Comprendiendo la complejidad en la Salud Pública con inteligencia artificial y registros nacionales

Adrian G. Zucco
Assistant Professor in Health Complexity

COPENHAGEN HEALTH COMPLEXITY CENTER
UNIVERSITY OF COPENHAGEN

"Todos los modelos son <u>incorrectos</u> pero algunos son útiles*

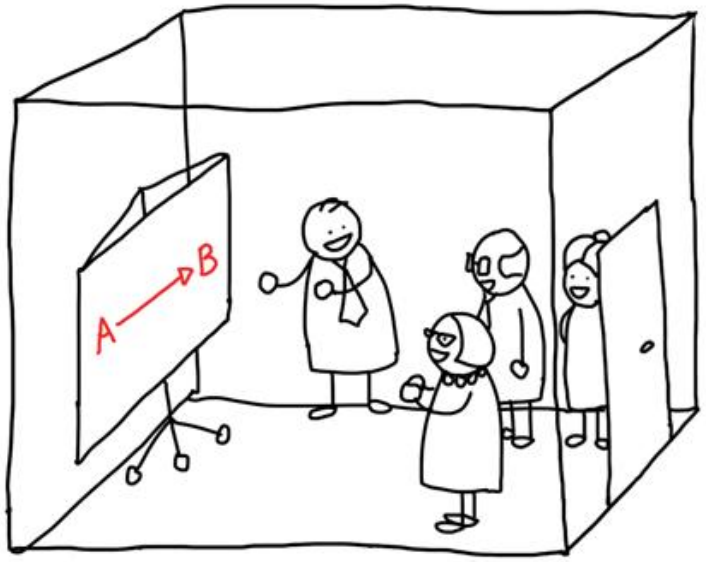
George Box, 1978

*Bajo términos y condiciones

Contenido

- La salud desde la perspectiva de los sistemas adaptativos complejos
- La aprendizaje automático como herramienta para estudiar estos sistemas
- Ejemplo usando registros de salud nacionales daneses
- Algoritmos justos
- Grandes datos en LATAM

