

圖片來源：[維基百科](#)

AI科技對學術研究與誠信的影響

柯皓仁

國立臺灣師範大學圖書資訊學研究所
教授兼學習資訊專業學院副院長

clavenke@ntnu.edu.tw

內容大綱

- ▶ 人工智慧(AI)是甚麼?
- ▶ AI與學術誠信 – 用AI
- ▶ AI與研究倫理 – 發展AI
- ▶ AI研究與IRB/REC
- ▶ 人體研究倫理委員會審查涉及人工智慧研究的考慮因素
- ▶ 結論



人工智慧是甚麼？

人工智慧

▶ 泛指普通電腦程式來呈現人類智慧的技術

- ✿ 長期目標為建構能夠跟人類似甚至超卓的推理、知識、規劃、學習、交流、感知、移物、使用工具和操控機械的能力

▶ 弱人工智慧

- ✿ 「不可能」製造出能「真正」地推理和解決問題的智慧機器，這些機器只不過「看起來」像是智慧的，但是並不真正擁有智慧，也不會有自主意識
 - ✓ 影像辨識、語言分析、棋類遊戲

▶ 強人工智慧

- ✿ 「有可能」製造出「真正」能推理和解決問題的智慧機器，並且，這樣的機器將被認為是具有知覺、有自我意識的

人工智慧(續)

EU guidelines on ethics in artificial intelligence: Context and implementation

- ▶ 人工智慧包括機器學習、神經網絡、計算機視覺、專家系統、自然語言處理和機器人技術
 - ✿ 搜索和分析大量資料的機器學習技術
 - ✿ 旨在使用知識表示和推理解決複雜問題的專家系統
 - ✿ 處理可程式化機器的概念、設計、製造和操作的機器人技術
 - ✿ 能夠預測人類和機器行為並做出自主決策的演算法和自動決策系統

Ethical Issues in Research with Artificial Intelligence Systems

- ▶ 根據其自主程度和智慧水準，可以將AI系統分為不同類別，從簡單的基於規則的系統到能夠學習並適應新情況的高級系統
- ▶ 人工智慧系統被理解為能以類似智慧人類行為的方式，自動處理數據和資訊的能力，通常包括推理、學習、感知、預測、規劃或控制等方面

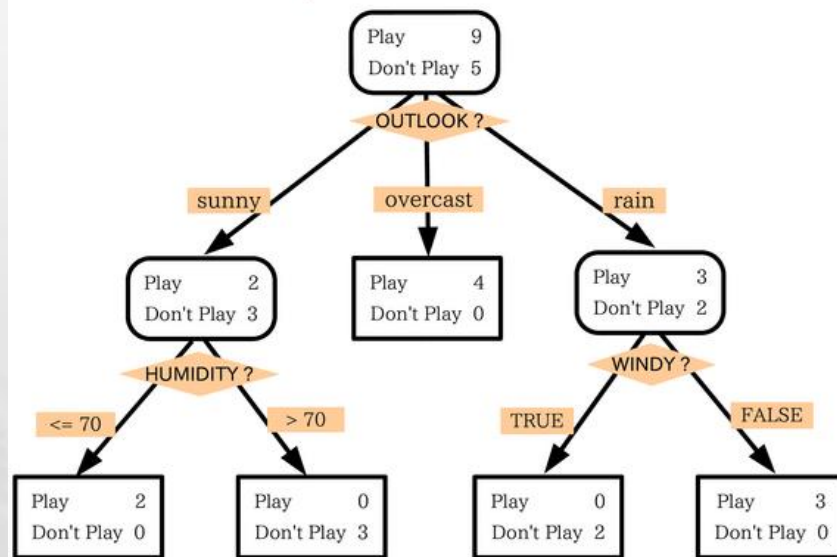
Principles for the Ethical Use of Artificial Intelligence in the United Nations System

決策樹 (Decision Tree)

Play golf dataset

Independent variables				Dep. var
OUTLOOK	TEMPERATURE	HUMIDITY	WINDY	PLAY
sunny	85	85	FALSE	Don't Play
sunny	80	90	TRUE	Don't Play
overcast	83	78	FALSE	Play
rain	70	96	FALSE	Play
rain	68	80	FALSE	Play
rain	65	70	TRUE	Don't Play
overcast	64	65	TRUE	Play
sunny	72	95	FALSE	Don't Play
sunny	69	70	FALSE	Play
rain	75	80	FALSE	Play
sunny	75	70	TRUE	Play
overcast	72	90	TRUE	Play
overcast	81	75	FALSE	Play
rain	71	80	TRUE	Don't Play

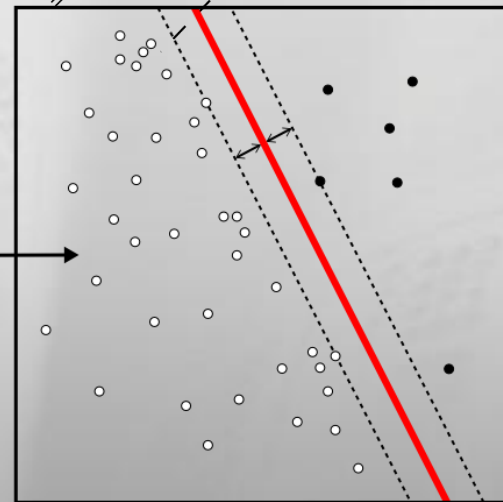
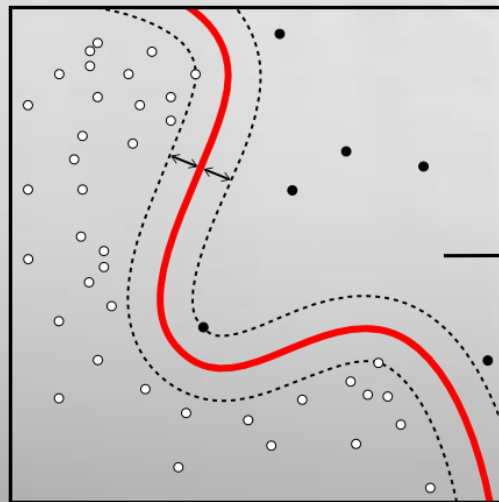
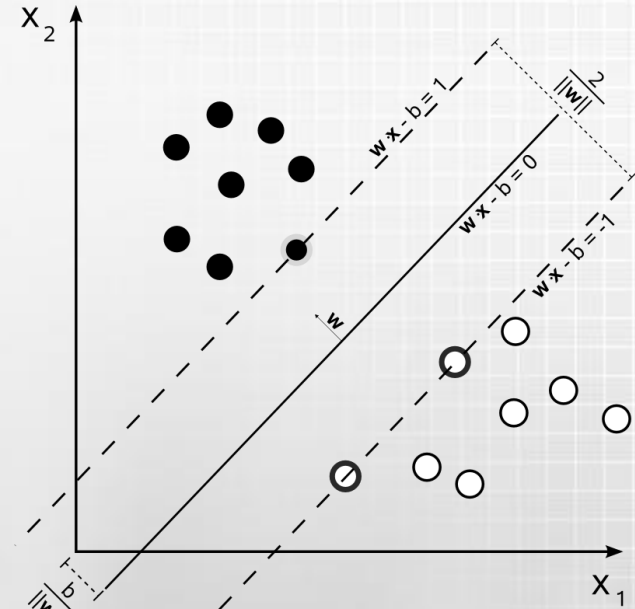
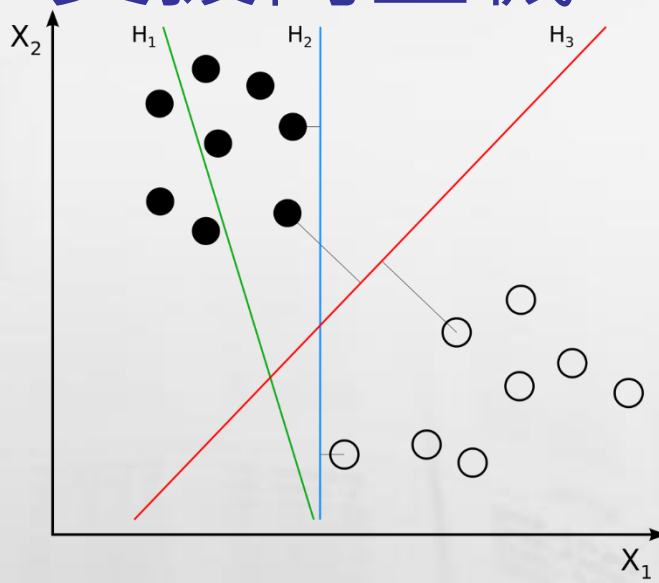
Dependent variable: PLAY



