臺北榮民總醫院 臨床研究受試者保護中心
Human Research Protection Center

人工智慧風險評估架構：由歐盟人工智慧法案談起

吳俊穎 講座教授

國立陽明交通大學 生物醫學資訊研究所 所長
國立陽明交通大學 健康創新中心 主任
國立陽明交通大學 微菌叢研究中心 主任
國家人體微菌核心設施 主任
台灣微菌聯盟 理事長

大綱

1. 統計分析 vs. 人工智慧
2. 人工智慧的可能偏差
3. 歐盟人工智慧法案的風險分層
4. 智慧醫療的法律責任

Pros of traditional statistics (1)

- Interpretability:
  - Results are generally easy to understand and explain
- Theoretical grounding:
  - Based on well-established mathematical theory
- Hypothesis-driven:
  - Encourages structured thinking through hypothesis testing
- Clear assumptions:
  - Models require known & testable assumptions (e.g., normality, independence)
- Reproducibility:
  - Given the same data and method, results are easily reproducible

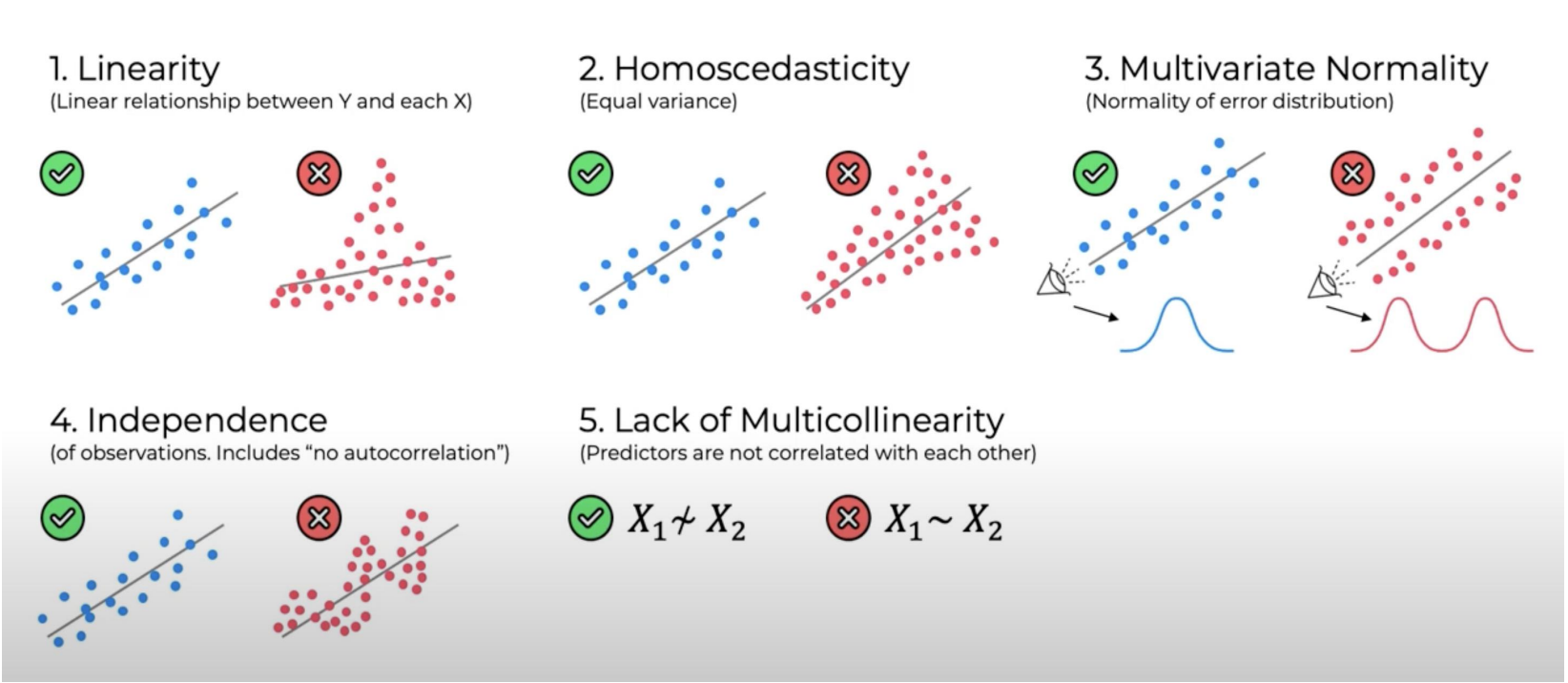
Pros of traditional statistics (2)

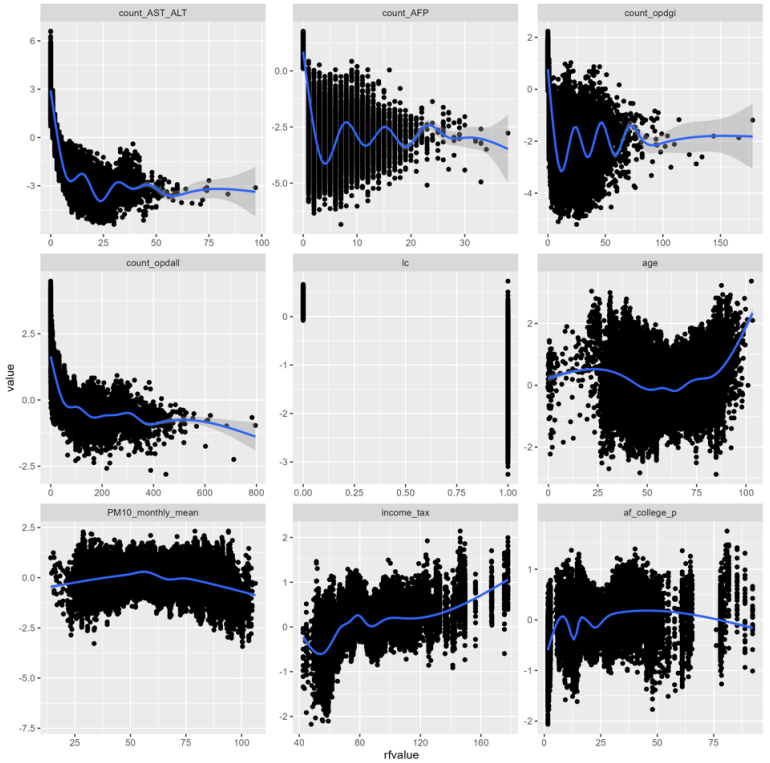
- Transparency:
  - Model parameters have direct and intuitive meanings
- Low computation cost:
  - Suitable for small datasets and requires less processing power
- Strong inference capabilities:
  - Suitable for establishing causality in controlled settings
- Wide acceptance:
  - Standard in many fields like medicine, law, and social sciences
- Statistical significance testing:
  - Provides p-values and confidence intervals for decisions

Cons of traditional statistics (1)

- Rigid assumptions:
  - Violations can invalidate results
- Limited with complex data:
  - Not ideal for high-dimensional or unstructured data
- Manual feature engineering required:
  - Needs explicit variable selection and transformation
- Not adaptive:
  - Struggles to learn from new data automatically
- Lower predictive power:
  - Less accurate for certain types of pattern recognition tasks

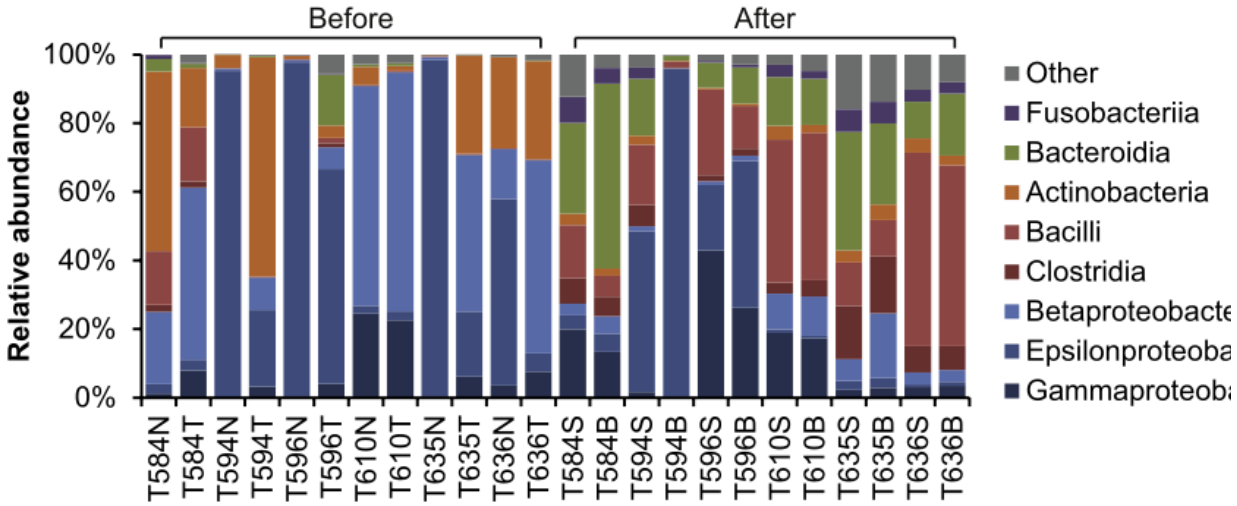
Limitation of statistics: Non-linear relationships



