114學年第1學期隨班附讀學分班第一期
進修目的說明書

2025年美國碩士班申請準備計畫

個人檔案與相關連結

- **GitHub：** github.com/goog-msft-fb-nflx-nvda-aapl
- **LinkedIn：** linkedin.com/in/james-goog-jp-ms-cl-uw/
- **學業成績・認定資格：** https://bit.ly/4fpCrMp
- **Credly 證書：** credly.com/users/james_hnd_2025/badges

一、進修目的與申請目標

本人有志於申請2025年底美國頂尖大學碩士班，專注於人工智慧、電腦科學、語言科技等前沿領域。此進修計畫旨在：

- **學術準備：**透過系統性課程學習，建立堅實的理論基礎與實作能力
- **研究經驗累積：**完成高品質期末專案，展現獨立研究與創新能力
- **申請資料強化：**豐富個人陳述(SOP)與研究作品集(Portfolio)內容
- **未來研究準備：**為碩士階段的深入研究奠定基礎

目標申請學校與科系

MS CS, Stanford, CA　MS CL, UW, Seattle, WA　MS AI, UT Austin, TX　MS CS in DS, UIUC, IL

二、修課計畫與理由

經詳細研讀課程大綱與評分標準，選擇以下三門核心課程，涵蓋 大型語言模型應用 、 機器學習安全 、 強化學習與RLHF 等關鍵領域：

① ADL — 深度學習之應用

課程網址：csie.ntu.edu.tw/~miulab/f113-adl/

課程內容重點：

- Transformer與BERT架構深入解析
- 自然語言生成技術與應用
- Prompt-based Learning方法學
- 大型語言模型微調與推論最佳化

實作項目：

- 三次程式作業：BERT實作、自然語言生成、LLM微調
- 期末專案展示（佔總成績35%）
- 探索主題：Retrieval-Augmented Generation、Pre-trained Model robustness

預期收穫：

掌握PyTorch與Hugging Face框架，具備大型語言模型開發與應用能力，為未來研究portfolio與SOP提供關鍵素材。

② SPML — 機器學習安全特論

課程網址：csie.ntu.edu.tw/~stchen/teaching/spml24spring/

課程內容重點：

- 對抗性機器學習理論與實務
- 隱私保護機器學習技術
- 涵蓋evasion attacks、certified defenses、poisoning、model stealing
- 差分隱私(Differential Privacy)與公平性(Fairness)議題

評量方式：

- 多次主題簡報與技術溝通訓練
- 期末專題提案與展示
- 研究論文閱讀與批判性思考能力培養

研究興趣連結：

與本人對大型語言模型與對話系統安全的研究興趣密切相關，特別是模型反演攻擊防範、對抗性提示攻擊對多模態模型的影響與防禦策略。

③ RL — 強化學習

課程資訊：docs.google.com/forms/d/e/1FAIpQLSdfBCjKbBkZRSI1114_Ye6UUDmRVWjroSZAcAjc3a7ogmsA/viewform

課程架構：

- 強化學習理論基礎建立
- 深度強化學習進階技術
- Model-based RL方法學
- 從人類回饋學習強化學習(RLHF)

特色安排：

- 多項程式作業與期末專案實作
- 邀請MIT、Stanford、Berkeley、DeepMind專家講座
- 提供產學研接軌機會

與既有經驗互補：

可與本人於Stanford修習之XCS234 RL課程相互補強，深化對policy gradient、function approximation與RL系統設計的理解。

三、相關修課準備與學術背景

詳細成績與證書記錄：http://bit.ly/3H3TxCK

國內課程成績 (國立台灣大學)

課程代碼	課程名稱	成績
CSIE5400	人工智慧	A
EE5184	機器學習	A-
EE5200	生成式人工智慧導論	A+
CSIE7430	高等深度學習	A+
CSIE7421	高等電腦視覺	A+
CSIE5042	自然語言處理	A+
DATA5014	自然語言與資訊檢索於社群網路之應用	A+
HDAS7004	統計與機器學習	A
CSIE5137	網路資訊檢索與探勘	A

國外課程成績 (Stanford AI Professional Program)

XCS234 Reinforcement Learning

- **講師：**Prof. Emma Brunskill
- **成績：**293.5 / 300
- **課程內容：**tabular methods、function approximation、policy gradient、fast RL
- **完成作業：**四項程式作業
- **成績報告：**https://bit.ly/49ZKVXz
- **區塊鏈證書：**https://bit.ly/4ombPQj

XCS224N NLP with Deep Learning

- **講師：**Prof. Christopher Manning
- **成績：**220 / 200
- **完成作業：**五項程式作業（GloVe、Word2Vec、Parsing、Translation、Transformers）
- **成績報告：**https://bit.ly/4bQZVrA
- **區塊鏈證書：**https://bit.ly/4ovpqA3

XCS224U Natural Language Understanding

- **講師：**Prof. Christopher Potts
- **區塊鏈證書：**https://stanford.io/4fpjnh9

四、預期學習成果與未來規劃

短期目標（課程期間）

- 深入掌握大型語言模型架構設計與應用開發技術
- 建立機器學習安全與隱私保護的理論基礎與實務經驗
- 強化強化學習理論理解，特別是RLHF在對話系統中的應用
- 完成三個高品質期末專案，展現跨領域整合能力

中期目標（申請準備階段）

- 將課程專案成果整理為研究作品集，強化申請競爭力
- 撰寫具體且具說服力的個人陳述，展現研究熱忱與學術潛力
- 建立與指導教授的學術關係，獲得推薦信支持
- 參與相關研究社群與學術活動，擴展專業網絡

長期目標（碩士階段研究方向）

- 專注於可信賴AI系統開發，結合安全性與實用性考量
- 探索多模態大型語言模型在對話系統中的應用與挑戰
- 研究人類回饋機制在AI系統調整中的理論與實務問題
- 推進AI安全與公平性技術的產業應用與社會影響評估

結語

本人誠摯希望能同時修習上述三門核心課程，透過系統性的理論學習與實務訓練，設計並實作具有學術價值與產業應用潛力的期末專案。這些學習經驗將成為未來研究生涯的重要基石，不僅能夠強化申請材料的競爭力，更重要的是為投入AI領域的前沿研究做好充分準備。

我相信透過這三門課程的深入學習，能夠在大型語言模型應用、機器學習安全、以及強化學習等關鍵領域建立紮實的理論基礎與實作能力，為實現在人工智慧領域貢獻創新研究的長期目標奠定堅實基礎。