圖片來源

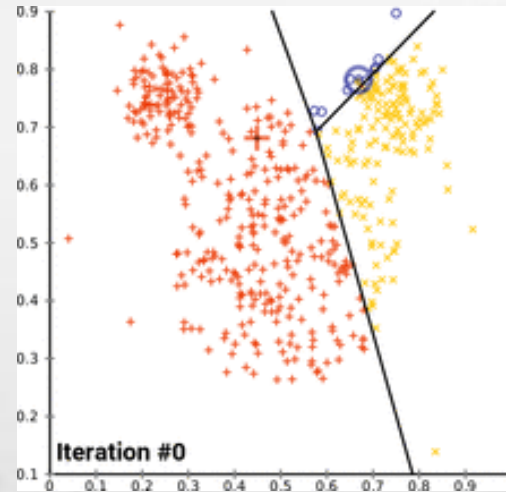
分類 (Classification) —

支援向量機 (Support Vector Machine)



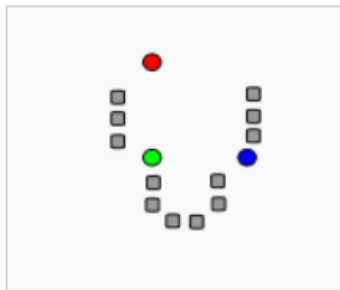
圖片來源

分群 (Clustering) – K-Means

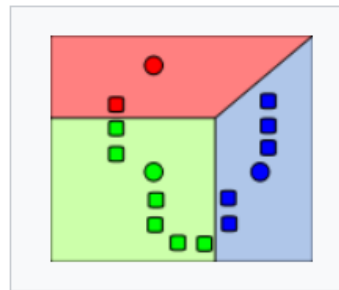


圖片來源

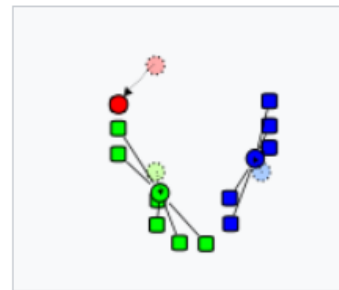
Demonstration of the standard algorithm



1. k initial "means" (in this case $k=3$) are randomly generated within the data domain (shown in color).



2. k clusters are created by associating every observation with the nearest mean. The partitions here represent the [Voronoi diagram](#) generated by the means.

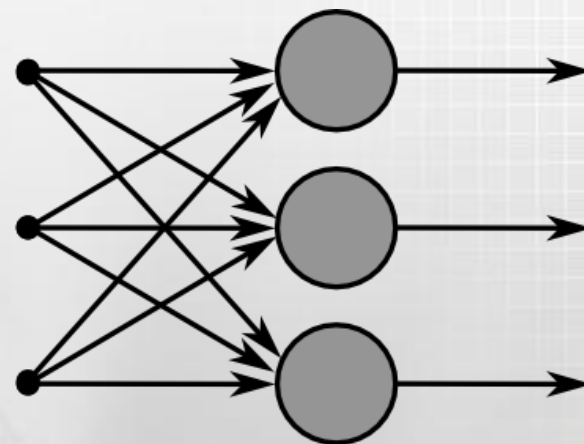
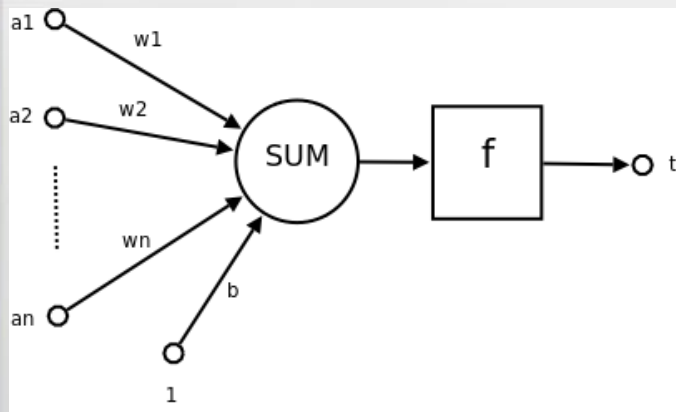


3. The [centroid](#) of each of the k clusters becomes the new mean.



4. Steps 2 and 3 are repeated until convergence has been reached.

類神經網路 (Artificial Neural Networks)



output layer

深度學習 (Deep Learning)
大型語言模型 (Large Language Model)

圖片來源

生成式AI (Generative AI)

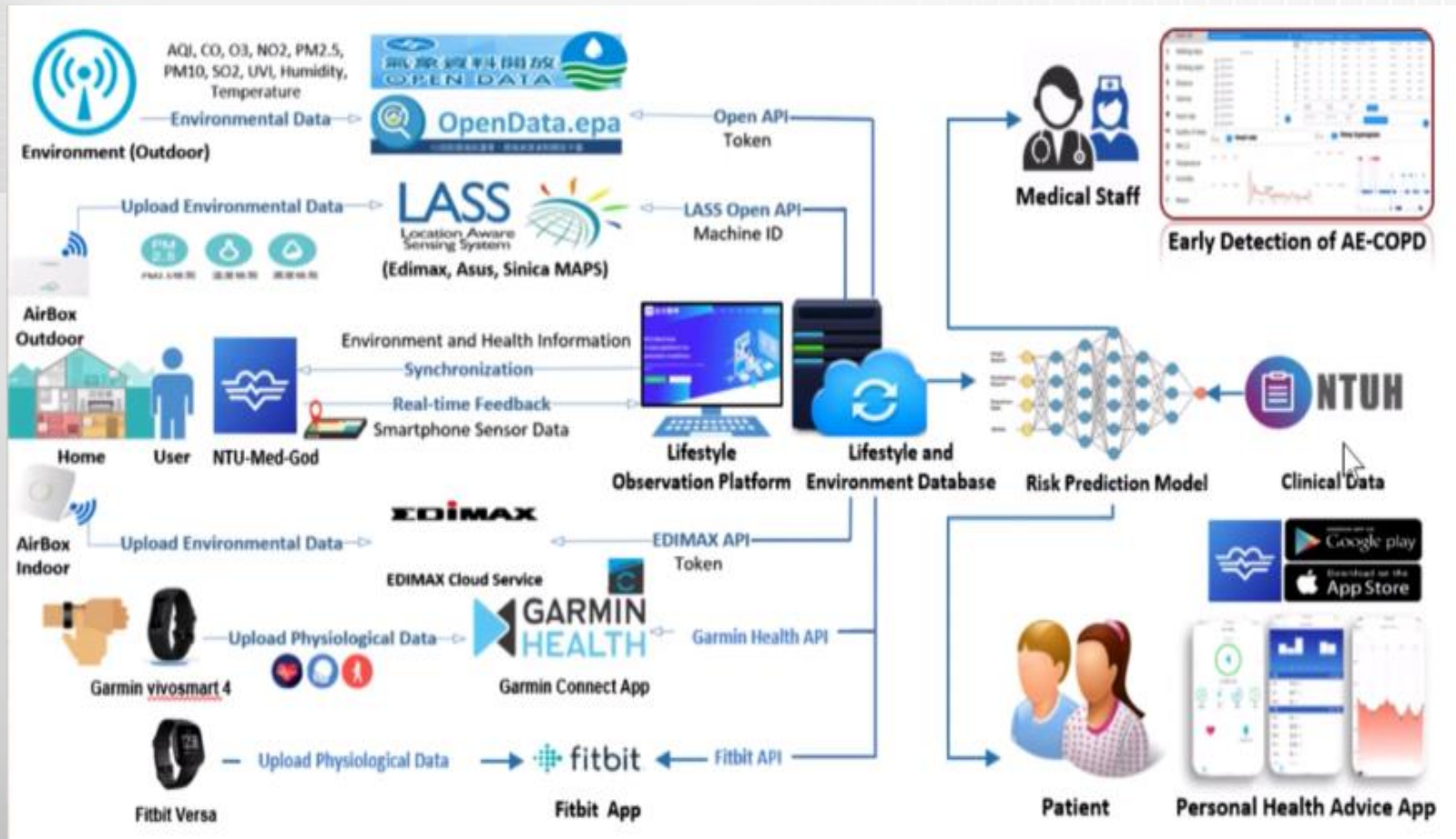
- ▶ BARD、ChatGPT、DALL-E、Midjourney...
- ▶ GPT: Generative pre-trained transformers

ChatGPT是一種基於人工智慧的語言生成模型，運用了深度學習技術。其運作原理是通過訓練大規模的神經網路，該網路能夠預測文本序列中的下一個單詞或字符，使其能夠生成自然流暢的文本回應。ChatGPT能夠理解和處理自然語言輸入，並根據上下文產生符合語境的回應，且具有一定的語言理解能力。其功能包括根據給定提示生成文本、提供多樣性的回答、支持不同主題的對話、並能模擬人類對話。ChatGPT被廣泛應用於對話生成、自然語言處理和應用，並在教育、研究和商業領域中發揮重要作用。

-- Response by ChatGPT

API整合AI加速精準醫療

資料來源：[iThome](http://iThome.com)