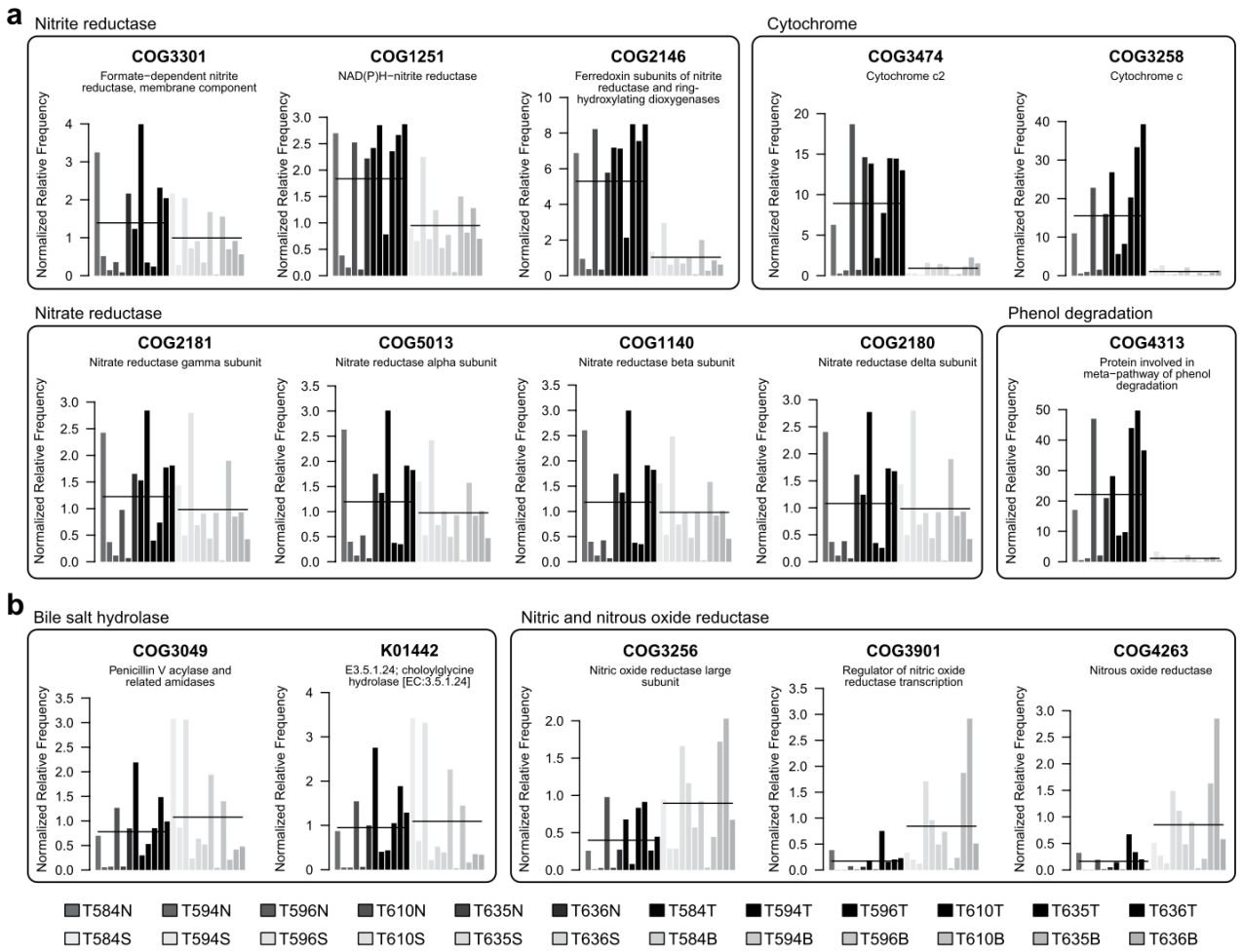


Assumption of Linear Regression

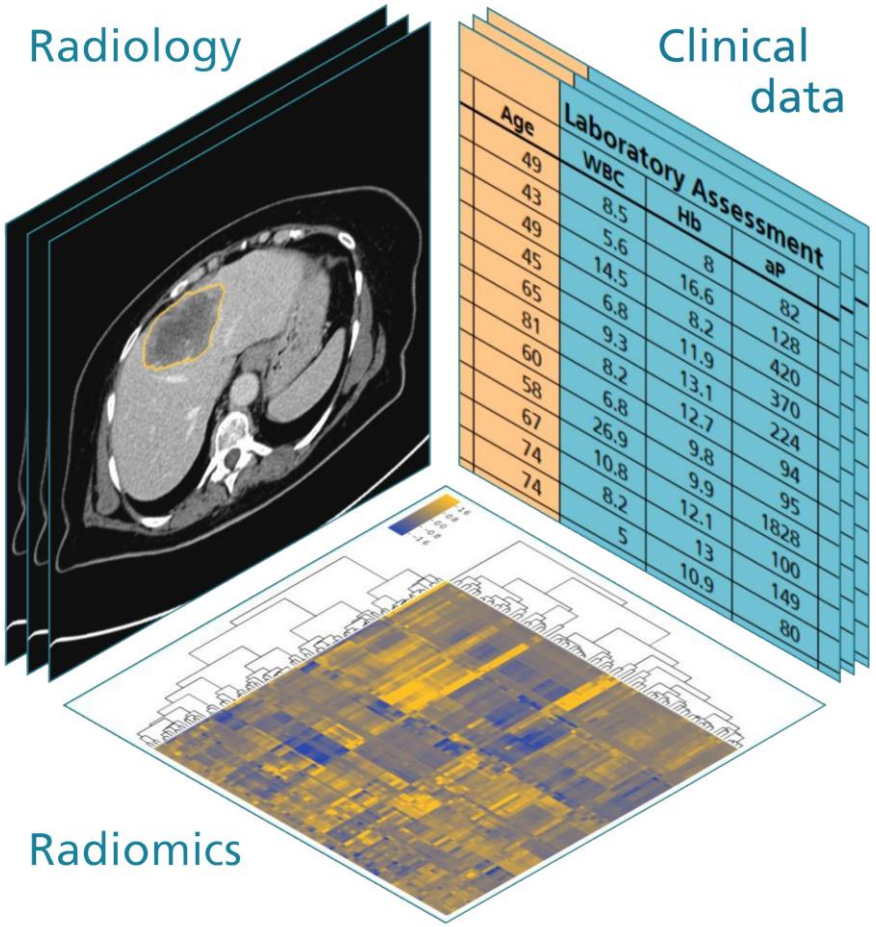
Limitation of statistics: high-dimensional data

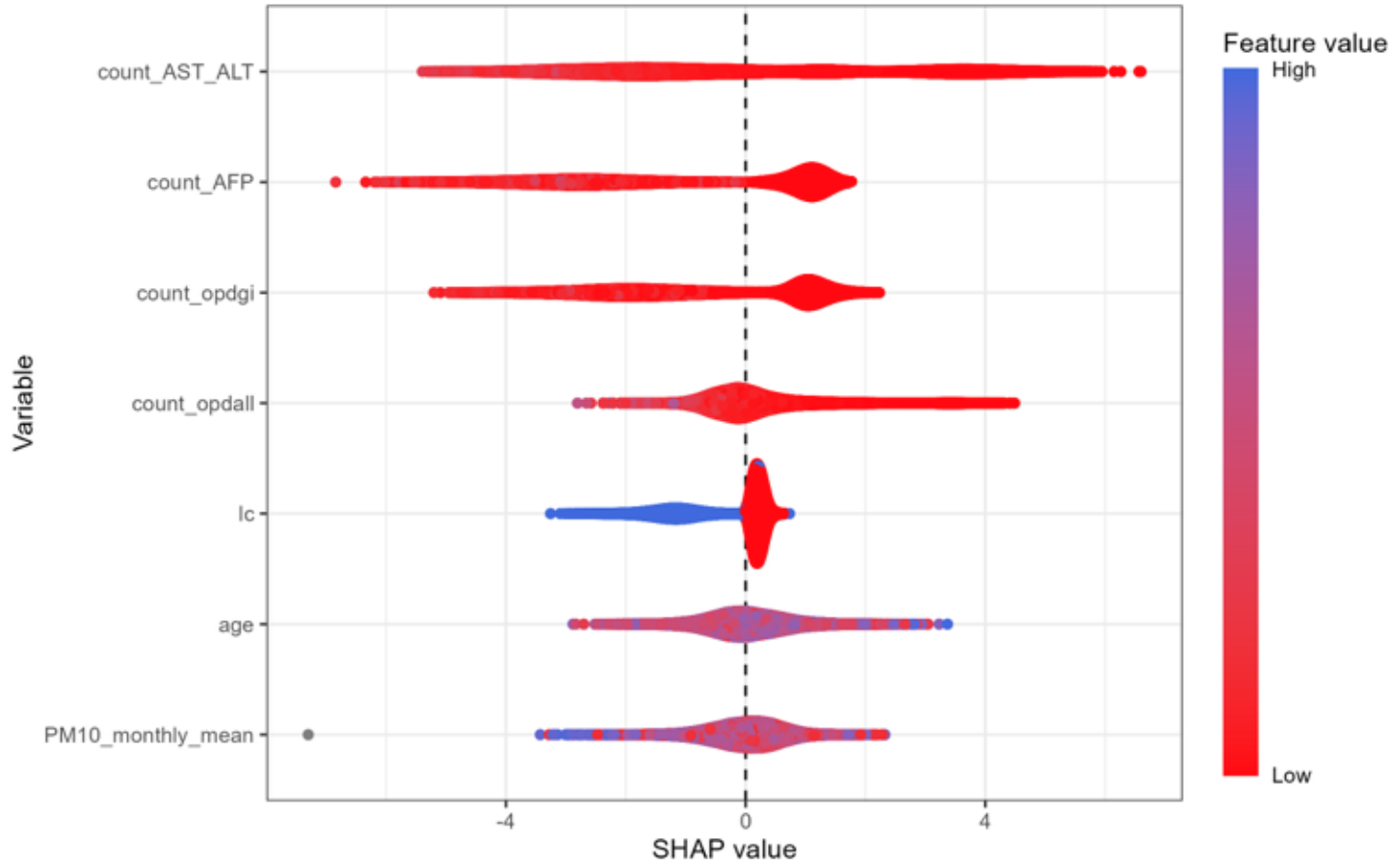


Microbiota data



Limitation of statistics: Complex datasets





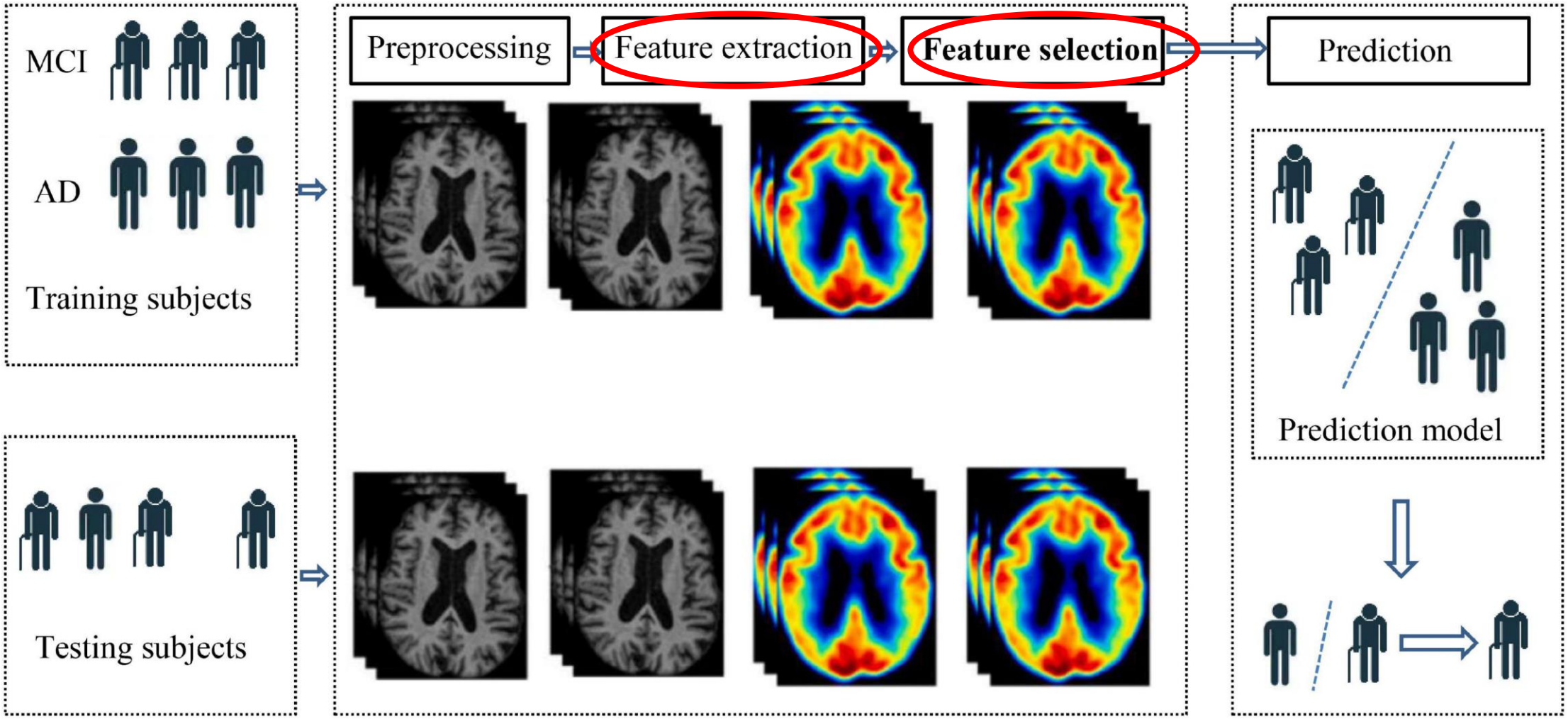
Cons of traditional statistics (2)

- Data must be clean and structured:
  - Poor handling of missing or noisy data
- Can overlook nonlinear relationships:
  - Many models assume linearity
- Model selection is manual and complex:
  - Choosing the right model is skill-intensive
- Less automation:
  - Cannot scale well without expert intervention
- Limited interaction modeling:
  - Hard to capture complex feature interactions

