) Introduction to GAI (text): Decoding Strategies	

		and Evaluations for Natural Language Generation	
11	11/11、11/13	大語言模型簡介與訓練 (1/4): Large language model concept and training (GPT-3, InstructGPT, RLHF)	
12	11/18、11/20	大語言模型簡介與訓練 (2/4): Parameter Efficient Fine-Tuning (PEFT)	
13	11/25、11/27	大語言模型簡介與訓練 (3/4): Introduction and Review technique of Retrieval Augmented Generation (RAG) I	
14	12/2、12/4	大語言模型簡介與訓練 (4/4): Introduction and Review technique of Retrieval Augmented Generation (RAG) II	
15	12/9、12/11	Application Implementation & Challenges	
16	12/16、12/18	Application Implementation & Challenges	

成績評量方式

- Homework x 5 75%
- Term Project x1 25%

課程要求

建議學生須先具 Python 程式語言的先備知識。

協同老師負責工作

條件式授權衛星課程的協同老師，需要獨立完成該盟校所有學生的評分，包含考試卷批改、分組簡報評分、期末專題評分等等，各盟校評分獨立作業。

【自然語言處理】的協同老師，不需要同步跟課，但是需要找尋適合的助教以及協助下列評分，也因此盟校至少要給予一半的授課學分數，若該校學生除了主導課程老師的上課內容外，還需要協同老師另外補充上課，則需給予更高的學分數。