利用環境資料、生活數據和患者電子病歷，來預測患者未來7天AECOPD的發生率，模型準確率達87.7%

AI 中心核心目標

- 將人工智慧落地臨床醫學
- 輔助醫師診斷
- 提升人力資源的效率
- 遠距醫療AI
- 智慧醫院之建構
- 提高治療準確度



AI 中心研發走向



醫療影像
分析

- AI醫學影像輔助診斷 (如：X-ray、CT、MR、超音波、步態分析影片…等)，提供相當重要資訊供醫師參考，並協助醫療團隊給予相應的醫療處置。



生理數據
訊號分析

- 長時間的連續性紀錄，或數據之動態變化，確認疾病診斷。



自然語言
處理技術

- 將非結構化的醫療文本轉化為包含重要醫學信息之結構化數據，從而提高醫療系統的品質，減少運行成本。
- 亦可輔助醫務管理應用。



生物資訊
領域應用

- 以單一核苷酸多型性 (Single Nucleotide Polymorphism, SNP) 為基因標記進行遺傳疾病的全基因體關連性分析。
- 基因資訊搭配醫學影像的多體學研究 (Multi-Omics)。
- 多種癌症基因突變類別之預測、治療反應、局部復發或遠端轉移等預後指引。

2019-2021重要研發成果

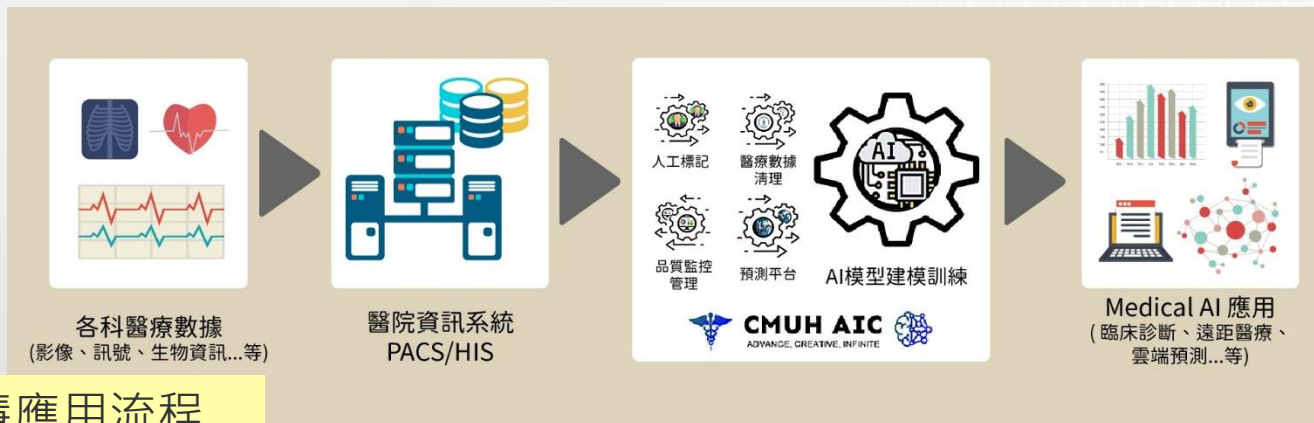
- 4大 AI臨床運用
- 12項 AI門診上線
- 7項 專案取得專利
- 2項 專案申請專利中



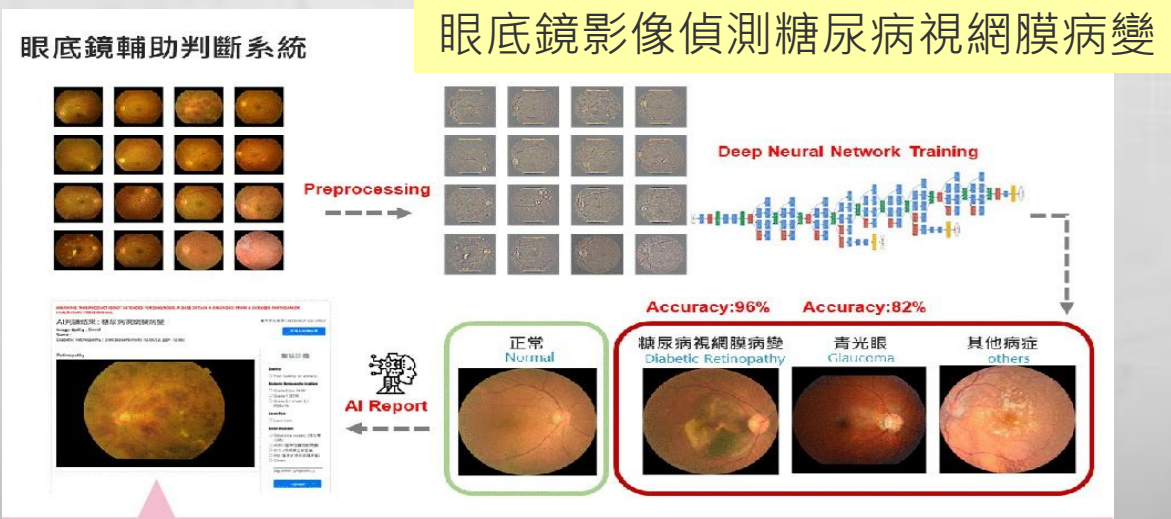
中國醫藥大學附設醫院
大數據智慧醫療

中國醫藥大學附設醫院

大數據智慧醫療 (續)



醫療數據人工智慧應用流程



醫師回饋

臨床科醫師若對當下AI判讀結果有相左的意見，可輸入醫師自己評斷之分級，回饋給系統進行模型優化校正。

中國醫藥大學附設醫院

大數據智慧醫療 (續)



Inference亦可提供『是否有結石』之結果！

WARNING: THIS PRODUCT IS NOT INTENDED FOR DIAGNOSIS. PLEASE OBTAIN A DIAGNOSIS FROM A LICENSED PHYSICIAN OR HEALTHCARE PROFESSIONAL.
僅供研究使用 | RESEARCH USE ONLY

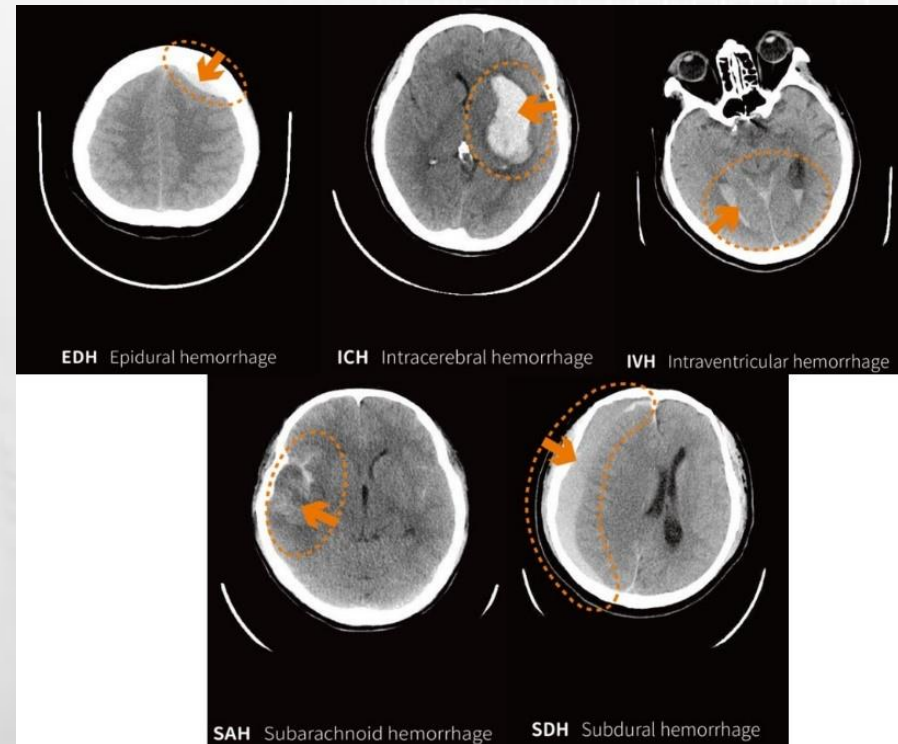
Inference Results
 Not CALC STONE

WARNING: THIS PRODUCT IS NOT INTENDED FOR DIAGNOSIS. PLEASE OBTAIN A DIAGNOSIS FROM A LICENSED PHYSICIAN OR HEALTHCARE PROFESSIONAL.
僅供研究使用 | RESEARCH USE ONLY

Inference Results
 BMD = 0.843777046029689 g/cm²
 診斷結果為：骨量充足。建議於中後期定期檢查。
 備註：骨質疏鬆、骨質增生或Osteoporosis患者不宜定期檢查骨質疏鬆程度。

醫師可參考 AI Inference Report中之BMD 值輔助診斷

骨密度判定模型



顱內出血AI判讀



生成式AI與教育

(生成式)AI的好處 – 學生

▶ 課業學習

- ✿ 文章重點整理
- ✿ 完善知識體系
- ✿ 提高學習成效
- ✿ 提供多元觀點

▶ 作業報告

- ✿ 文章架構擬定
- ✿ 文章改寫
- ✿ 文章校對

(生成式)AI的好處 – 教師

- ▶ 課前準備
 - ✿ 課綱設計
 - ✿ 教材準備
- ▶ 教學現場導入
 - ✿ 討論文本生成
 - ✿ 統整結論
 - ✿ 人機討論
- ▶ 建立多元評量方式
- ▶ 精進評量品質

(生成式)AI的好處 — 學者

- ▶ 起草/修改 電子郵件與回覆
- ▶ 撰寫一流的研究人員簡歷
- ▶ 簡化複雜主題
- ▶ 文章重點摘要
- ▶ 為獎助申請腦力激盪
- ▶ 撰寫社交媒體貼文
- ▶ 為您的研究提出類比
- ▶ 訂正語法
- ▶ 文章翻譯

Opperman, K. (2023). [How to use ChatGPT: Opportunities and Risks for Researchers](#)

AI幫你找參考文獻



Company ▾ Keenious for ▾ Universities ▾ Resources ▾

Sign In

Get Keenious!