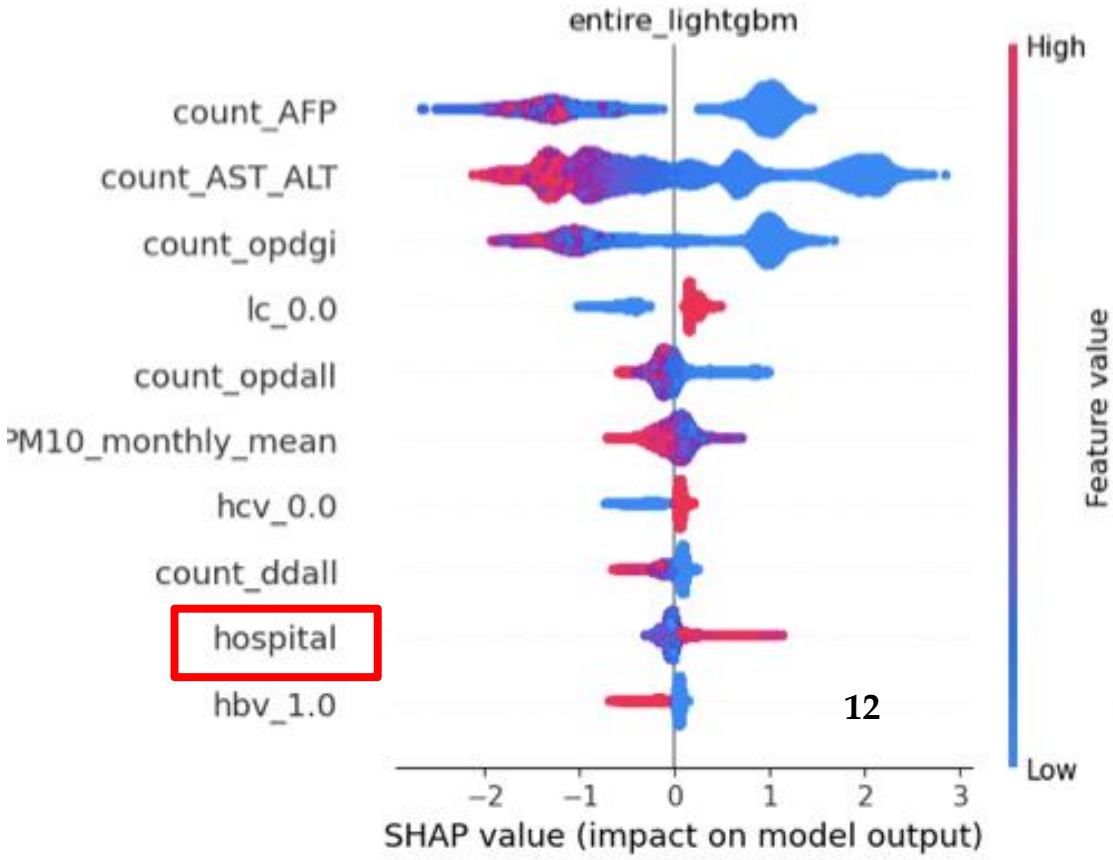
Pros of AI/ML analysis (1)

- High predictive accuracy:
  - Excels in pattern recognition and classification
- Handles large, complex datasets:
  - Suitable for big data and unstructured formats
- Learns from data:
  - Models can improve with more data
- Captures nonlinear relationships:
  - Neural networks and tree models model complex interactions
- Automated feature extraction:
  - Especially in deep learning

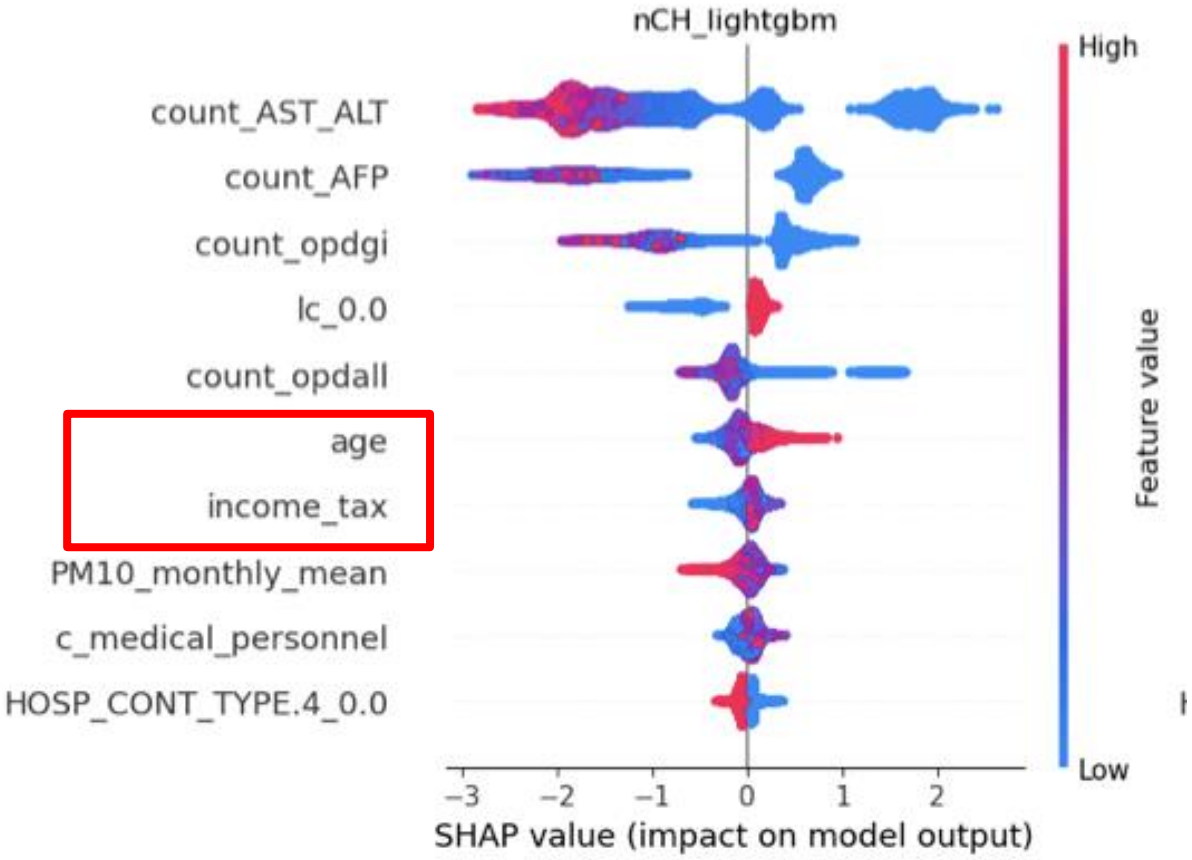
Automation of feature extraction and selection



Automation of feature extraction and selection



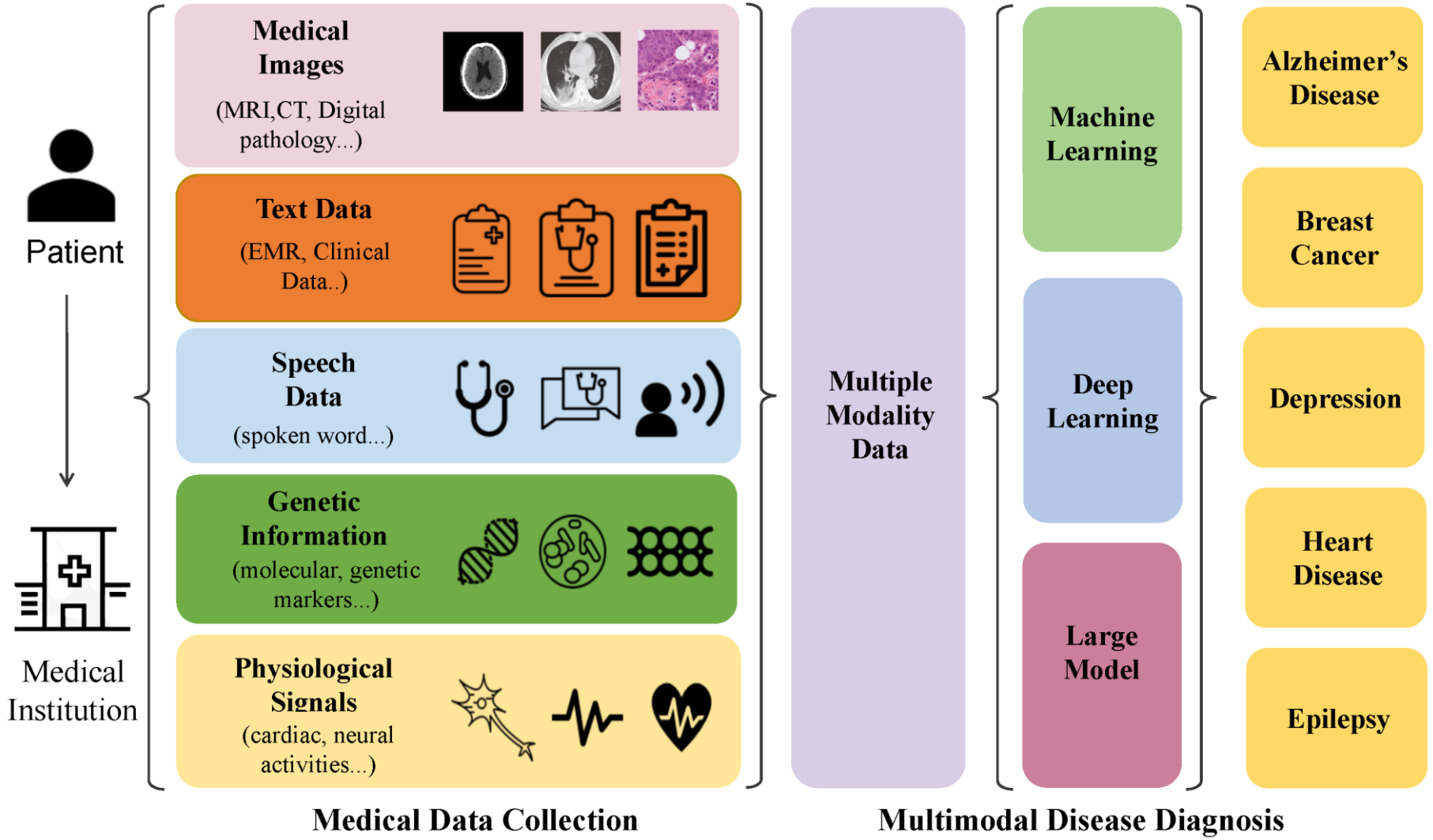
12



Pros of AI/ML analysis (2)

- Scalable:
  - Can be deployed at scale in real-time systems
- Flexible algorithms:
  - Wide range of models for different problems
- No need for strict assumptions:
  - Works even when traditional assumptions are violated
- Customizable learning:
  - Prioritize outcomes like precision or recall
- Useful in diverse domains:
  - Applied in vision, language, genomics, finance, etc.

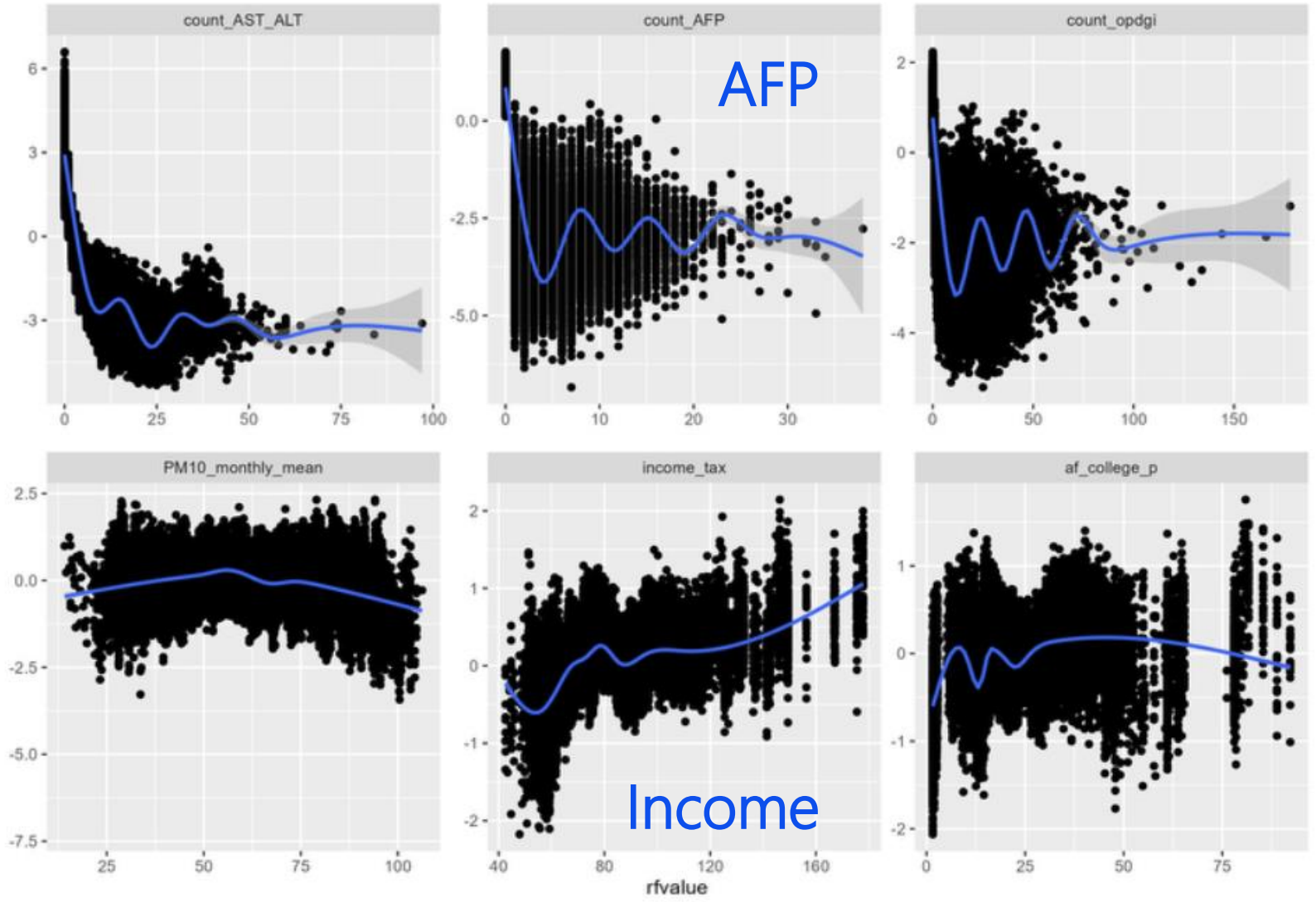
Flexibility and adaptability to complex datasets



