

國立勤益科技大學通識教育學院

114 學年度 1 學期 教學大綱

部別	<input checked="" type="checkbox"/> 日間部 <input type="checkbox"/> 進修部 <input type="checkbox"/> 推廣部	學制	<input checked="" type="checkbox"/> 四技 <input type="checkbox"/> 二技 <input type="checkbox"/> 二專
授課教師	郭宇婕	教師學歷	國立臺灣師範大學圖書資訊博士
教師經歷	國立勤益科技大學資訊工程系助理教授	教師級職	助理教授
科目名稱(中)	人工智慧倫理與治理		
科目名稱(英)	Artificial Intelligence Ethics and Governance		
開課單位	<input type="checkbox"/> 基礎通識教育中心 <input checked="" type="checkbox"/> 博雅通識教育中心	學分/學時數	2/2
領域	<input type="checkbox"/> 人文藝術 <input type="checkbox"/> 社會科學 <input checked="" type="checkbox"/> 自然科技		
優質課程類別	<input checked="" type="checkbox"/> 一般課程、 <input type="checkbox"/> 智慧財產權、 <input type="checkbox"/> 內涵式服務學習課程、 <input type="checkbox"/> 性別平等、 <input type="checkbox"/> 綠色課程 <input checked="" type="checkbox"/> 創新、創意課程、 <input type="checkbox"/> 工作(職場)倫理課程、 <input type="checkbox"/> 工具機技術研發 創新、創意課程定義：課程目標為「激發學生獨特的想像與創意思考，透過企劃與執行以創新模式解決實際問題。」		
科目與通識核心能力關聯	<input checked="" type="checkbox"/> 知識統整能力 <u>30</u> % <input checked="" type="checkbox"/> 創意思維能力 <u>35</u> % <input checked="" type="checkbox"/> 溝通表達能力 <u>35</u> % <input type="checkbox"/> 美感鑑賞能力 _____ % <input type="checkbox"/> 邏輯推理能力 _____ % <input type="checkbox"/> 法治思辨能力 _____ % <input type="checkbox"/> 博通宏觀能力 _____ % <input type="checkbox"/> 倫理關懷能力 _____ % (核心能力定義請參見附件一，請選擇 2~3 項相關程度較高之核心能力)		
科目屬性	<input type="checkbox"/> 核心課程 <input type="checkbox"/> 跨領域課程(須符合附件二定義，並請勾選下一欄) <input checked="" type="checkbox"/> 生活性課程 <input type="checkbox"/> 學術性課程 <input checked="" type="checkbox"/> 通論性課程 <input type="checkbox"/> 經典性課程 (屬性定義請參見附件二，可複選)		
跨領域課程	<input type="checkbox"/> 人文藝術領域： <input type="checkbox"/> 社會科學領域： <input checked="" type="checkbox"/> 自然科技領域： <u>100%</u> (以上總和百分比須達 100%)		
教科書	自編教材		
參考書目	<p>Borg, J. S., Sinnott-Armstrong, W., & Conitzer, V. (2024). Moral AI: And how we get there. Random House.</p> <p>Coeckelbergh, M. (2020). AI ethics. MIT Press.</p> <p>Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017, February). Moral decision making frameworks for artificial intelligence. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).</p> <p>Leslie, D., Rincon, C., Briggs, M., Perini, A., Jayadeva, S., Borda, A., ... & Fischer, C. (2024). AI ethics and governance in practice: An introduction. arXiv preprint arXiv:2403.15403.</p>		