✔ Cross-language recommendations are now available as a Free Research Preview!

[Learn more](#)

Find research relevant to any document!

Try it now!



PDF

URL

Text

↑ Choose PDF

Drop your PDF file here or use an [example PDF](#)

運用(生成式)AI的注意事項

- ▶ 檢核內容正確性
- ▶ 不可靠的參考資料
- ▶ 風險和法律問題
- ▶ 謹守學術誠信
- ▶ 訓練偏差
- ▶ 注意資訊安全
- ▶ 避免過度依賴
- ▶ 目前生成式AI對2021年後的知識較為缺乏

不要全盤接受



請推薦給我五部精采的日本棒球漫畫



以下是五部精采的日本棒球漫畫，它們在棒球迷和漫畫愛好者中都非常受歡迎：



1. 《大甲子園》 ("Ookiku Furikabutte")

這部漫畫由新谷洋平創作，講述了一支高中棒球隊的成長故事，強調團隊合作和友誼。

2. 《鑽石之王牌》 ("Diamond no Ace")

由寺島優創作的這部漫畫以一名年輕投手的棒球生涯為中心，探討了他在高中棒球隊的成長和挑戰。

3. 《棒球大聯盟》 ("Major")

這部漫畫由滿田拓也創作，講述了主角從兒童棒球開始一直到成為職業棒球選手的旅程，強調奮鬥和成長。

4. 《夏目友人帳》 ("Natsume's Book of Friends")

由綠川ゆき創作的這部漫畫雖然主要講述妖怪和人類之間的故事，但也包括了一些觸動人心的棒球場景。

5. 《草薙舞衛》 ("Koushien")

這部漫畫由有河サキ創作，講述了一支來自不同背景的高中棒球隊的成長故事，充滿

🔄 Regenerate



AI與學術誠信 – 用AI

學術誠信原則

- ▶ 可靠性 (Reliability)：確保研究的品質，反映在設計、方法論、分析和資源的使用上
- ▶ 誠實 (Honesty)：以透明、公平、充分和公正的方式發展、執行、審查、提報和傳播研究
- ▶ 尊重 (Respect)：尊重同事、研究參與者、研究對象、社會、生態系統、文化遺產和環境
- ▶ 問責 (Accountability)：對於研究從想法到出版、管理和組織，培訓、監督和指導，以及對其更廣泛的社會影響承擔責任

ALLEA (2023) The European Code of Conduct for Research Integrity – Revised Edition 2023. Berlin. DOI 10.26356/ECOC

學術誠信核心價值觀

- ▶ 誠實 (Honesty) : 誠實追求真理和知識
- ▶ 信任 (Trust) : 培養互信、鼓勵交流想法，使所有人能夠發揮潛能
- ▶ 公平 (Fairness) : 建立明確的標準、慣例和程序，體現公平公正
- ▶ 尊重 (Respect) : 尊重多元意見和想法
- ▶ 責任 (Responsibility) : 提倡個人責任，面對不法行為時要有所行動
- ▶ 勇氣 (Courage) : 勇敢面對壓力和逆境

AI與學術誠信

- ▶ 作者端：使用諸如ChatGPT之類的應用程式來生成資料 並/或 編寫論文
- ▶ 期刊出版社：使用AI進行文章的品質檢查，例如剽竊偵測
- ▶ AI具有潛在的好處，但為了維護學術誠信，科學家和機構需要透明地使用AI工具

[European research integrity code updated to reflect advances in artificial intelligence](#)

運用(生成式)AI機器翻譯的倫理問題

- ▶ 抄襲 – 翻譯的隻字片語是哪來的？
- ▶ 資料保密 – 餵資料先
- ▶ 無偏見語言 – 文化差異和敏感度
- ▶ 捏造資訊 – 似是而非

- ◆ 基本能力還是很重要的
- ◆ 不要仰賴AI進行原創、創見處的詮釋
- ◆ AI訓練資料集的著作權歸屬

好的研究實務

Good Research Practices

▶ 2.3 研究程序 (Research Procedures)

- ✿ 研究人員以符合該學科的公認標準，有助於驗證或複製的方式，報告他們的研究結果和方法，包括外部服務、**人工智慧和自動化工具**的使用(Researchers report their results and methods, including the use of external services or AI and automated tools, in a way that is compatible with the accepted norms of the discipline and facilitates verification or replication, where applicable)

▶ 2.8 審查和評估 (Review and Assessment)

- ✿ 研究人員、研究機構和組織以透明和可證明的方式審查和評估提交的出版物、資助、任命、晉升或獎勵，並披露使用**人工智慧和自動化工具**的情況

ALLEA (2023) The European Code of Conduct for Research Integrity – Revised Edition 2023. Berlin. DOI 10.26356/ECOC

違反研究誠信

Violations of Research Integrity

▶ 3.1 研究不端行為和其他不可接受的實務

- ✿ 隱瞞在創建內容或起草出版物時使用人工智慧或自動化工具。

AI寫的文章能被偵測出來嗎？



Products ▾ Solutions ▾ Resources Support

Turnitin's AI detector capabilities

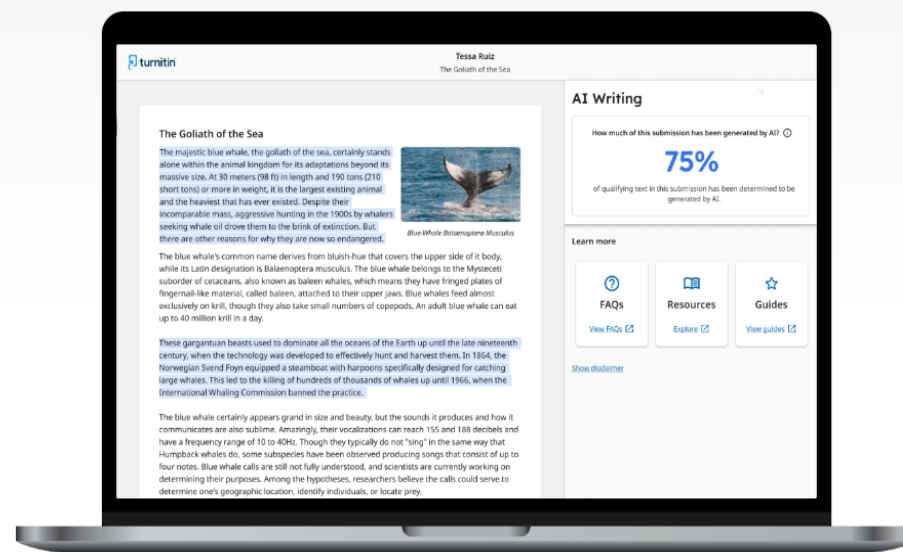
Rapidly innovating to uphold academic integrity

Identify when AI writing tools such as ChatGPT have been used in students' submissions.

AI writing detection is available to customers using Turnitin Feedback Studio (TFS), TFS with Originality, Turnitin Originality, Turnitin Similarity, Simcheck, Originality Check, and Originality Check+.

From January 2024, the AI detector will only be available to Turnitin customers when licensing Originality with their existing product.

[Learn more about Originality](#)



偵測原理：Writing Style

《Science》對人工智慧的聲明

- ▶ 在《Science》期刊發表的論文中，不得使用AI、機器學習或類似的算法工具生成的文本，除非經編輯明確許可。此外，論文中的圖表、圖像或圖形也不得來自這些工具，同樣需要編輯的明確許可。此政策也規定，AI程序本身不得作為《Science》期刊論文的作者。違反此政策將視為科學不端行為。

《 Nature 》 對人工智慧的聲明

Tools such as ChatGPT threaten transparent science; here are our ground rules for their use

- ▶ 任何語言模型(LLM)工具將不被接受為研究論文的正式作者。這是因為對作者身份的歸因意味著對工作的負責，而人工智慧工具無法承擔這種責任
- ▶ 使用LLM的作者應該在論文的方法或致謝記錄此使用情況。如果一篇論文未包含這些部分，則可以記錄在引言或其他適當的部分

Authors. Corresponding author(s) should be identified with an asterisk. Large Language Models (LLMs), such as [ChatGPT](#), do not currently satisfy our [authorship criteria](#). Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript.