Ability to capture non-linear relationships

GOT/GPT

PM10



AFP

GI OPD

Income

Education

Cons of AI/ML analysis (1)

- Low interpretability:
  - Often black-box in nature
- Data-hungry:
  - Requires large, high-quality datasets
- Overfitting risk:
  - Can memorize noise in training data
- Computationally intensive:
  - Needs powerful hardware and time
- Difficult to validate causality:
  - Often correlation-based

Cons of AI/ML analysis (2)

- Harder to debug:
  - Errors not always obvious
- Bias amplification:
  - Can reproduce systemic biases
- Complex tuning required:
  - Needs model selection and hyperparameter tuning
- Opaque decision-making:
  - Not ideal for accountability-heavy fields
- Model drift:
  - Performance may degrade if data patterns change

大綱

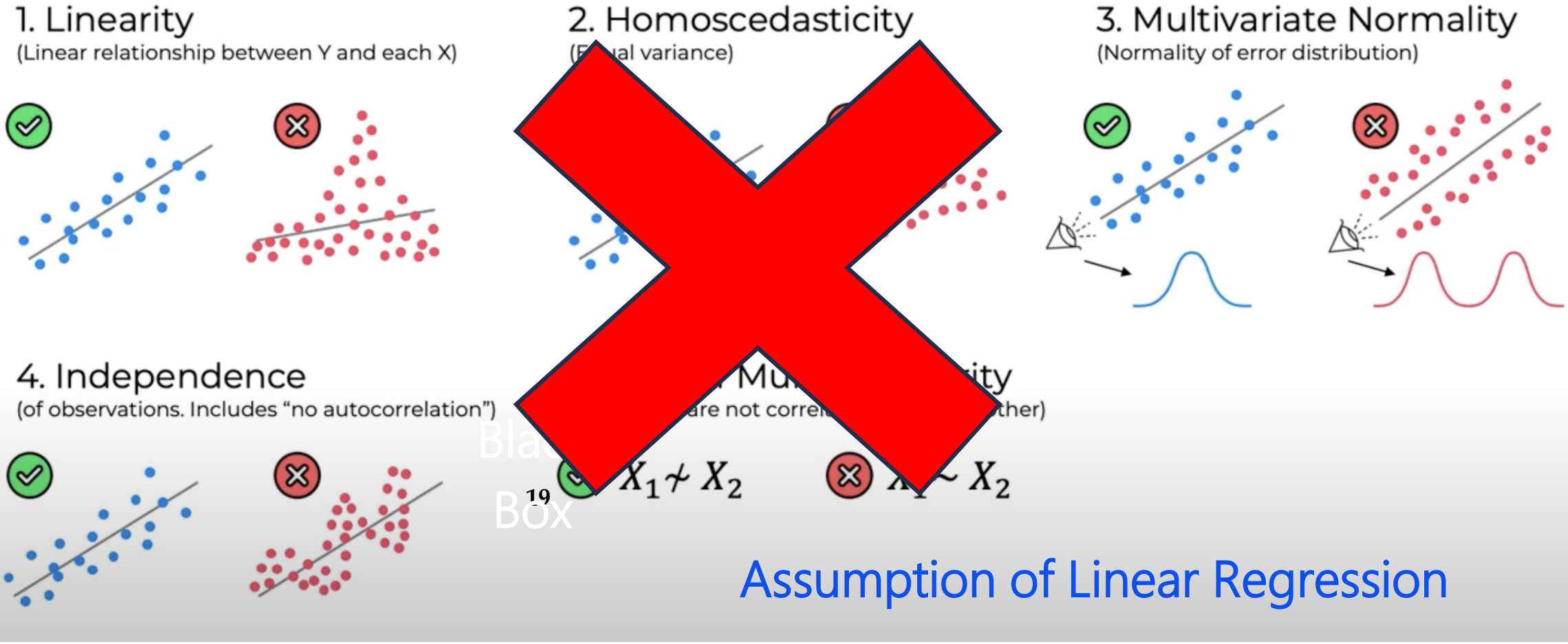
1 統計分析 vs. 人工智慧

