

淡江大學 114 學年度第 1 學期課程教學計畫表

課程名稱	人工智慧倫理	授課 教師	林嘉琪 LIN, CHIACHI
	AI ETHICS		
開課系級	A I 四 A	開課 資料	實體課程 選修 單學期 3學分
	TKFXB4A		
課程與SDGs 關聯性	SDG4 優質教育 SDG10 減少不平等 SDG16 和平正義與有力的制度		
系 (所) 教育目標			
<p>一、教育學生運用程式、數學及人工智慧知識以分析科學與應用之相關問題。</p> <p>二、訓練學生透過問題分析、實驗執行、數據解釋與推導演繹規劃與實作人工智慧系統，以解決科學與應用之相關問題。</p> <p>三、教導學生能夠獨立完成任務及具備團隊合作精神之人工智慧工程師，使其專業素養與工作倫理能充分發揮於職場。</p> <p>四、督促學生具備全球競爭的基本技能，以面對不同的生涯發展，並能持續終身學習。</p>			
課程簡介	<p>人工智慧 (AI) 在我們的生活中愈來愈普及，本課程將從人工智慧的歷史、基本原理、技術發展，討論人工智慧和機器學習中出現的倫理問題。例如AI的責任問題、算法中的公平性和偏見，以及使用AI進行決策。我們還將探討更具體的問題，例如自動駕駛車輛的風險、隱私和虛假信息。本課程的目標是使學生瞭解人工智慧的基本原理、演進歷史和學會批判性地思考我們世界中出現的許多重要倫理問題。</p> <p>Artificial intelligence(AI) is becoming more and more popular in our lives. This course will discuss the history, basic principles, and technological development of artificial intelligence and the ethical issues that arise in artificial intelligence and machine learning. The goal of this course is for students to understand the basic principles and evolutionary history of artificial intelligence and learn to think critically about the many important ethical issues that arise in our world.</p>		

本課程教學目標與認知、情意、技能目標之對應

將課程教學目標分別對應「認知 (Cognitive)」、「情意 (Affective)」與「技能(Psychomotor)」的各目標類型。

- 一、認知(Cognitive)：著重在該科目的事實、概念、程序、後設認知等各類知識之學習。
- 二、情意(Affective)：著重在該科目的興趣、倫理、態度、信念、價值觀等之學習。
- 三、技能(Psychomotor)：著重在該科目的肢體動作或技術操作之學習。

序號	教學目標(中文)	教學目標(英文)
1	了解人工智慧的倫理和社會影響，發展道德觀點，並能夠參與討論和決策相關議題。	Understand the ethical and social implications of artificial intelligence, develop a moral perspective, and be able to participate in discussions and decision-making on relevant issues.

教學目標之目標類型、核心能力、基本素養教學方法與評量方式

序號	目標類型	院、系(所)核心能力	校級基本素養	教學方法	評量方式
1	情意	ABCD	12345678	講述、討論	測驗、作業、討論(含課堂、線上)

授課進度表

週次	日期起訖	內容 (Subject/Topics)	備註
1	114/09/15~114/09/21	導論：AI倫理有什麼好談的？課程與AI發展史簡介	
2	114/09/22~114/09/28	AI倫理發展背景、技術相關哲學倫理：人工智慧的過去、現在與未來：倫理基本概念與AI國際倫規範簡介	閱讀文獻 (AIE) Ch1_Introduction: Why AI Ethics?
3	114/09/29~114/10/05	批判性倫理思維方法與工具介紹 人類、智能機器與倫理價值：AI資料的來源與分類之倫理議題	閱讀文獻 (AIE)Ch2_The Ris of AI Ethics
4	114/10/06~114/10/12	規範道德理論與AI相關哲學倫理：AI模型的倫理與社會議題	閱讀文獻 (AIE)Ch5_Humans and Intelligent Machines: Underlying Values
5	114/10/13~114/10/19	人類自主與AI	閱讀文獻 (AIE)Ch6_Normative Ethical Theory and AI Ethics
6	114/10/20~114/10/26	AI偏誤與公平性	閱讀文獻： (AIE) Ch8_Persons and AI 參考文獻
7	114/10/27~114/11/02	歐盟AI法案	閱讀文獻： (MAI)Ch4_Can AI be fair?