Nature – Initial submission



AI與研究倫理 – 發展AI

AI的偏見

紐西蘭想打造大數據演算模型受惠國家，但最後卻變成一個帶有「偏見」的無情機器

2014年中國大陸國務院提出《社會信用體系建設規畫綱要》，為在2020年達成全面實施社會信用體系的目標，各地政府陸續廣泛蒐集人民的資料，利用大數據分析方式，從政務誠信、商務誠信、社會誠信、司法公信四大領域，對個人和企業組織進行評分。據2018年3月澳洲廣播公司報導，已有民眾在不明原因的情況下被列入失信名單，以至於旅行嚴重受阻或無法借貸，甚至資產被凍結。

無獨有偶地，據2018年4月紐西蘭媒體報導，紐西蘭移民署（Immigration New Zealand）從十八個月前開始實行一項實驗計畫（pilot programme），藉由國家簽證申請程序蒐集移民的年齡、性別、種族、犯罪紀錄、醫療費用欠繳紀錄等資料，目的在建立一個演算模型——傷害模型（harm model），用以預測居住在紐西蘭的移民如果續留是否從事犯罪行為或帶來更多醫療支出。一旦模型顯示某些移民可能對國家造成治安疑慮或增加醫療負擔，這些移民將無法重新申請簽證，更甚者，將直接被驅逐出境。

紐西蘭移民署表示這個實驗計畫有利於政府及早採取行動，然而不少移民質疑演算模型帶有偏見。一位印度裔移民的家屬向紐西蘭移民署申請居留被拒，原因在於，移民署透過演算模型相信該印度裔移民的母親過去曾經患有乳癌，但實際上，他的母親不曾患過任何關於乳房組織的病變。此外，一些律師與人權團體亦因為演算模型納入種族資料，強烈抨擊移民署的計畫背後隱藏著種族歧視。

何謂研究倫理

- ▶ 進行學術研究時必須遵守的行為規範，亦是用來評估研究者從事研究的各項行為是否符合社會規範的原則
 - ✿ 抄襲、洩漏隱私、不實論文作者掛名、代寫論文、引用不實數據、扭曲研究結果、未經同意採取檢體、刻意將論文鎖住不公開讓人閱覽、藉由審查論文/計劃書之便剽竊對方構想
- ▶ 研究倫理為何受到重視？
 - ✿ 來自人權與婦女等運動對於人性價值的反思
 - ✿ 為防止違反倫理，惡名昭彰的研究案例再度發生
 - ✓ 納粹醫學實驗等

研究倫理實踐準則

- ▶ 傷害最小化 (無害性)
- ▶ 知情同意 (資格能力、志願性、資訊完全、理解)
- ▶ 隱私與保密
- ▶ 避免欺騙
- ▶ 避免雙重關係及利益衝突

人工智慧與機器人倫理的主要爭點

- ▶ 隱私與監視
- ▶ 行為操控
- ▶ AI系統的不透明性
- ▶ 決策系統中的偏見
- ▶ 自動化與就業(automation and employment)
- ▶ 自主系統(autonomous systems)
- ▶ 機器倫理(machine ethics)
- ▶ 人造道德代理程序(artificial moral agents)
- ▶ 奇異點(singularity)

隱私與監視

Privacy & Surveillance

- ▶ 對於隱私資料、可辨別個人身份資料的取用
- ▶ 感應技術產生更多有關個人生活的非數位化資料
- ▶ 人工智慧增加了智慧數據收集和數據分析的可能性
- ▶ 照片與影片中的人臉識別，進而蒐集或搜尋個人資料
- ▶ 設備指紋(Device Fingerprinting)、數位蹤跡
- ▶ 電腦/系統比我們更瞭解自己
- ▶ 人工智慧搭配與物聯網、智慧城市、智慧治理所衍生的數據收集議題

行為操控

Manipulation of Behavior

- ▶ 人工智慧時代的利用數據引導、操縱、欺騙個人或群體
- ▶ 從商品、遊戲的廣告、行銷到政治操作
- ▶ DeepFake圖片到文本
- ▶ 人工智慧中的機器學習技術依賴於大量的資料培訓。這意味著通常會在隱私和資料權利與產品技術品質之間存在權衡
 - ✿ TO BE OR NOT TO BE
 - ✿ EITHER...OR...

AI系統的不透明性

Opacity of AI Systems

- ▶ 不透明和偏見是資料倫理的核心議題
- ▶ 用於自動決策支援和預測分析的人工智慧系統引發了關於法律正當程序、問責、社區參與和稽核不足的重大擔憂
- ▶ 受影響的人通常無法知道系統是如何產生這個結果的，也就是說，對於這個人來說，系統是**不透明的**。如果系統涉及機器學習，即使是專家也通常無法知道特定模式是如何被識別的，甚至不知道模式是什麼

決策系統中的偏見

Bias in Decision Systems

- ▶ 商業、醫療保健和其他領域的決策與預測分析
 - ✿ 餐廳的喜好、醫療診斷、信用卡核卡、保釋...
- ▶ 這些系統可能延續了已經存在於用於建立系統的數據中的偏見
 - ✿ 歧視婦女、歧視黑人
- ▶ 用來訓練的資料集是否有偏差？
 - ✿ 如果資料集有太多白人男性的相片...

使用人工智慧的倫理原則

Principles for the Ethical Use of Artificial Intelligence in the United Nations System

- ▶ 不造成傷害 (Do not harm)
- ▶ 定義目的、必要性和適當性 (Defined purpose, necessity and proportionality)
- ▶ 安全和保密 (Safe and security)
- ▶ 公平和非歧視 (Fairness and non-discrimination)
- ▶ 永續性 (sustainability)
- ▶ 隱私權、數據保護和數據管理 (Right to privacy, data protection, and data governance)
- ▶ 人類自主權和監督 (Human autonomy and oversight)
- ▶ 透明度和可解釋性 (Transparency and explainability)
- ▶ 責任和問責 (Responsibility and accountability)
- ▶ 參與和包容 (Inclusion and participation)

人工智慧的關鍵倫理要求

- ▶ 人類代理機制與監管
- ▶ 技術的穩固性與安全性
- ▶ 隱私與數據保護
- ▶ 透明度
- ▶ 多樣性、非歧視與公平性
- ▶ 社會與環境的福祉
- ▶ 問責

以人為中心
(Human Centric)

[EU guidelines on ethics in artificial intelligence: Context and implementation](#)

人類代理機制與監管

Human Agency and Oversight

▶ 尊重人類自主權和基本權利

- ✿ 開方AI系統前應進行基本權利影響評估。之後應建立檢核機制，並允許回饋反饋
- ✿ 使用者應能夠滿意地理解AI系統並與之互動。使用者的權利不應受到僅基於自動處理的決策的影響
- ✿ 機器無法完全控制，因此，應總是需要人類監督。人們應始終具有最終推翻系統決策的可能性

技術的穩固性和安全性

Technical Robustness and Safety

- ▶ 擁有安全、可靠、和穩健的系統和軟體，能夠應對AI系統整個生命週期中出現的錯誤或不一致性
 - ✿ 確保網路安全(cybersecurity)：漏洞、網路攻擊、駭客入侵
 - ✿ AI開發人員應建立能夠評估安全風險的流程，以防有人將他們正在構建的AI系統用於有害目的
 - ✓ 人類控制接管並中止系統繼續運作

隱私和數據保護

Privacy and Data Protection

▶ AI利害相關者必須遵守GDPR

✿ 最小限度利用個資、最大程度賦權用戶、最透明的決策機制

▶ 在構建和運行AI系統時確保隱私和個人數據受到保護

▶ 公民應對自己的數據擁有完全控制權，他們的數據不應用於傷害或歧視他們

▶ AI開發人員應應用設計技術，如數據加密和數據匿名化。此外，他們應確保數據的品質，即避免社會構建的偏見、不準確、錯誤和失誤。

✿ 數據收集不應有偏見，AI開發人員應建立監督機制來控制數據集的品質

透明度

Transparency

- ▶ 透明度對確保AI不偏見至關重要
- ▶ 用於構建AI系統的資料集和過程應該被記錄和追蹤
- ▶ AI系統應該被識別為AI系統，人們需要知道他們正在與AI系統互動... 尤其是一個攸關個人權益、基本自由、服務、福利的決定...
- ▶ AI系統和相關的人類決策應受到可解釋性原則的約束，根據該原則，人類應該能夠理解和追溯AI系統的決策

多樣性、非歧視和公平性

Diversity, Non-Discrimination, and Fairness

- ▶ 設計AI產品和服務時要避免不公平偏見
- ▶ AI開發人員應確保他們的演算法設計不帶偏見 (例如，不使用不恰當的資料集)
- ▶ 可能會直接或間接受到AI系統影響的利害相關者應被諮詢並參與其開發和實施
- ▶ 應考慮人類的所有能力、技能和需求，確保身心障礙人士能夠取用AI系統

社會和環境的福祉

Societal and Environment Well-being

- ▶ 應該使用AI系統來促進積極的社會變革，鼓勵AI系統的永續性和環保責任
- ▶ 鼓勵採取措施來確保AI系統對環境友好(例如，選擇較少有害的能源消耗方法)，並應監控和考慮這些系統對社會和民主的影響(包括選舉情境下的影響)

問責

Accountability

- ▶ 應建立機制來確保AI系統及其結果的責任和問責性
- ▶ 應設立內部和外部獨立審計，特別是對於使用影響基本權利的AI系統
- ▶ 應提供AI系統的負面影響報告，並應使用影響評估工具
- ▶ 在實施關鍵道德要求之間可能產生衝突的情況下，應持續評估(傾向道德要求的)權衡決策
- ▶ 應實施可執行的救濟機制

微軟的負責任AI實踐



Fairness

AI systems should treat all people fairly.