2 人工智慧的可能偏差

3 歐盟人工智慧法案的風險分層

4 智慧醫療的法律責任

Cons of AI: No assumptions

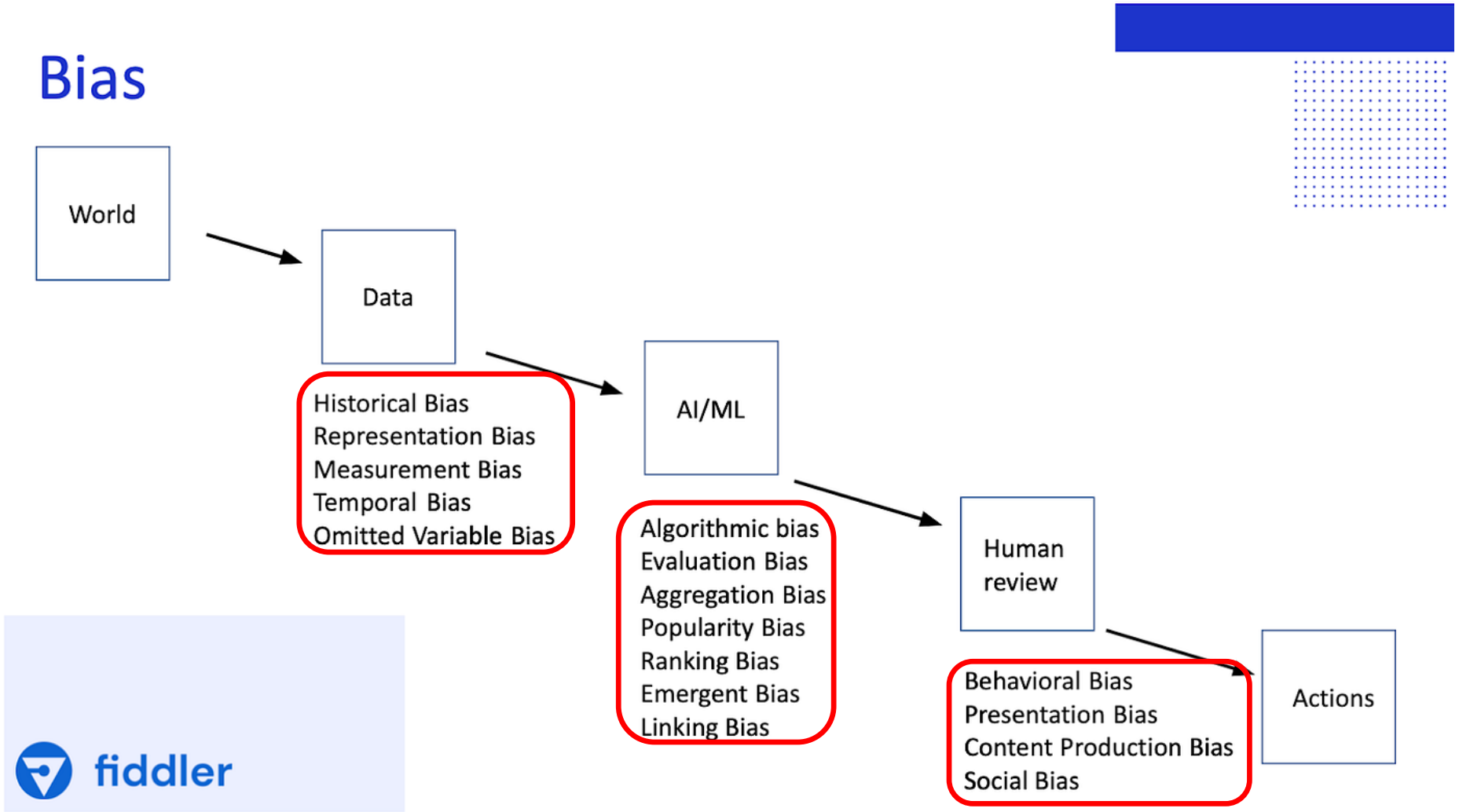


Bla
Box

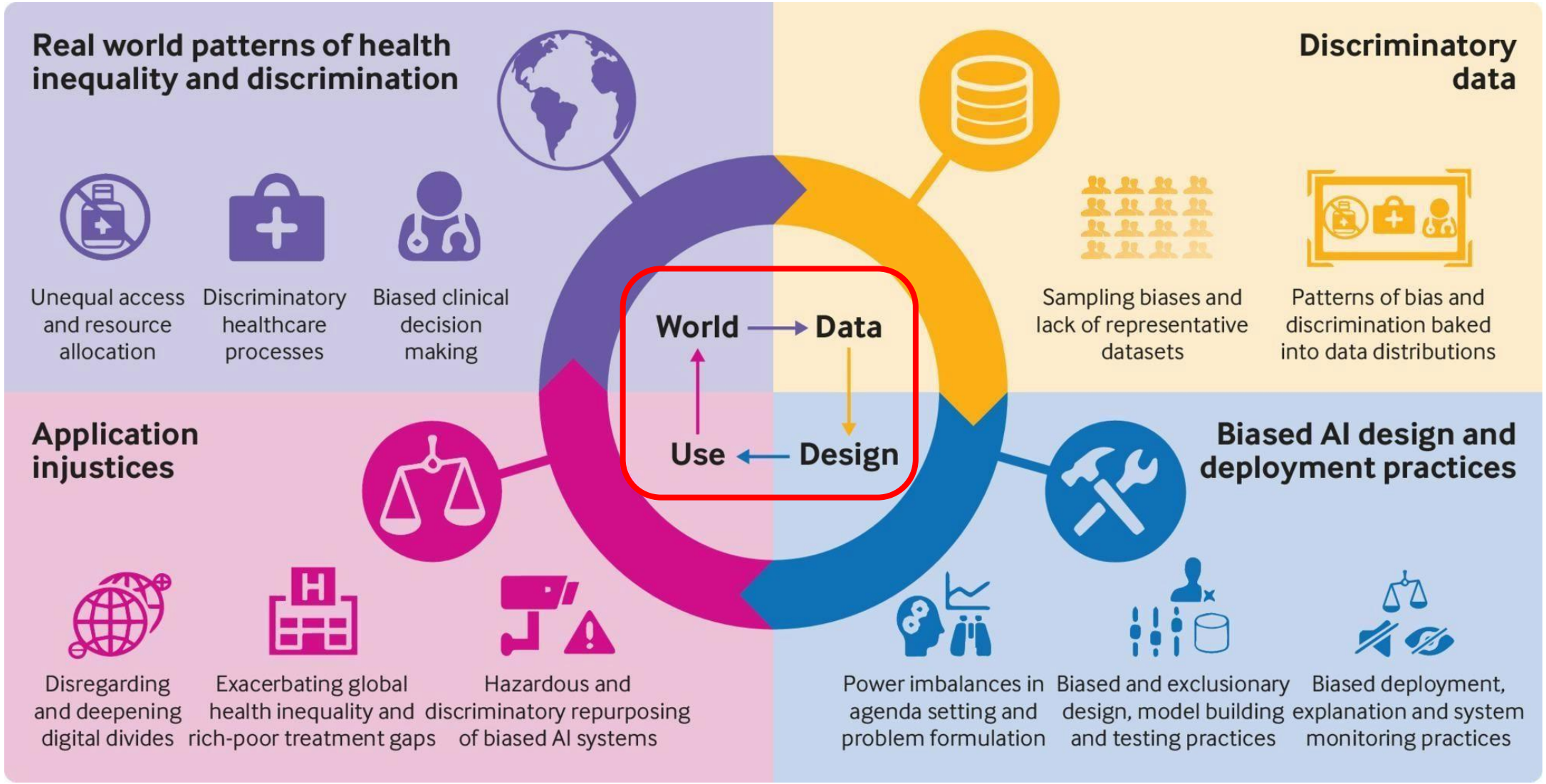
19

Assumption of Linear Regression

Cons of AI: Prone to bias and error



Sources of AI Bias



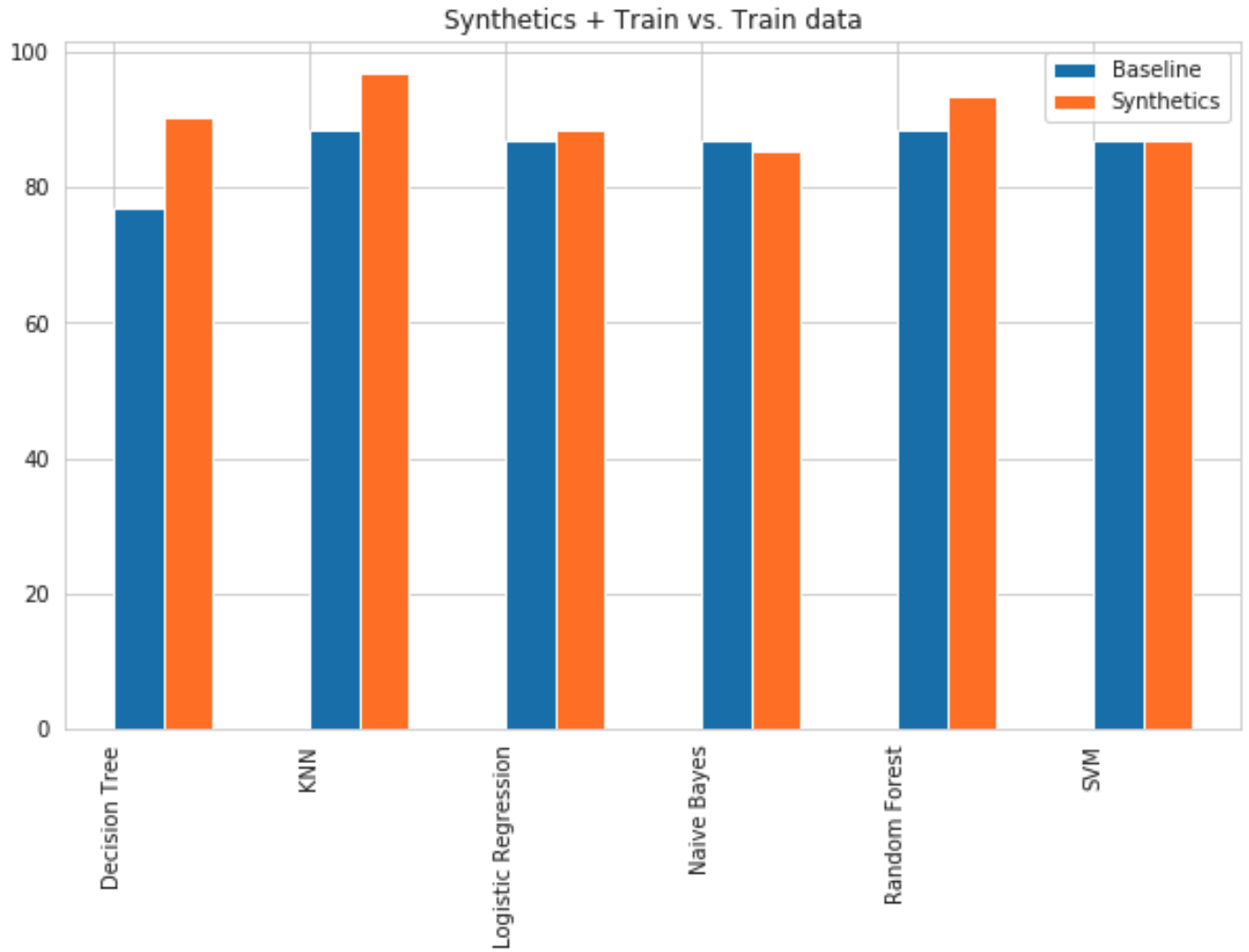
Imbalanced data

heart.csv (11.06 KB) — 14 of 14 columns — Views

	# age age in years	# sex (1 = male; 0 = female)	# cp chest pain type	# trestbps resting blood pressure (in mm Hg on admission to the hospital)	# chol serum cholestoral in mg/dl	# fbs (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)	# restecg resting electrocardiog results
	29 – 77	0 – 1	0 – 3	94 – 200	126 – 564	0 – 1	0
1	63	1	3	145	233	1	
2	37	1	2	130	250	0	
3	41	0	1	130	204	0	
4	56	1	1	120	236	0	
5	57	0	0	120	354	0	
6	57	1	0	140	192	0	
7	56	0	1	140	294	0	