	<p>Liu, Z., & Zheng, Y. (2022). AI ethics and governance. Springer Singapore.</p> <p>Wang, W., & Siau, K. (2018). Ethical and moral issues with AI.</p> <p>甘偵蓉. (2022). 人工智慧的倫理與風險：以遠距監考系統為例. Journal of Education Research (1680-6360), (341).</p> <p>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，(37)，167-219。 https://doi.org/10.6464/TJSSTM.202310_(37).0004</p> <p>甘偵蓉. (2023). 為何應該讓人工智能強化倫理衝突的緊急決策?. 資訊社會研究, (45), 19-50.</p> <p>甘偵蓉. (2024). AI 開發過程的倫理權衡：自駕車決策案例研究. EurAmerica, 54(1).</p> <p>蔡甫昌, & 胡嘉輝. (2020). 人工智慧醫療應用與倫理準則. 澄清醫護管理雜誌, 16(2), 4-8.</p>
教學目標	<p>在完成本課程後，學生應該能夠：</p> <ol style="list-style-type: none"> 1. 認識人工智慧倫理與治理的概念和意義。 2. 瞭解人工智慧倫理治理的核心與原則、挑戰和發展。 3. 理解人工智慧的可問責性(Accountability)與透明度(transparency) 4. 嚴謹評估 AI 治理策略，批判性地評估其起源、應用及有效性 5. 利用上課所學知識與技能，完成一個 AI 倫理或 AI 治理專題報告
評量方式	<p>課程參與(40%) 期中報告(30%) 期末報告(30%)</p> <p>學生將被要求積極參與作業相關的討論。課堂參與度的評分將根據學生討論表現來評量。</p>
內容綱要	<p>這門課程旨在深入探討人工智慧 (AI) 技術所引發的倫理困境及其在現代社會中的治理需求。通過結合理論框架與實際案例分析，課程將幫助學生理解 AI 對社會、經濟、法律等各方面的深遠影響。學習重點包括如何尊重人類的自主權、預防潛在傷害、維護公平性與可解釋性等基本倫理原則。此外，課程將挑戰學生思考如何將道德決策內建於人工智慧系統中，並剖析這一過程中的倫理障礙。學生還將學習一套符合公平、透明、問責且尊重人權的原則，並針對未來在 AI 領域可能面臨的倫理與社會挑戰，進行深入探討與分析。</p>
教學方式	<p>講授與互動式討論</p>
創新教學活動設計	<p>(若有的話，請敘述本科目融入那些創新的教學活動設計)</p>

科目進度與內容

(勿只填寫單元名稱，請簡述內容)

週次	教學內容 ※申請跨領域課程時，將特別針對教學內容細節審核，請詳細說明，以作為審核依據	備註 (課程活動與作業) ※請務必填寫	※若勾選「跨領域課程」請標註每週次涵蓋領域，可複選
1	課程介紹與基本概念 1. 課程簡介與評估方式 2. 人工智慧 (AI) 的基礎知識與應用範疇	討論 AI 倫理現狀並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
2	AI倫理的核心 1. 尊重人類自主性原則 2. 預防傷害原則 3. 公平原則 4. 可解釋性原則	批判性分析 AI 倫理的核心並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
3	公平性 (FAIRNESS) 1. AI 技術效率與公平性之間的取捨 2. AI 部署中出現偏見的道德義務	批判性分析 AI 公平性並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
4	透明性 (TRANSPARENCY) 1. AI 所具備的能力與製造目的必須經過公開的協商 2. 可解釋性原則 (The Principle of Explicability)	批判性分析 AI 透明性並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
5	當責 (ACCOUNTABILITY) 1. 將道德決策編入機器的倫理挑戰 2. 當這些決策導致傷害時的法律與倫理當責架構。	批判性分析 AI 當責並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
6	人工智慧幻覺 (AI Hallucination) 1. AI 偏見與歧視及虛構引用 2. AI 真假難辨的隱憂	互動式討論 AI 幻覺並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
7	AI 責任歸屬 1. 當 AI 出錯時，誰該負責？ 2. AI 造成傷害時責任如何歸屬？	批判性分析 AI 責任歸屬問題並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
8	期中回顧與專題報告預備 1. 期中報告預備 2. 小組互動式專題討論	選擇一個 AI 倫理或 AI 治理待解決問題，提出可能的解決方案並課後提交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
9	期中研討與專題報告 1. 期中報告 2. 小組互動式專題討論	專題計劃書撰寫與上課學習反思	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域