Reliability and safety

AI systems should perform reliably and safely.



Privacy and security

AI systems should be secure and respect privacy.



Inclusiveness

AI systems should empower everyone and engage people.



Transparency

AI systems should be understandable.



Accountability

People should be accountable for AI systems.

GDPR與AI

為了因應大數據與人工智慧時代對個人資料保護之衝擊，歐盟一般資料保護規則（General Data Protection Regulation, GDPR）針對自動化決策（automated decision-making）和資料剖析（profiling）已明文規定。依據 GDPR 第 22（1）條規定，純粹基於自動化資料處理（包括資料剖析）所作成的決定，對於資料當事人（data subject）產生法律效果或類似重大影響時，資料當事人應有權免受該決定的拘束。所謂「純粹基於自動化資料處理所作成之決策」，指的是在沒有人為參與的情況下，透過技術方式作出的決策。

然而，GDPR 並不是完全禁止純粹基於自動化資料處理所作成的決策，前提是這種決策必須對資料當事人產生「法律效果或類似重大影響」。例如：人才招聘系統使用預先排程的演算法與標準進行性向測驗，並純粹依此結果決定是否面試特定人時，當事人可能不明究理地被阻擋在求職大門之外，嚴重影響就業機會，這樣的決策將被歐盟所禁止；而借助演算法了解客戶收視習慣進而推薦電視節目的情況，因可能不會產生法律效果或重大影響，則不在 GDPR 第 22 條禁止之列。

此外，如果是基於以下三種情況：（1）簽訂或履行契約所必要（2）經過歐盟或其會員國法律所授權（3）經過資料當事人明示同意，仍可藉由純粹自動化資料處理方式作決策。但是，資料控制者（data controller）和資料處理者（data processor）必須盡到以下義務：（1）主動將自動化決策和資料剖析一事通知資料當事人；（2）實施適當安全措施，例如預先進行資料保護影響評估（Data Protection Impact Assessment, DPIA）並做好風險管理；（3）提供資料當事人權利行使管道，例如當事人異議或申訴程序。

人工智慧決策也會產生「偏見」，人類該如何用法律做好把關？



AI研究與 IRB / REC

Friesen, P., Douglas-Jones, R., Marks, M., Pierce, R., Fletcher, K., Mishra, A., ... & Sallamuddin, T. (2021). Governing AI-driven health research: Are IRBs up to the task? *Ethics & Human Research*, 43(2), 35-42.

AI可以幫助研究倫理審查嗎？

- ▶ 電腦會揀土豆 ~
- ▶ A role for machines?
- ▶ A human in the loop

資料來源：[AI could transform ethics committees](#)

案例一

- ▶ OO公司計劃開發一種演算法來識別自殺風險的用戶。平台的人工智慧將使用自然語言處理來篩選用戶的發文；被分類為高風險的文章將被發送給人類審核者，當他們確定需要干預時，將通知當地當局
 - ✿ 如果在學術環境：涉及敏感的心理健康數據、易受傷害族群，利用的是研究人員無法控制的侵入性干預手段，而用戶並沒有完全被告知該計劃或其風險
 - ✿ 然而，OO公司不在需要IRB審查的法規範圍內，所以僅進行的倫理審查是由OO公司開發和實施的一個過程，即使審查過程在某些方面類似於傳統的IRB，但完全由OO公司的員工進行，通常只涉及直接的管理者

案例二

- ▶ 一位學術研究者計劃開發一個臉部識別工具，用於識別患有胎兒酒精譜系障礙(FASD)的個人。她希望從網路中收集公開可用的患有FASD的個人圖像，將這些照片與沒有FASD的個人圖像混合，並使用這個數據集來訓練一個識別患有FASD的個人的演算法
 - ✿ 有關或可能來自仍然存活的個人的數據/樣本：YES
 - ✿ 所有有關樣本/數據的資訊是否都公開可用：YES
 - ✿ →本計畫不是人類受試者研究→不需要提交IRB
 - ✿ 一旦發表，她開發的分類模型被一個學校委員會採納，以便及早識別患有FASD的兒童，以便老師可以得知哪些學生可能缺乏衝動控制

案例一與案例二的啟示

- ▶ 商業公司倫理委員會的運作是否具透明性、權威性、獨立性
- ▶ AI時代的資料生命週期、研究產出的再利用與誤用、研究產出應用方式超出原先預期、致易受傷害族群於風險中

AI時代IRB/REC的調整建議

▶ 獨立性

- ✿ IRB成員應該與研究團隊及資金提供者有所區別
- ✿ 確保倫理問題的分析不受企業目標的影響

▶ 權威性

- ✿ 商業公司諮詢形式的倫理委員會，但這種委員會將決策權留在公司參與者手中
- ✿ 建議不同於監督

▶ 多樣性

- ✿ IRB成員的多樣化代表性一直是核心問題
- ✿ 為了減少偏見和剝削，有必要包括最有可能受到AI健康研究影響的人群的代表。邀請對曾經進行弱勢、易受傷害族群相關研究學者的參與也可能減少此類研究的不良影響

AI時代IRB/REC的調整建議 (續)

▶ 符合比例原則的審查

- ✿ 倫理審查的存在和程度應與研究中存在的風險和不確定性成比例，並且不依賴於誰提供資金以及研究是否發生在商業背景下
 - ✓ 不僅是指對每個研究參與者的風險
 - ✓ 應包括整個研究計畫的風險，包括數據或技術的潛在次級用途的風險
- ✿ 應避免過度擴展倫理監督。低風險項目可能需要適當的自我審查
 - ✓ 例如，使用數據保護影響評估來證明符合GDPR規範
- ✿ 具不確定影響的高風險研究可能需要更全面的審查
- ✿ 對於提出獨特倫理問題的新形式的研究，由國家、國際或機構委員會監督可能是理想的(例如，與胚胎幹細胞研究委員會)

AI時代IRB/REC的調整建議 (續)

▶ 透明性

- ✿ 在AI驅動的健康研究中劃定容許和不容許的界線的過程應該是理性的、合理的，並且應該公開發表
- ✿ 如此可允許研究者、IRB和外部觀察者互相學習，並使審查過程隨著時間的推移得以改進

▶ 持續監控

- ✿ 資料生命週期差異很大，並且在研究背景中開發的模型可以稍後應用於新資料，從而為毫無戒備的用戶創造新的健康數據
- ✿ IRB的預期審查可能不足夠，在某些情況下可能需要持續監控。這種監控可以由IRB或不同的委員會監督，可能是模仿資料安全監控委員會的運作


AI時代IRB/REC的調整建議 (續)

▶ 標準和工具

- ✿ 採取措施在IRB程序內部和跨部門之間發展一致性
- ✿ 藉由針對研究情境脈絡制定清晰的程序指南，可以提高一致性
- ✿ 工具可用於幫助委員會成員和研究者進行倫理分析和預測，例如演算法影響評估

▶ 效果證據

- ✿ 有關IRB的評鑑



SACHRP 是指美國國家衛生研究院 (NIH) 下屬的「人體研究保護顧問委員會」 (Secretary's Advisory Committee on Human Research Protections) 。該委員會的使命是提供對NIH和衛生部長提供有關人體研究保護政策和規定的專業建議和指導。該委員會由一組來自各學科領域的專家組成，他們就涉及人體研究倫理和法律的問題提供意見和建議。 SACHRP 的目標是確保進行的人體研究符合倫理標準，保護參與者的權益和福祉。

人體研究倫理委員會審查涉及人工智慧研究的考慮因素

資料來源：[Considerations for IRB Review of Research Involving Artificial Intelligence](#)

對SACHRP的指控/批評


- ▶ 在什麼情況下，為AI或AI驗證活動收集數據會符合「旨在發展或貢獻於可推廣知識」的《通用規則》（Common Rule）對研究的定義？
- ▶ 當AI涉及研究私人可識別資訊（PII）時，何時這些人被視為研究對象？該研究是否涵蓋了對於研究對象的「關於誰」的部分？是否有其他對於這些人的道德考慮？
- ▶ 在《通用規則》下，何時收集數據為AI或AI驗證活動通常可獲得豁免？
- ▶ 對於受《通用規則》規定需進行審查的研究，對於在用於AI開發的數據集中包含資訊的人們，最重要的人體研究保護考慮是什麼？這些考慮在研究集中於測試或驗證AI時是否不同？對於那些不是研究對象的人是否還有其他道德考慮？
- ▶ 是否有現有的框架或工具，資助機構、調查人員、HRPP人員和IRB可以使用以闡明並減輕人類導向AI研究和開發的道德問題？