32% Female

Imbalanced data vs. synthetic data



AI amplifies Bias

COOKING

ROLE		VALUE
AGENT	▶	WOMAN
FOOD	▶	PASTA
HEAT	▶	STOVE
TOOL	▶	SPATULA
PLACE	▶	KITCHEN

COOKING

ROLE		VALUE
AGENT	▶	WOMAN
FOOD	▶	FRUIT
HEAT	▶	—
TOOL	▶	KNIFE
PLACE	▶	KITCHEN

COOKING

ROLE		VALUE
AGENT	▶	WOMAN
FOOD	▶	MEAT
HEAT	▶	GRILL
TOOL	▶	TONGS
PLACE	▶	OUTSIDE

COOKING

ROLE		VALUE
AGENT	▶	WOMAN
FOOD	▶	VEGETABLES
HEAT	▶	STOVE
TOOL	▶	TONGS
PLACE	▶	KITCHEN

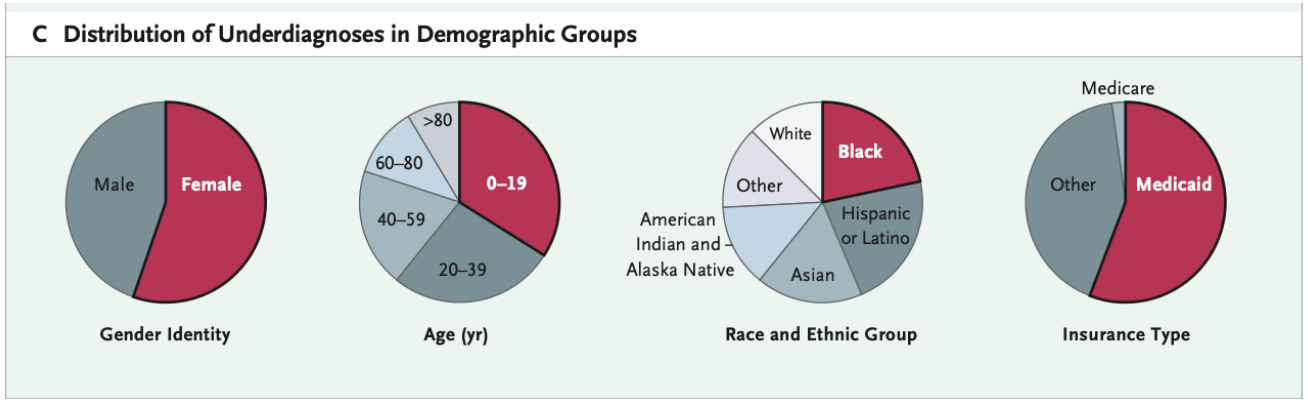
COOKING

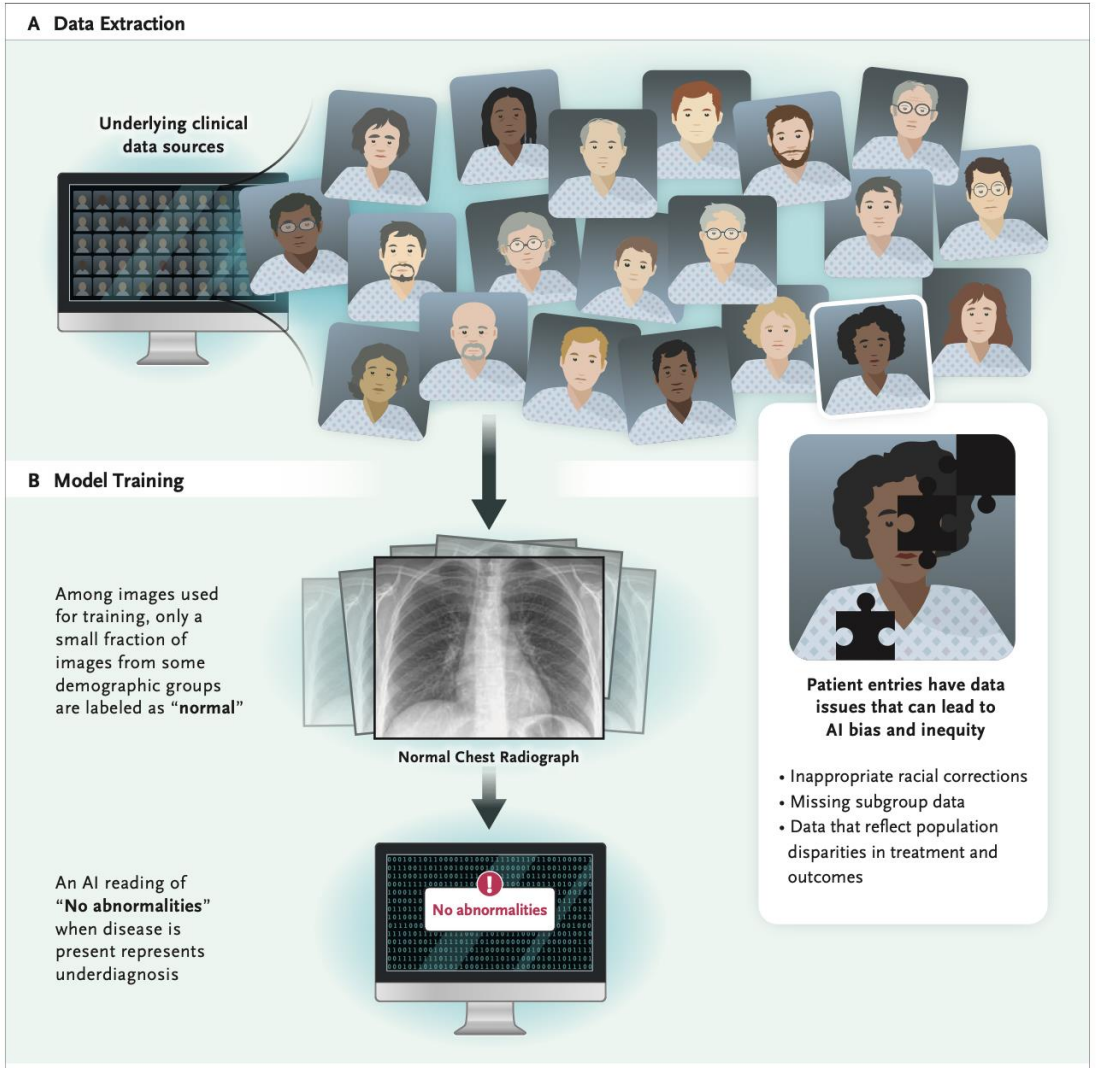
ROLE		VALUE
AGENT	▶	MAN
FOOD	▶	—
HEAT	▶	STOVE
TOOL	▶	SPATULA
PLACE	▶	KITCHEN

Biased data leads to AI artifacts

The NEW ENGLAND JOURNAL of MEDICINE

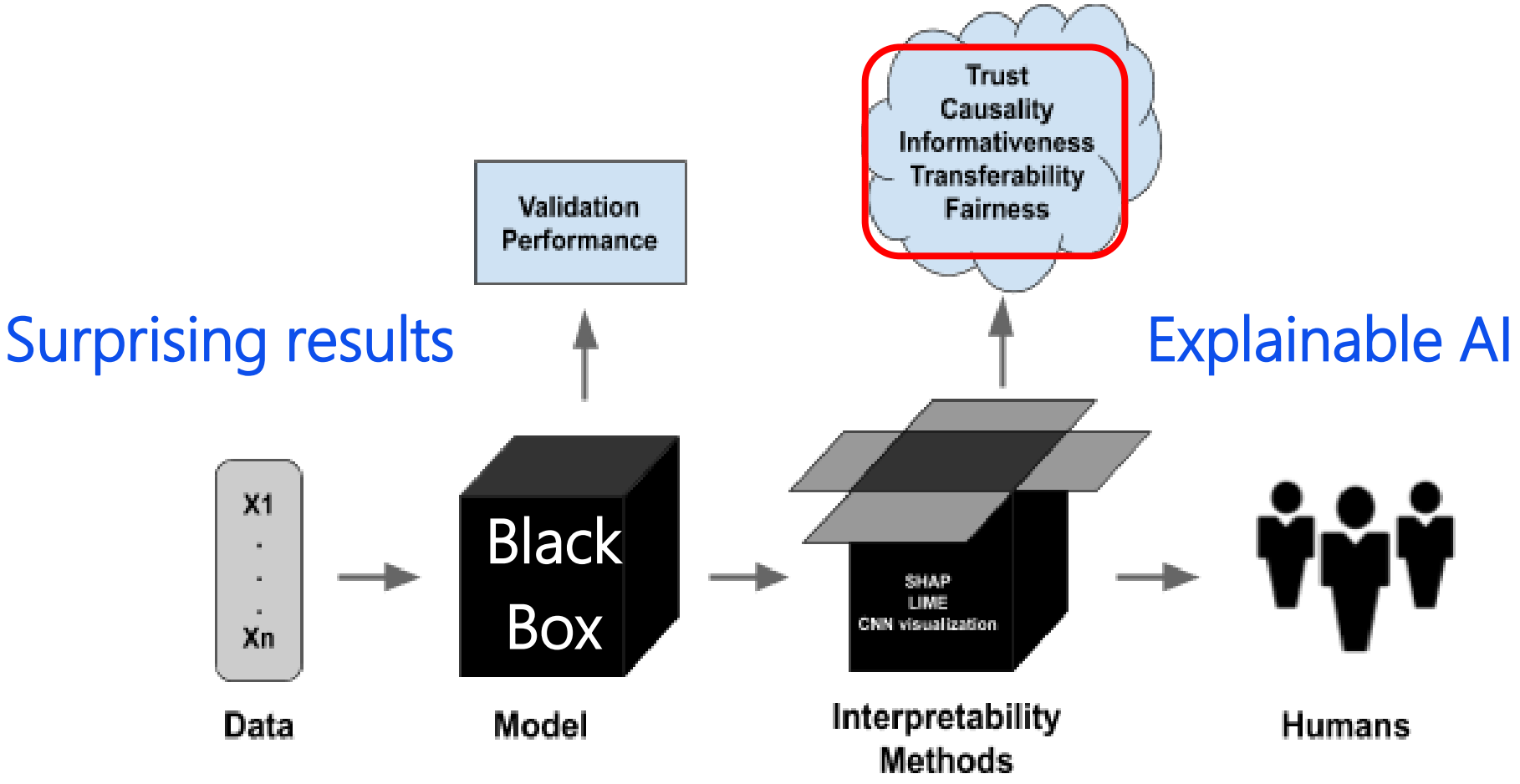
Considering Biased Data as Informative Artifacts in AI-Assisted Health Care





Ferryman K, et al. NEJM, 2023

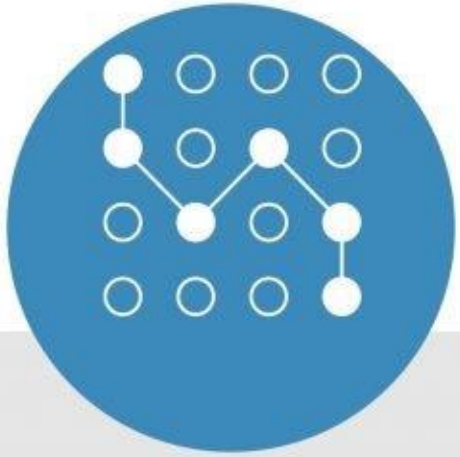
Cons of AI: Difficult to explain



Cons of AI: Difficult to explain

Explainable Data

What data was used to train the model and why?



Explainable Predictions

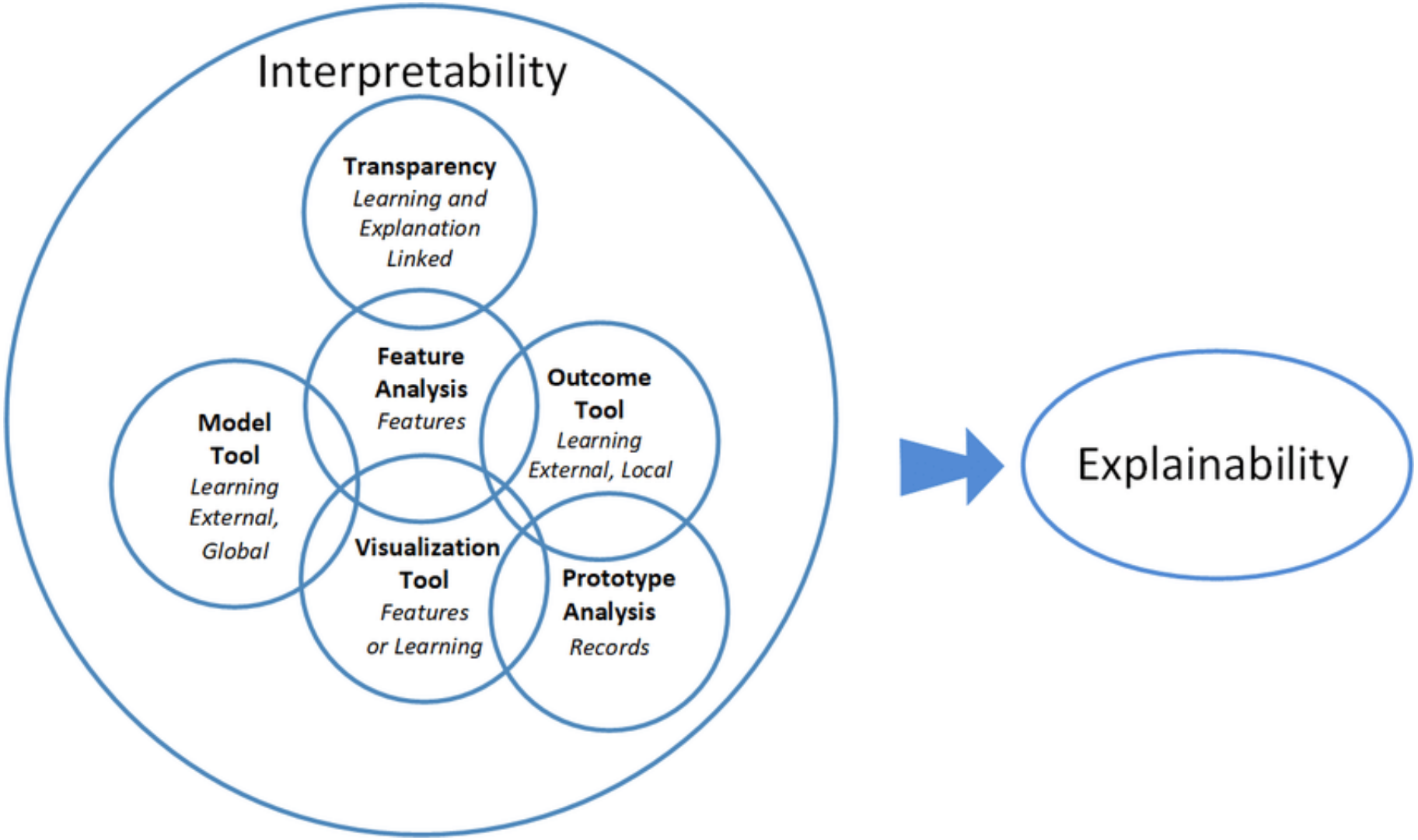
What features and weights were used for this particular prediction?

Explainable Algorithms

What are the individual layers and the thresholds used for a prediction?

Explainable AI

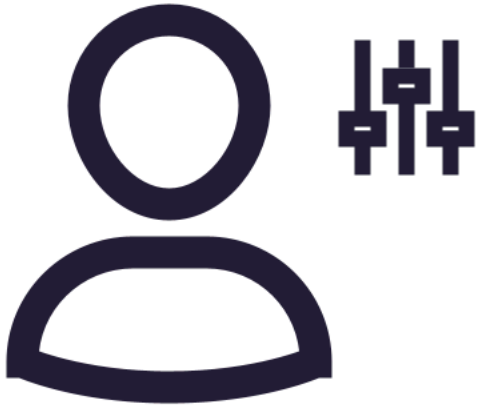
Interpretability vs. explainability



Interpretability vs. explainability

Interpretability

Understand the relationship between variables and the predicted response at a **global level**.



Explainability

Understand which variables were most relevant for the prediction for a **given record**.