10	人工智慧治理 (AI Governance) 1. 制定公平、透明、問責和尊重人權的原則 2. 建立AI法規框架以規範AI的開發和使用	互動式討論AI治理並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
11	人工智慧對使用者隱私的威脅 1. 如何減少對使用者資料隱私的侵犯並且價值最大化 2. 如何保護AI使用者資料隱私	批判性分析AI對隱私之威脅並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
12	人工智慧控制 (CONTROL) 1. AI治理與潛在風險之間的平衡 2. AI演算法偏差	互動式討論AI控制並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
13	AI 資料倫理 1. AI技術應用的倫理挑戰 2. 探討AI系統資料的管理和保護，包括資料隱私和安全問題	互動式討論AI資料倫理並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
14	人工智慧強化倫理衝突的緊急決策 1. 人類如何借助AI做出較佳的倫理決策 1. 從資料驅動到AI驅動	互動式討論並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
15	人工智慧開發過程的倫理權衡 1. AI設計的倫理權衡問題 2. AI自動決策系統開發過程涉及倫理權衡	互動式討論並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
16	人工智慧的倫理與社會責任 1. AI技術應用中的倫理挑戰 2. AI倫理框架與社會參與	互動式討論AI在社會責任中的倫理界限並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
17	課程學期總結與反思 1. 回顧整個課程的關鍵概念與學習重點 2. AI在倫理與治理未來挑戰	互動式討論AI倫理與治理未來挑戰並課後繳交上課參與單	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域
18	期末評估與課程結束 1. 期末專題研討 2. 反思與課程回顧 3. 總結與未來學習建議	期末專題研討	<input type="checkbox"/> 人文藝術領域 <input type="checkbox"/> 社會科學領域 <input type="checkbox"/> 自然科技領域

請遵守智慧財產權觀念，不得非法影印

附件一 核心能力定義

核心能力	定義與說明
A. 知識統整能力	學生能在各種知識與文化脈絡中，尋得恆久不變的價值觀，並將此價值觀融入其生活，進而認識、欣賞、尊重與珍惜生命的意義。
B. 創意思維能力	學生能認知各知識領域與多元文化間的差異處與鏈結點，進而具備跨領域思維與評判能力，使其能在固有的架構中，呈現嶄新的創造力。
C. 溝通表達能力	學生能釐清自我思想，並藉由正確且清楚的語文表達理念，以建立與他人良好的溝通。
D. 美感鑑賞能力	學生能認知、接收並傳達多元藝術美感，具備敏銳的鑑賞能力，並運用在不同領域的統整中。
E. 邏輯推理能力	學生能依據自身認知和客觀事實，運用邏輯分析與量化推理，進行反思與論證，進而做出合理判斷。
F. 法治思辨能力	學生能正確認知人權、民主、與法治之互動關聯，進行獨立思辨與論辯且基於人本關懷精神，以確立其自身與社會群體之關係。
G. 博通宏觀能力	學生能以基礎知識為本，培養前瞻性的觀點並開拓宏博的視野，以建立整全之人生觀。
H. 倫理關懷能力	學生能認知自身與所處環境的關係，並進而願意以己身之力與專業知識參與社會與環境的改造，提升正向能量。

附件二 課程屬性定義

核心課程：全校性共同必修之通識課程。

跨領域課程：課程內容須跨人文藝術/社會科學/自然科技三領域其中之二項以上。課程須有一主領域，其授課內容須達 60%以上，上限為 70%以下。(依 110 年 4 月 7 日 109-2 博雅通識中心第 1 次教評會議決議)

生活性課程：課程重點強調知識應用與人類生活相關之課程。

學術性課程：課程重點偏重理論發展之脈絡、思想之沿革、與歷史文化背景之因素。

通論性課程：針對特定領域或時代的知識與思想做綜觀性的介紹，與廣博性的探討。

經典性課程：針對特定領域或時代具有代表性的人物、思想、典籍做較為深入之探討、剖析、或導讀。