對SACHRP的指控/批評 (續)

- ▶ AI特有的考慮是否會影響研究知情同意書中研究活動的充分披露？
- ▶ AI研究是否有任何獨特之處會需要IRB思考並確定《通用規則》的適用性，而這並不是對於所有人體研究都已經考慮過的。
- ▶ 在具有AI的研究中，45 CFR 46.111的具體部分需要特別關注，例如隱私和保密性、知情同意、風險等方面？
- ▶ 對於機構/HRPP責任而言，與IRB所管轄的其他研究責任相比，AI相關的具體考慮有哪些？
- ▶ 在研究中使用AI是否存在更大的偏見和/或缺陷可能性，IRB應如何考慮這種潛在情況？（例如，臉部識別演算法可能主要基於白人男性，但使用該演算法的研究人員可能不知道這一點）


在什麼情況下，為AI或AI驗證活動收集數據會符合「旨在發展或貢獻於可推廣知識」的《通用規則》（Common Rule）對研究的定義？

- ▶ 當數據收集明確屬於研究計劃書的一部分時，這種收集完全符合《通用規則》對研究的定義。
- ▶ 但是，AI通常使用為其他目的收集的數據，例如醫療紀錄或社交媒體貼文。根據目前的監管框架，這種收集本身並不屬於研究，並且隨後對這些數據的次級使用通常被認為符合《通用規則》第45 CFR 46.104(d)(4)的豁免。當像使用從社交媒體貼文中收集的素材進行研究時，這種豁免的使用尤其棘手，因為這些數據被認為是「公開可得的」，或者像「去識別化」的醫療紀錄一樣，它們是「以不能輕易確定人體研究對象身份的方式進行記錄」。
- ▶ 這種監管方法不一定是錯誤的，但是在大數據和AI普及之前就已開發出來；目前使用這些工具進行的研究正在利用並非針對此目的開發的研究豁免。這一限制在2018年《通用規則》中被明確意識到，並承諾定期重新審查可識別性概念（102(e)(7)(i)）。
- ▶ 因此，許多AI研究是符合要求的，但不一定充分保護研究參與者的權利和福祉。



當AI涉及研究私人可識別資訊 (PII) 時，何時這些人被視為研究對象？該研究是否涵蓋了對於研究對象的「關於誰」的部分？是否有其他對於這些人的道德考慮？

- ▶ 《通用規則》在102(e)(1)條款中將研究對象定義為「一個存活中的個體，研究人員...進行研究時：(i) 通過干預或與該個體互動獲得資訊或生物樣本，並使用、研究或分析資訊或生物樣本；或 (ii) 獲得、使用、研究、分析或產生可識別的私人資訊或可識別的生物樣本。」
- ▶ 此問題明確假設AI研究涉及私人可識別資訊。在這種情況下，應該將這些個體視為研究對象。
- ▶ 如對問題1的回答中所述，當所謂的私人資訊被認為是公開可得時，監管語言存在歧義。在由網際網路、全球資訊網和元宇宙創建的新環境中，傳統上私人的資訊必須作為參與的成本分享，這已成為預期的社會規範，因此傳統上的「私人」和「公共」的定義不應被認為適用。




在《通用規則》下，何時收集數據為AI或AI驗證活動通常可獲得豁免？

- ▶ 見問題一

對於受《通用規則》規定需進行審查的研究，對於在用於AI開發的數據集中包含資訊的人們，**最重要的人體研究保護考慮是什麼？**這些考慮在研究集中於測試或驗證AI時是否不同？對於那些不是研究對象的人是否還有其他道德考慮？

- ▶ 根據目前的監管解釋，只有一小部分研究需要在《通用規則》下進行審查。這樣的研究將被定義為那些被視為「可識別的私人信息」的研究，但不屬於次級使用的豁免範圍。換句話說，這些數據不會「公開可得」，研究對象的身份必須由研究人員「輕易確定」，而且這些數據不能受到其他監管體系的保護，具體指HIPAA或聯邦隱私法。符合這些標準的研究可能會被認為是最低風險的，因為數據的收集和使用已成為日常生活的普遍現實，並且符合加速審查的標準，即類別5(僅為非研究目的收集的材料涉及的研究)。
- ▶ 鑒於目前的實踐和研究的最低風險性質，保護措施可能主要源於對於111(a)(3) - 研究對象的公平選擇和111(a)(7) - 隱私保護的考慮。很可能AI研究，即使符合所有要求以使其受到規則的積極監管，也會符合知情同意豁免，因為如果沒有這種豁免，該研究可能無法實際進行。請注意，根據111(a)(2)的風險/利益計算所提供的保護將受到限制，因為該研究很可能被認為是最低風險，大多數IRB會解釋禁止考慮將所獲得的知識應用於長期影響以防止由個人偏見或現有偏見加強而導致的群體傷害的規定。



對於受《通用規則》規定需進行審查的研究，對於在用於AI開發的數據集中包含資訊的人們，最重要的人體研究保護考慮是什麼？這些考慮在研究集中於測試或驗證AI時是否不同？對於那些不是研究對象的人是否還有其他道德考慮？(續)

- ▶ 在那些信息包含在用於AI開發的數據集中以及驗證和測試的人的保護方面，沒有明顯的區別，儘管至少有一個領域的數據被例行收集用於AI開發，即行動健康(mHealth)。如果數據收集本身是研究的一部分，如行動設備的開發，則研究參與者將被認為具有附加的保護，因為他們需要自願且知情地同意參與。
- ▶ 最後，對於根據規定不被認為是人體研究對象的人來說，還存在倫理考慮。如前所述，這些包括群體傷害、個人剖繪和潛在將公共資源轉移開來應對疾病和邊緣化問題的可能性。

是否有現有的**框架或工具**，資助機構、調查人員、HRPP人員和IRB可以使用以闡明並減輕人類導向AI研究和開發的道德問題？

- Keans, M., Roth, A. (2020) *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press

A book that provides a non-technical discussion of computing approaches to building principles and values into AI/ML algorithms themselves.

- Hutson, M. (2022, Feb 26). The Future of Computing. *Science News*, 201(4), 16-22. <https://www.sciencenews.org/century/computer-ai-algorithm-moore-law-ethics> ↗

A lay article that reviews the history of computing technology, the history and fundamental concepts behind AI/ML, and some of the ethical issues raised.

- Wolfram U. *Zero to AI in 60 Minutes*. <https://www.wolfram.com/wolfram-u/machine-learning-zero-to-AI-60-minutes/> ↗ (last viewed 4.8.2022)

A short online course that illustrates how easy it is for anyone with access to the appropriate tools to use AI/ML without deep knowledge of software design or coding.

- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication (2020). https://dash.harvard.edu/bitstream/handle/1/42160420/HLS%20White%20Paper%20Final_v3.pdf?sequence=1&isAllowed=y-PDF ↗


A white paper that presents an international survey of AI governance documents.

- Bernstein, M. S. et al. Ethics and society review: *Ethics reflection as a precondition to research funding*. *Proc Natl Acad Sci U S A* 118, e21117261118 (2021).

A description of one university's approach to addressing potential harms of AI/ML research.

AI特有的考慮是否會影響研究知情同意書中研究活動的充分披露？

- ▶ 需要自願知情同意的研究很可能只是所有AI研究的少數。對於確實需要知情同意的AI研究，這類研究的風險和收益性質不適合於當前所要求的知情同意書要素。特別是，116(b)(2)要求披露「對研究對象的任何可預見的風險或不適」，而116(b)(3)要求披露「對研究對象或他人可能合理期待的任何好處」。風險和收益的這種不對稱考慮反映了IRB批准標準中的情況。損害風險可能對所有人造成，因為預期的好處也將對所有人產生影響，但只有後者被允許在目前的法規中被考慮。目前AI研究的進行受益於這種不一致，因為最重要的損害影響的是群體，而不是個人。受管制的研究是一個公共企業；風險和好處應該平衡私人和公共利益。當前的監管結構只把這部分任務交給了IRB。
- ▶ 在更新的《通用規則》中添加的116(b)(9)處的披露要求也不適合於AI或BD研究，因為它反映了對可識別性的過度簡化的概念。刪除個人身份符號不再意味著個人無法被識別，也不意味著私人和敏感信息不會被披露，並可能在將來與個人相連。這一風險應該明確地披露。



AI研究是否有任何獨特之處會需要**IRB**思考並確定《通用規則》的適用性，而這並不是對於所有人體研究都已經考慮過的。

- ▶ 對於**IRB**來說，在判定《通用規則》是否適用方面幾乎沒有太多的「彈性空間」
- ▶ 更好的問題可能是，當前對於人體研究和人體研究的定義是否允許**IRB**在AI研究下充分保護個人和群體