8	114/11/03~ 114/11/09	AI與安全問題、責任鴻溝及應對策略	閱讀文獻： (MAI)Ch2_Can AI be safe? (MAI)Ch5_Can AI (or its creators or users) be responsible? (MAI)Ch7_What can we do?
9	114/11/10~ 114/11/16	期中考/期中評量週(老師得自行調整週次) AI與人類價值對齊	
10	114/11/17~ 114/11/23	AI演算法與社群媒體	
11	114/11/24~ 114/11/30	AI演算法與社群媒體	閱讀文獻： (MAI)Ch6_Can AI incorporate human morality?
12	114/12/01~ 114/12/07	AI對工作及未來的社會影響	閱讀文獻： (AIE)Ch10_Towards the Future with AI: Work and Superintelligence
13	114/12/08~ 114/12/14	AI機器人相關倫理議題	閱讀文獻：Lin, P., Abney, K., & Jenkins, R. (2017). Robot ethics 2.0: From autonomous cars to artificial intelligence.
14	114/12/15~ 114/12/21	大型語言模型與生成式AI相關倫理議題 Earp, et al., 'Credit and blame for AI-generated content: Effects of personalization in four countries', New York Academy of Sciences, 1542(1): 51-57	閱讀文獻： (MAI)Ch3_Can AI respect privacy?
15	114/12/22~ 114/12/28	AI影像相關倫理議題	
16	114/12/29~ 115/01/04	期末多元評量週	
17	115/01/05~ 115/01/11	期末多元評量週/教師彈性教學週	
18	115/01/12~ 115/01/18	教師彈性教學週	
課程培養 關鍵能力	自主學習、資訊科技、人文關懷、跨領域		
跨領域課程	STEAM課程(S科學、T科技、E工程、M數學，融入A人文藝術領域)		
特色教學 課程			
課程 教授內容	A I 應用 永續議題		

<p>修課應注意事項</p>	<p>1.請勿睡覺、飲食、看漫畫、打撲克牌、玩掌上型電玩、手機、聊天喧嘩等等。以上皆列入平時考評分數，一次扣總成績3分。 2.點名時間原則上在課程開始後20分鐘、第一節下課、第二節下課之間不定期施行，點名當下未到者，不予補點名。 3.點名未到達全學期點名次數之記錄二分之一以上者，總成績不予及格。 4.作業絕對不得抄襲，轉引資料必定要註明出處，而且禁止整段複製-貼上，抄襲查獲者，此部分成績即不予計算。 5.有問題非常歡迎隨時寫信給老師，報告與信件皆請註明姓名、系級、星期幾、第幾節課程的學生，以方便老師解答，回覆。 6.本課程講義，教師視上課進度，保有異動權。請勿轉載。</p>
<p>教科書與教材</p>	<p>自編教材:簡報 採用他人教材:簡報 教材說明: 參考資料: 1.第三屆點子松 金點子作品簡介 2.東海大學哲學系粉絲頁(畢業作品展示) 3.數位發展部公部門人工智慧應用參考手冊(草案) 4.甘偵蓉&陳煒林(2024/2/23), 天下雜誌獨立評論; 甘偵蓉在科學月刊有關ChatGPT評論1,2,LaMDA事件</p>
<p>參考文獻</p>	<p>凱特.克勞馥(Kate Crawford) (2022), 第二、三、四章, 出自《人工智慧最後的秘密》, 臉譜文化, pp. 71-108。 泰娜.布策(Taina Bucher) (2021), 第一、二、四章, 出自《被操弄的真實: 演算法中隱藏的政治與權力》, 台灣商務印書館。 凱西.歐尼爾(Cathy O’Neill)(2109), 第一、十章, 出自《大數據的傲慢與偏見》, 大寫出版。 約蘭德(Yolande Strengers) & 珍妮.甘迺迪(Jenny Kennedy) (2023), 第一、二章, 《智慧妻子》, 陽明大學交通出版社。 尼克.伯斯特隆姆(Nick Bostrom)(2016), 第十三、十四、十五章, 出自《超智慧》, 八旗文化。 甘偵蓉(2023), 〈人工智慧科研倫理與風險之基本認識〉, 《科技、醫療與社會》季刊, 37: 167-220。 甘偵蓉(2024), 〈AI開發過程的倫理權衡: 自駕車決策案例研究〉, 中研院《歐美研究》季刊第54卷第1期, 頁1-68。DOI: https://doi.org/10.7015/JEAS.202403_54(1).0001 (THCI/TSSCI) 甘偵蓉(2024), 〈人工智慧系統應該內建倫理嗎? 人工道德行為者之探討〉, 《危機時代的哲學—「後」疫情時期的反思》, 五南圖書出版。ISBN: 978-626-393-008-7 Borg, J. S., Sinnott-Armstrong, W., & Conitzer, V. (2024). Moral AI: And How We Get There. Random House. Gabriel, I. (2020). Artificial intelligence, values, and alignment. Minds & Machines, 30, 411 - 437. https://doi.org/10.1007/s11023-020-09539-2 Russell, S., & Norvig, P. (2021). Chap. 1 & Chap. 27. In Artificial intelligence: A modern approach (4th ed.). University of California, Berkeley. Russell, S. (2019). Chap. 7 & Chap. 10. In Human compatible: Artificial intelligence and the problem of control. Penguin.</p>
<p>學期成績計算方式</p>	<p>◆出席率: 25.0 % ◆平時評量: 25.0 % ◆期中評量: 25.0 % ◆期末評量: 25.0 % ◆其他〈 〉: %</p>
<p>備考</p>	<p>「教學計畫表管理系統」網址: https://web2.ais.tku.edu.tw/csp 或由教務處首頁→教務資訊「教學計畫表管理系統」進入。 ※「遵守智慧財產權觀念」及「不得非法影印、下載及散布」。請使用正版教科書，勿非法影印他人著作，以免觸法。</p>