大綱

1 統計分析 vs. 人工智慧

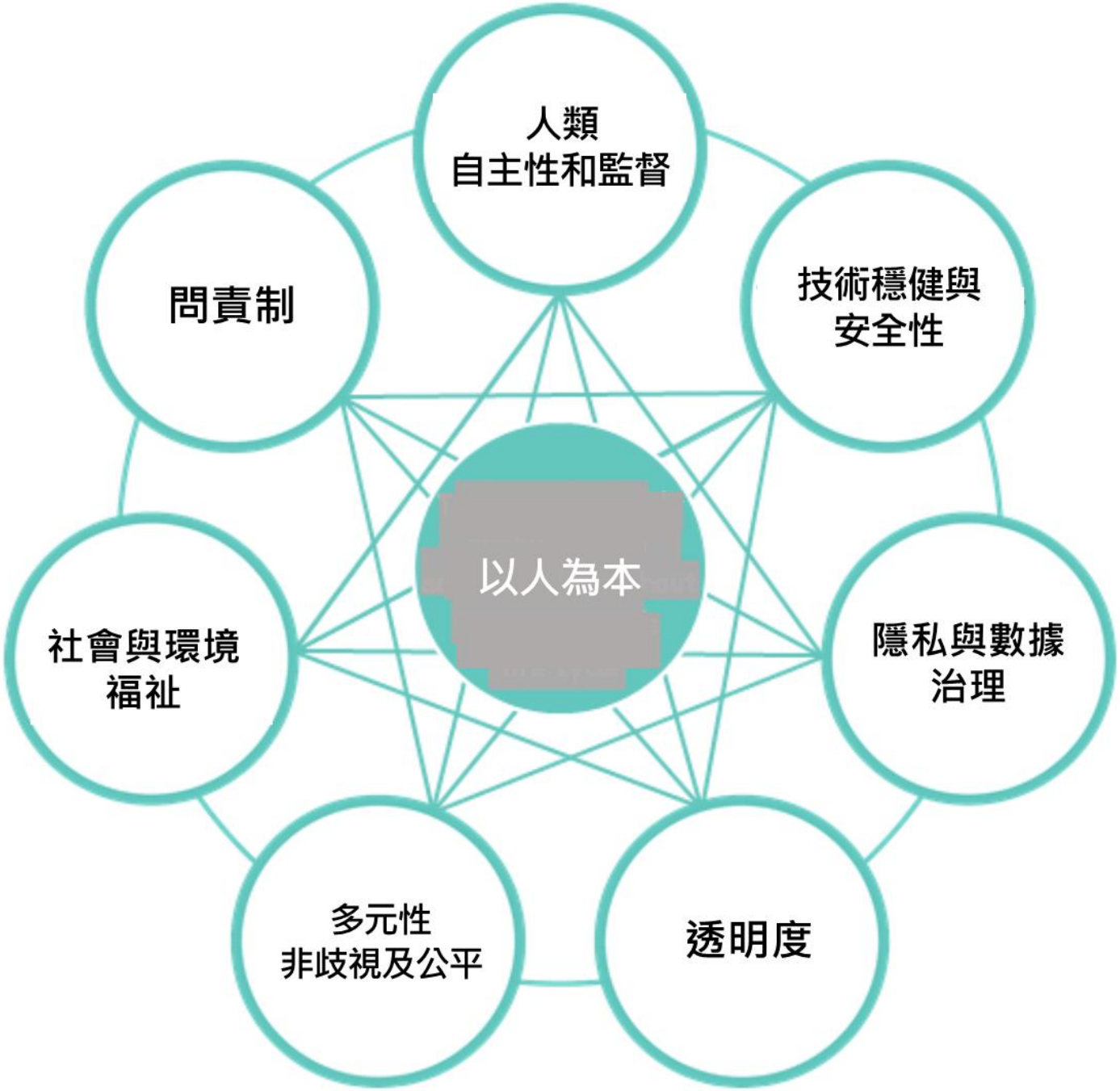
2 人工智慧的可能偏差

3 **歐盟人工智慧法案的風險分層**

4 智慧醫療的法律責任

可信賴的人工智慧倫理準則

Ethics Guidelines for Trustworthy AI



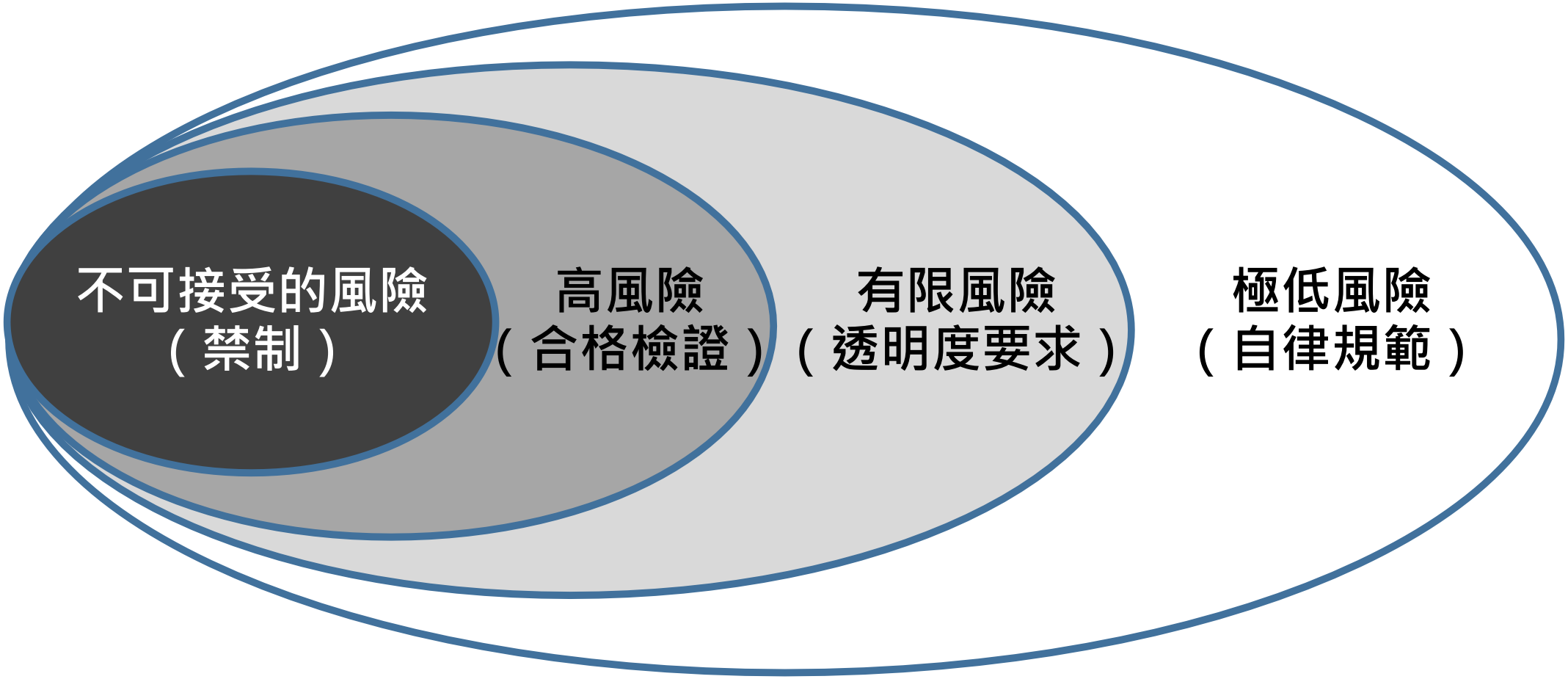
歐盟人工智慧法

EU Artificial Intelligence Act

—於2024年5月21日正式通過

項目	可信賴的AI倫理準則	歐盟AI白皮書	歐盟人工智慧法
制定時間	2019年4月發表（草案2018年12月）	2020年2月發表	2024年3月歐洲議會通過，5月歐盟理事會通過
制定單位	歐盟高階專家小組（AI HLEG） 由歐盟執委會設立的獨立團體	歐盟執委會（European Commission）	歐洲議會與歐盟理事會共同立法
法律定位	非約束性倫理準則（Soft law）	政策文件、戰略建議（非正式法律文件）	強制性歐盟法規（Regulation） 具法律拘束力
主要立法意旨	建立可信賴AI三大核心： 合法、倫理、健全	發展歐洲版AI策略： 促進創新與信任並重	建立風險導向AI分類系統與監管架構， 保護基本權利並促進AI發展
主要法條內容	四大倫理原則： ① 人類自主 ② 預防傷害 ③ 公平 ④ 可解釋性 七項要求：監督、健全性、隱私、透明、公平、多元、問責 信賴AI評估清單	強調以人為本、符合歐洲價值的AI治理 高風險AI風險導向監管 與資料戰略、倫理準則整合	四類風險分類（不可接受、高風險、有限、極低） 高風險AI合規檢查 禁止社會評分與操控行為 設立AI辦公室與AI委員會

AI風險評估架構



不可接受風險的AI

定義：

- 不可接受風險的AI系統，是指其使用方式嚴重侵犯基本人權或對社會構成廣泛危害，因此全面禁止。
- 此類系統通常涉及操控、歧視、監控或其他對人類尊嚴與自主性構成實質威脅的用途，無論其潛在效益如何，都被視為無法容忍。

不可接受風險的AI

範例：

- 基於社會行為或個人特質進行「社會評分」（social scoring）系統
- 基於種族或性別等特徵進行歧視性篩選的 AI 系統
- 用於公共空間中即時遠端生物辨識監控的 AI 系統（例如臉部辨識），除非符合極端例外
- 嘗試解讀個人情緒或意圖以操控其行為的情感推論系統（尤其用於工作或教育環境中）
- 基於未經合理懷疑的個人特質進行犯罪預測的 AI 風險分析工具

高風險AI

定義：

- 高風險AI系統指的是其使用對個人生命、健康、安全或基本權利具有顯著影響力。
- 此類系統仍允許使用，但必須滿足嚴格的合規性、透明性、風險管理與人類監督要求。主要涵蓋與人身安全、基本服務、公權力、司法等密切相關的領域。

高風險AI

範例：

- 醫療診斷輔助系統，例如放射影像 AI 分析工具。
- 金融信貸評分系統，用於個人信用核定。
- 雇主用來篩選履歷、評估績效的 AI 工具。
- 移民與庇護審查中的 AI 系統，例如危險評估或文件審查工具。
- 支援司法判決或爭議解決程序中的法律事實分析工具。

高風險

適用醫療器材安全專法：

- 歐盟醫療器材法
  (MDR, Regulation EU 2017/745)
- 歐盟體外診斷醫療器材法
  (IVDR, Regulation EU 2017/746)

兩者類同美國FDA對於「醫療器材軟體」(SaMD)採行分級監管

有限風險AI

定義：

- 有限風險 AI 系統指其可能引發的風險較低，主要需滿足特定透明度義務。
- 雖不要求高風險級別的強化規範，但需告知用戶自己正在與 AI 互動，並提供足夠的資訊以理解其功能。

有限風險AI

範例：

- 虛擬客服機器人（如語音助理）。
- 線上廣告推薦系統，使用者可辨識為 AI 推薦。
- AI 生成內容工具，如簡易的文風修改工具。
- 自動翻譯工具，用於非正式語境。
- 支援社群媒體平台內容分類的 AI 工具。

大型生成式AI模型

透明性

- AI系統的開發和使用應以允許適當的可追溯性和可理解性為前提，讓人們知道他們正在與AI系統進行交流或互動；
- 同時充分告知部署者該AI系統的能力和限制，以及告知受影響人員他們的權利。

通用人工智慧（GPAI）

一般要求（第53條）

- 制定**技術文件**，包括訓練和測試過程以及評估結果。
- 起草**使用說明**，提供給打算將 GPAI 模型整合到自己的人工智慧系統中的下游供應商，以便後者明瞭功能並能夠遵守相關限制。
- 尊重**版權指令**的政策。
- 發布有關用於**GPAI 模型訓練內容的詳足摘要**。

通用人工智慧（GPAI）

特別要求（第55條）

- 除上述四項要求外，具有**系統性風險**的GPAI 模型外加履行其他義務。
- 執行模型評估，包括進行**記錄**與**對抗性測試**，以識別和減輕系統風險。
- 追蹤、記錄並向人工智慧辦公室和相關主管機關**報告嚴重事件和可能的糾正措施**，不得無故拖延。
- **網路安全**保護等級。

大型生成式AI模型

版權保護

- 由於在開發與訓練模型時需要存取大量文字、圖像、影片和其他資料。凡此內容可能受**版權保護**。
- 根據這些規則，權利持有人可以選擇保留對其權利作品或其他主題以防止文字和資料探勘，除非這樣做是為了科學研究的目的。此外也可以**保留選擇退出的權利**。

其他類型

免費和開放框架下所發布的AI模型

- 它的提供者應把模型架構及模型使用資訊對外公開，並遵守與透明度相關的要求。

情緒辨識或生物辨識分類系統

- 部署者應將系統運作告知受影響的自然人。但此義務不適用在刑事犯罪之偵查、預防或調查等司法程序。

其他類型

超大型線上平台或線上搜尋引擎

- 業者有義務評估源自其服務的設計、運作和使用的系統性風險，並予採取適當的緩解措施（風險管理架構）。
- 重在識別出某些人為生成或操縱的虛假資訊，特別是對民主進程、公民討論和公民權利的實際或可預見的負面影響。

極低風險

定義：

- 極低風險 AI 系統屬於日常應用中幾乎無風險者，對個人基本權利或社會公共利益無實質危害，因此不受額外法規限制。
- 這類系統廣泛應用於消費性產品或輔助工具中。

極低風險

範例：

- 電子郵件中的自動完成建議。
- 智慧家電（如掃地機器人）中的導航系統。
- 個人化音樂或影片推薦演算法。
- 辦公軟體中的拼字檢查或語法建議功能。
- 根據使用習慣自動調整亮度的手機 AI 設定。

極低風險

非強制性義務（第56條）

- 主管機關應邀集民間團體、工商界、學術界及其他利害關係人共同參與研議。
- 鼓勵業者制定AI操作指引據以自我要求
- 附加績效評核指標
  （Key Performance Indicators，簡稱KPI）