在具有AI的研究中，45 CFR 46.111的具體部分需要特別關注，例如隱私和保密性、知情同意、風險等方面？

▶ 見問題四、六

對於機構/HRPP責任而言，與IRB所管轄的其他研究責任相比，AI相關的具體考慮有哪些？

- ▶ AI引發了有關群體傷害的問題，主要涉及到數據集的理解不足和使用AI工具可能掩蓋了疾病、邊緣化和不平等的潛在可解決原因。
- ▶ 此外，BD引發了隱私和可識別性方面的問題，在目前的法規中並未得到很好的解決。就機構對其服務社群或所在地社區負責任而言，這些考慮應該由這些機構解決，可能通過它們的HRPP。
- ▶ 然而，許多可以預見的損害超出了任何單一機構的範疇，最好在聯邦層面加以解決。此外，將這一責任交給單個機構的風險在於可能會造成一系列不一致的保護措施，這將不可避免地使那些得到更好保護的人受益，但其代價是其他人受損。



在研究中使用AI是否存在更大的偏見和/或缺陷可能性，IRB應如何考慮這種潛在情況？（例如，臉部識別演算法可能主要基於白人男性，但使用該演算法的研究人員可能不知道這一點）

- ▶ AI的潛在危害來自於數據集中未被認識到的限制或偏見，例如由系統性種族主義和歧視所引起的限制和其他情況，其中數據並不代表將應用其結論的人群。
- ▶ 在大多數AI研究中，初始數據集的組建是與AI研究分開進行的，這使得調查人員更有可能不知道他們結論的普遍適用性有限這一點。

建議—可識別性與隱私

- ▶ AI/ML和BD研究暴露了傳統可識別性概念的局限性，該概念是《通用規則》下隱私保護的基礎。從特定數據集中明確識別個人的能力是一種特徵，在數據被孤立分析時是適當的，當數據收集主要發生在明確定義的研究環境中時（即在廣泛使用電子健康紀錄和在醫療保健以外範圍進行普遍數據收集之前），以及在常規收集和使用基因組數據之前，這些數據可說是本質上具有識別性的。SACHRP敦促應跟進《通用規則》對定期重新審視可識別性含義的承諾，以應對不斷發展的技術和研究實踐。
- ▶ 大型數據集的組合使得可能獲取或推斷出個人未知的信息。從某種意義上說，這是AI/ML的目標之一，因為它利用數據中的模式來推斷這些個人的新奇或未披露的信息。如果研究基本上可以重新創建有關人們的私人和敏感信息，即使他們的身份不是明確的，個人是否會認為這違反了他們的隱私權？換句話說，BD和AI/ML是否允許研究人員創建“虛擬研究對象”，可以在此類研究中進行，而無需遵守規定的負擔，但與識別性數據上的研究沒有實質性差異？SACHRP建議考慮可識別性是否仍然是一個讓研究參與者和一般公眾認為在設定聯邦保護限制方面具有用途的概念。

建議—人體研究對象的定義

- ▶ 對於AI/ML最相關的《通用規則》中關於人體研究對象的定義是：
“一名研究者進行研究的存活個體：... (ii) 獲取、使用、研究、分析或生成可識別的私人信息或可識別的生物樣本。” 2018年對規則的更新將生成可識別的私人信息的可能性添加到了現有的定義中，恰當地認識到了數據集和基因組信息很少單獨使用，並且數據集的組合即使沒有單個數據集本身具有識別性也可以識別個人。
- ▶ 儘管如此，規則仍然依賴“公開”概念來排除那些公開披露信息的個人免受保護。這個擔憂並不新鮮；公共行為和私人行為之間的界限一直不明確，公共行為或言論是否針對公眾觀眾的問題也一直存在矛盾。互聯網和社交媒體使這種擔憂對更廣泛的人群變得更加嚴重。社交媒體邀請個人分享關於自己的信息，承諾其用戶他們將建立社區，但商業上的目的是收集數據並分析群體的行為。同樣，信用卡和簽帳金融卡作為現金的替代品為用戶提供了財務管理的便利和靈活性，但現在它們還具有數據收集和分析購買模式的附加目的。

建議一人體研究對象的定義(續)

- ▶ 事實上，現代社會以在任何可能的機會上收集個人數據為特徵。這種數據收集是否得到適當披露，個人是否真的可以選擇不允許這種數據收集而不會在社交和財務上受到嚴重損害，以及這種數據收集是否具有剝削性，這是一個比聯邦保護研究參與者更大的問題，但卻悄然存在於《通用規則》下AI/ML和BD考慮的背景中。該規則使我們可以避免考慮這些更深層次的問題，特別是聯邦資助的研究是否應該在這些領域比商業活動更高的標準下遵循，因為它將這些信息的大部分視為“公開”。
- ▶ SACHRP建議考慮更加細緻但明確的公開行為與私人行為以及私人信息的定義，以認識到自這些概念首次被立法確立以來，技術帶來的深刻變化。

建議 – 在設定新標準時加入包容性

- ▶ 最初的研究規定主要是為了應對生物醫學研究中發生的傷害而制定的，它們的要求不成比例地保護了所有社會成員都會認可的身體傷害。同樣，普遍認為，改善健康和減輕疾病負擔是一項值得追求的公共利益和聯邦政府的角色。
- ▶ 雖然AI/ML和BD可以應用於生物醫學和醫療保健研究領域，但它們所帶來的許多風險和所承諾的好處超出了這些範疇。從風險的角度來看，它們的許多潛在危害主要影響到了群體。依賴只能反映當前或過去做法的數據，它們的應用可能會使不恰當的群體差異和偏見被鞏固或虛假確認，而這些都是必然會被這些數據所捕捉到的。以保護這些群體的個人成員來解決此類危害，這在AI技術使推理變得不透明且“正當程序”困難時是不夠的。從利益的角度來看，AI研究的許多目標可能對所有社會成員的價值並不明顯且平等。這些不同的評估可能是由於一個歷史上的群體被排除在研究成果之外，對AI/ML進一步社會邊緣化的擔憂，或者不同的文化規範所導致的。

建議 – 在設定新標準時加入包容性 (續)

- ▶ 與美國原住民部落社區的研究經驗突顯了這一問題。這些社區被認為是主權獨立的，因此他們對自己文化價值觀的權利在法律和規定中得到了確立，明確允許他們將《通用規則》適應到他們自己的社區。亞利桑那州立大學使用來自哈瓦蘇派部落成員的生物樣本進行的基因組研究表明，對於可能被描述為生物醫學的研究，對群體危害和群體規範的擔憂是相關的；當目標是了解群體特徵時，評估研究價值的文化差異很可能存在。儘管美國原住民的主權賦予了該多樣化群體的成員一些獨特的司法和法律保護，但還有許多其他群體的成員也同樣強烈地感受到了共同社區價值觀，但他們卻沒有得到這樣的認可或保護。
- ▶ 如何在制定或解釋研究規定時納入相關聲音是一個困難的問題，不太可能有一個能滿足所有人的解決方案，這是定義個人、群體和政府之間關係的許多問題的特徵，在多元化民主體制中都如此。然而，這種困難不應成為不明確考慮問題並尋求公平解決群體關切的借口，特別是在研究獲得公共資助時。

建議 – 在設定新標準時加入包容性 (續)

- ▶ **SACHRP**建議考慮建立促進對話和最終規定指導的平台和機制，關於如何考慮和保護AI研究可預見地影響到的群體利益，並且保持科學完整性一致。此外，**SACHRP**建議，基於這些對話機會，應確立正式指導，以確保當**HHS**考慮資助使用AI的研究計畫或改進AI方法和演算法的研究時，考慮到受影響群體的預期利益以及危害，尤其是在生物醫學範疇之外的研究中，這些群體的利益和危害可能會被預見到。



結論

結論

- ▶ 用AI – 自己還是要有基礎能力，才能有效運用AI
- ▶ 發展AI – 人工智慧領域的希波克拉底誓言
- ▶ AI的可解釋性、可理解性、透明性至關重要
- ▶ 負責任的AI
- ▶ IRB/REC如何因應AI時代



Thank you for Listening



ChatGPT協助部分翻譯工作

Reference

- ▶ Madiega, T. A. (2019). *EU guidelines on ethics in artificial intelligence: Context and implementation*. Think Tank, European Parliament.
[https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EP_RS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EP_RS_BRI(2019)640163_EN.pdf).
- ▶ Müller, V. C. (2023). *Ethics of artificial intelligence and robotics*. *The Stanford Encyclopedia of Philosophy* (Fall 2023 Edition). In Edward N. Zalta & Uri Nodelman (eds.).
<https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=ethics-ai>.
- ▶ Rotaru, T. Ş., & Amariei, C. (2023). Ethical Issues in Research with Artificial Intelligence Systems. In M. Radenkovic (Ed.), *Ethics - Scientific Research, Ethical Issues, Artificial Intelligence and Education*.
<https://www.intechopen.com/chapters/1127065>.
- ▶ United Nations System (2022). Principles for the ethical use of artificial intelligence in the United Nations System.
https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.