治理機制

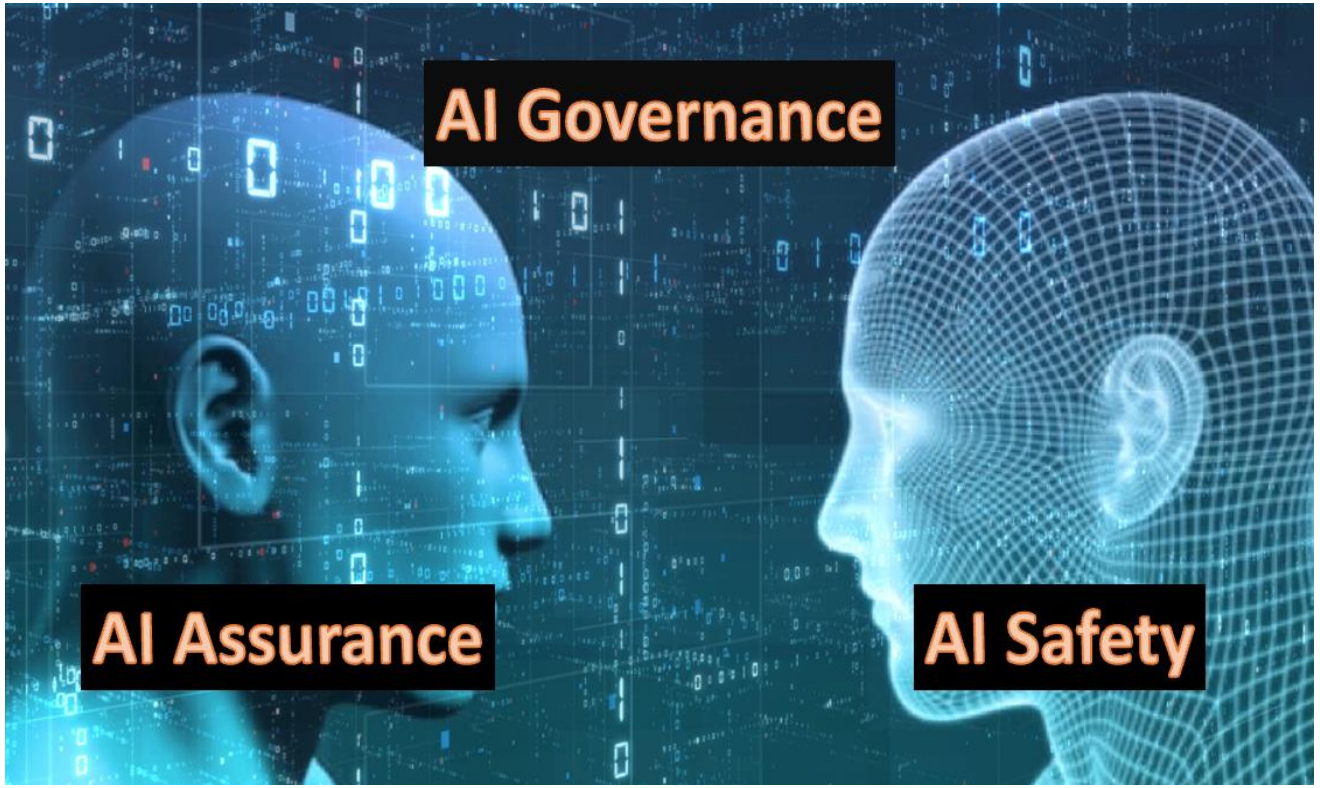
設置專責機關

制定合格標準、證書、註冊

高風險人工智慧系統資料庫

上市監控

治理機制



嚴重事故調查報告

創新沙盒制度

機密保護

處罰效果

大綱

1 統計分析 vs. 人工智慧

2 人工智慧的可能偏差

3 歐盟人工智慧法案的風險分層

4 **智慧醫療的法律責任**

產品責任 vs. 醫師責任 vs.醫療機構責任

醫療設備之產品責任

消費者保護法第7條：

- 如有違反者，致生損害於消費者或第三人時，應負連帶賠償責任。但製造商能**證明其無過失**者，法院得**減輕其賠償責任**。（第7條第3項）
- 製造商主張其商品於流通進入市場，或其服務於提供時，符合當時科技或專業水準可合理期待之安全性者，就其主張之事實**負舉證責任**。（第 7-1 條）

醫療品質管控

SaMD:

- **市售前進行臨床評估**（clinical evaluation），以求驗證產品具備安全性、有效性，乃至與其預想之特定用途相符。
- **市售後**也在真實世界找出相應顯現之性能數據（real world performance data）多加蒐集情報，據以改善設計而使SaMD滿足臨床需求 。

LLM vs. Physicians in diagnosis

JAMA Network Open

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Group	Median (IQR), %		
	Physicians plus LLM	Physicians plus conventional resources	LLM alone
All participants	76 (66 to 87)	74 (63 to 84)	92 (82 to 97)
Level of training			
Attending	79 (63 to 87)	75 (61 to 87)	
Resident	76 (68 to 84)	74 (63 to 84)	
LLM experience			
Less than monthly	76 (63 to 84)	76 (63 to 87)	
More than monthly	79 (68 to 90)	74 (63 to 84)	

92 %

76 %

8 %

24 %

LLM 診斷 ⟹ 醫師診斷

	醫師診斷正確 76%	醫師診斷錯誤 24%
LLM 診斷正確 92%	正確醫療 70%	醫療疏失 22%
LLM 診斷錯誤 8%	產品瑕疵 但不會有醫療糾紛 6%	產品瑕疵 + 醫療疏失 2%

醫療疏失責任歸屬探討

法政策分析：病患安全把關的角度

- 如果法院以「誰蓋章、誰負責」角度來裁判。當AI錯誤率遠低於醫師時，AI發生錯誤，醫師有動機及足夠能力去修正AI錯誤？
- 醫師的道德風險：醫師花時間去發現AI錯誤(6%)，是否值得？是否有足夠能力？萬一修正後反而是醫師出錯(22%)？
- 如此醫師隨著AI錯誤率不斷下降，而逐漸放鬆把關，將可能會與醫療法保障病患醫療安全的立法意旨背道而馳。

民事訴訟賠償金額

	判賠個案數	最小金額	最大金額	中位數	期望值*
被告身分					
醫院	11 (23.4%)	1	9,777,500	583,216	136,473
診所醫師	17 (26.6%)	46,000	11,626,754	1,200,000	319,200
醫院醫師	4 (15.4%)	300,000	6,720,710	4,352,301	670,254
醫師與醫院	53 (17.5%)	60,000	23,000,857	1,800,000	315,000

吳俊穎 等：月旦法學，2014年。收錄於 吳俊穎等，實證法學：醫療糾紛的全國性實證研究，元照出版社，2014, 151-188頁。

台灣醫療訴訟的特色

- 「以刑逼民」
  - 優點
    - 刑事訴訟無須負擔訴訟費用
    - 刑事訴訟中，檢察官免費
    - 藉由檢察官發動偵查以保全證據
    - 刑事訴訟可以附加民事賠償
  - 缺點？
    - 刑事訴訟病方敗訴率極高

吳俊穎 等：月旦法學，2014年。收錄於 吳俊穎等，實證法學：醫療糾紛的全國性實證研究，元照出版社，2014, 189-216頁。

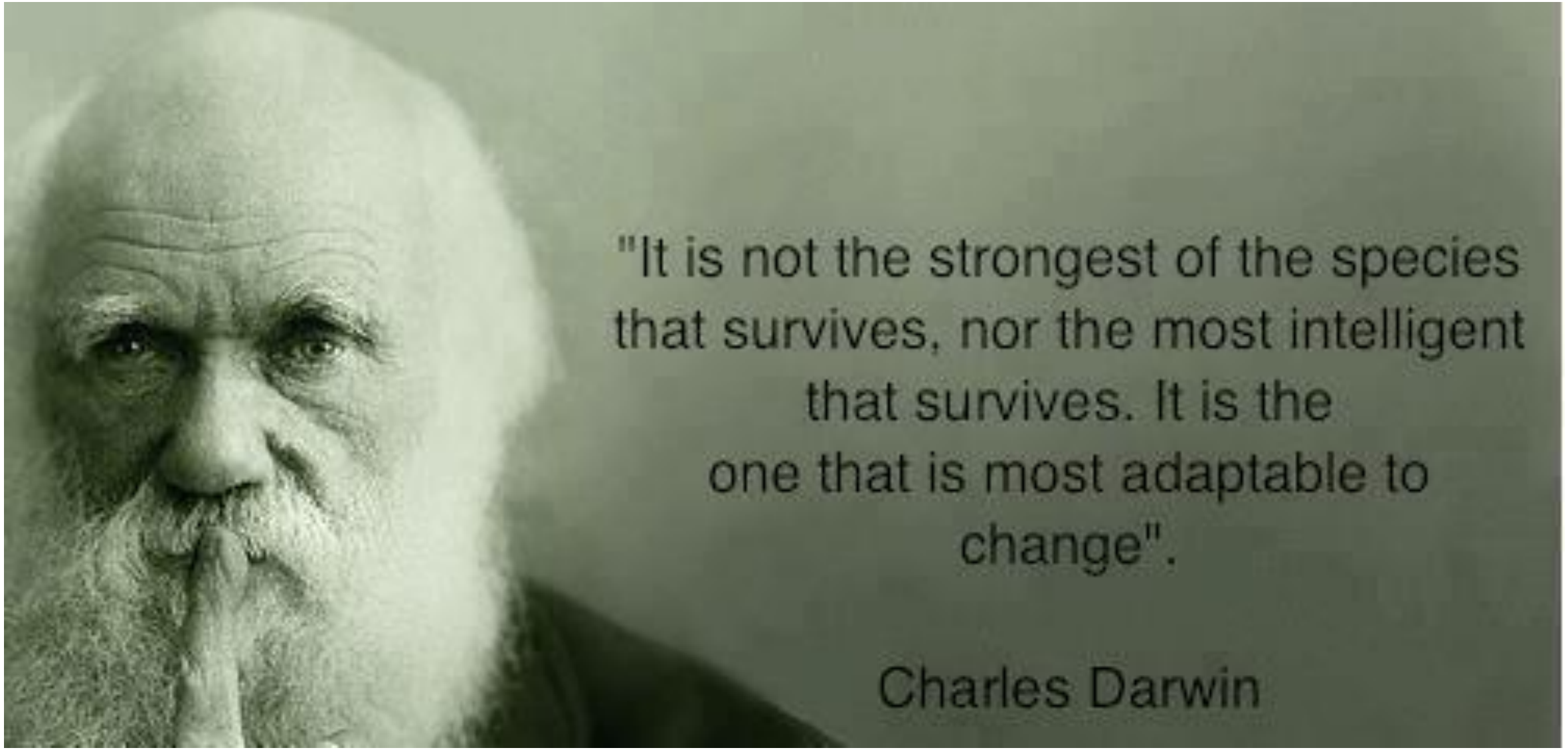
醫療機構之組織責任

侵權責任

- 智慧醫療有賴於儀器設備之正常運轉，就此部分往往必須憑藉醫療組織提供**企業化管理**，例如投入大規模的維修資金或為經營上的分工合作，方可提高臨床成效。
- 類此模式潛藏某種之**系統性風險**（systemic risks），不乏伴隨組織過失而肇生事故，醫療機構既為該危險源之製造者，唯有其本身才對危險發生具有相當的管控能力。

結 論

- 大數據與人工智慧，即將重塑醫療照護生態。
- 人工智慧的可能偏差，是需要進行規範的主要理由
- 歐盟人工智慧法，區分了不同法律規範密度的人工智慧
- 智慧醫療的法律責任，需要更多討論來釐清後關權利義務關係，並促進病患安全。


"It is not the strongest of the species
that survives, nor the most intelligent
that survives. It is the
one that is most adaptable to
change".
Charles Darwin

Acknowledgement

國立陽明交通大學
NATIONAL YANG MING CHIAO TUNG UNIVERSITY

Health Innovation Center

Director: Chun-Ying Wu

Big Data and AI research team

Microbiota research team

Community health research team