Estudiando problemas de salud



VIRPI/BUSINESSILLUSTRATOR.COM

La salud es un fenómeno complejo

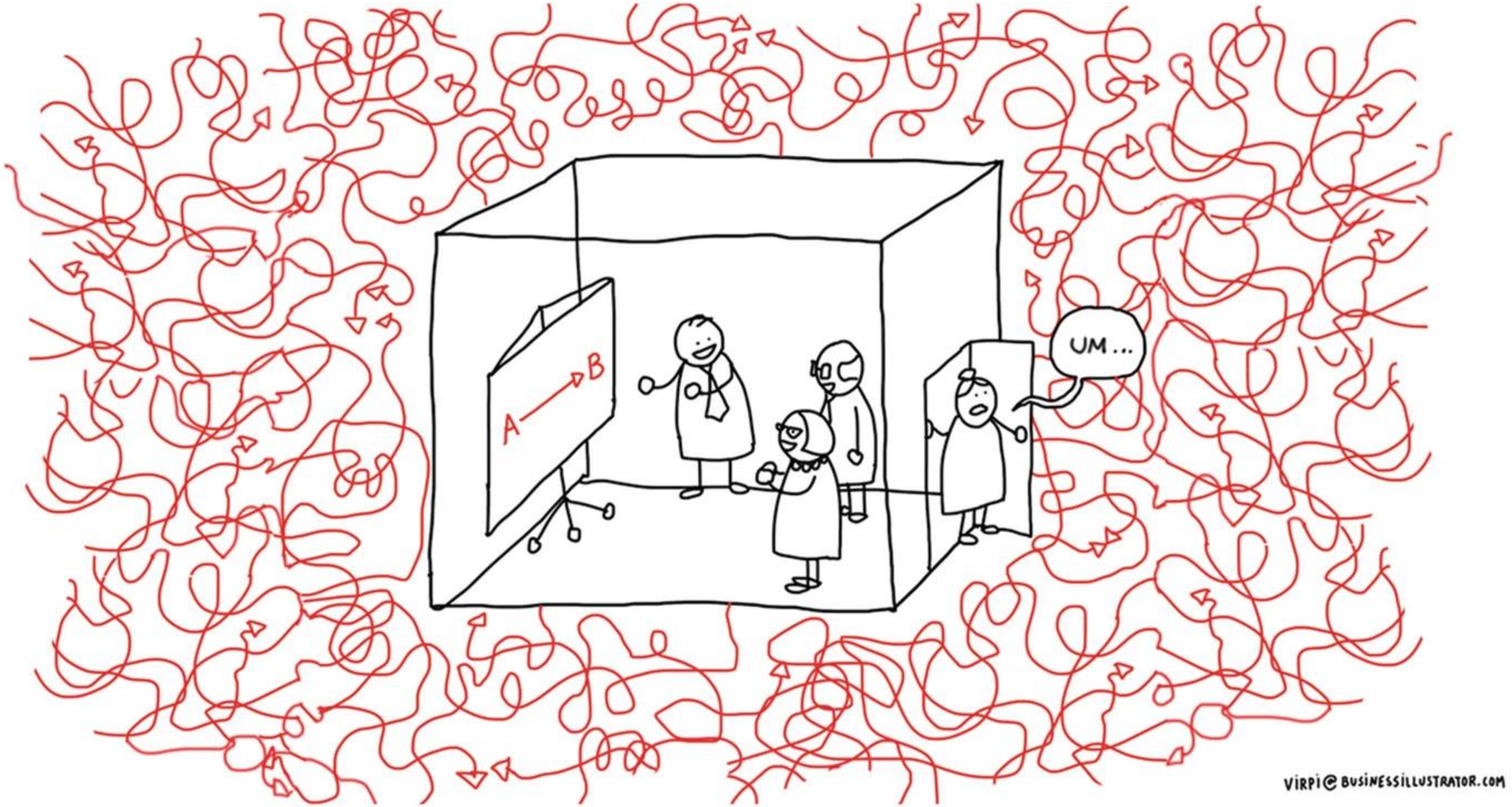
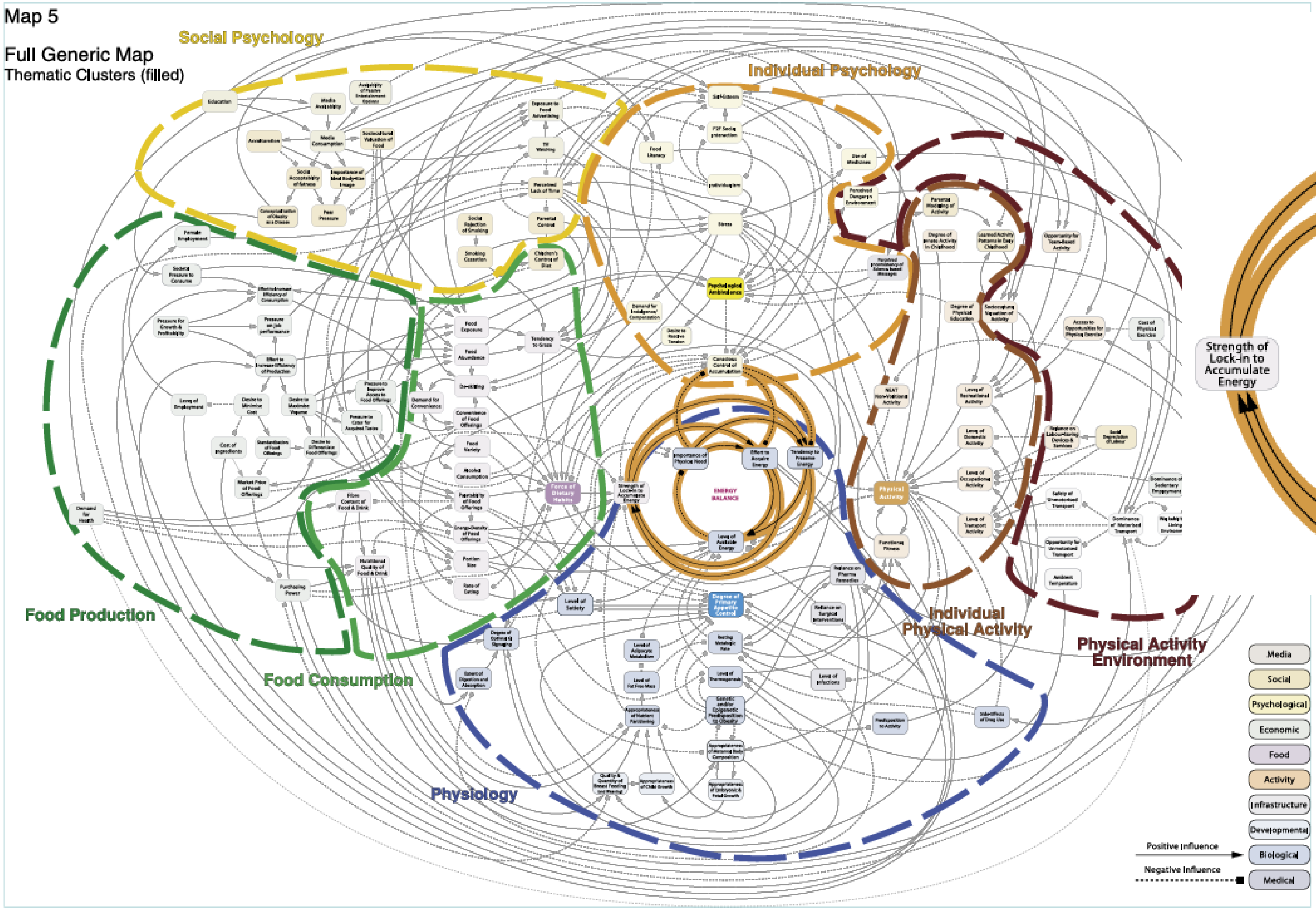
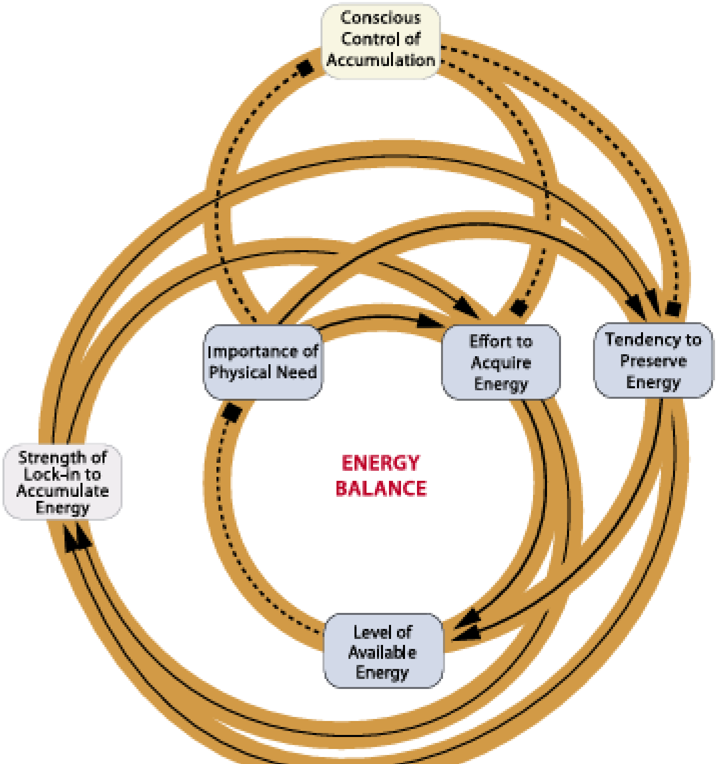


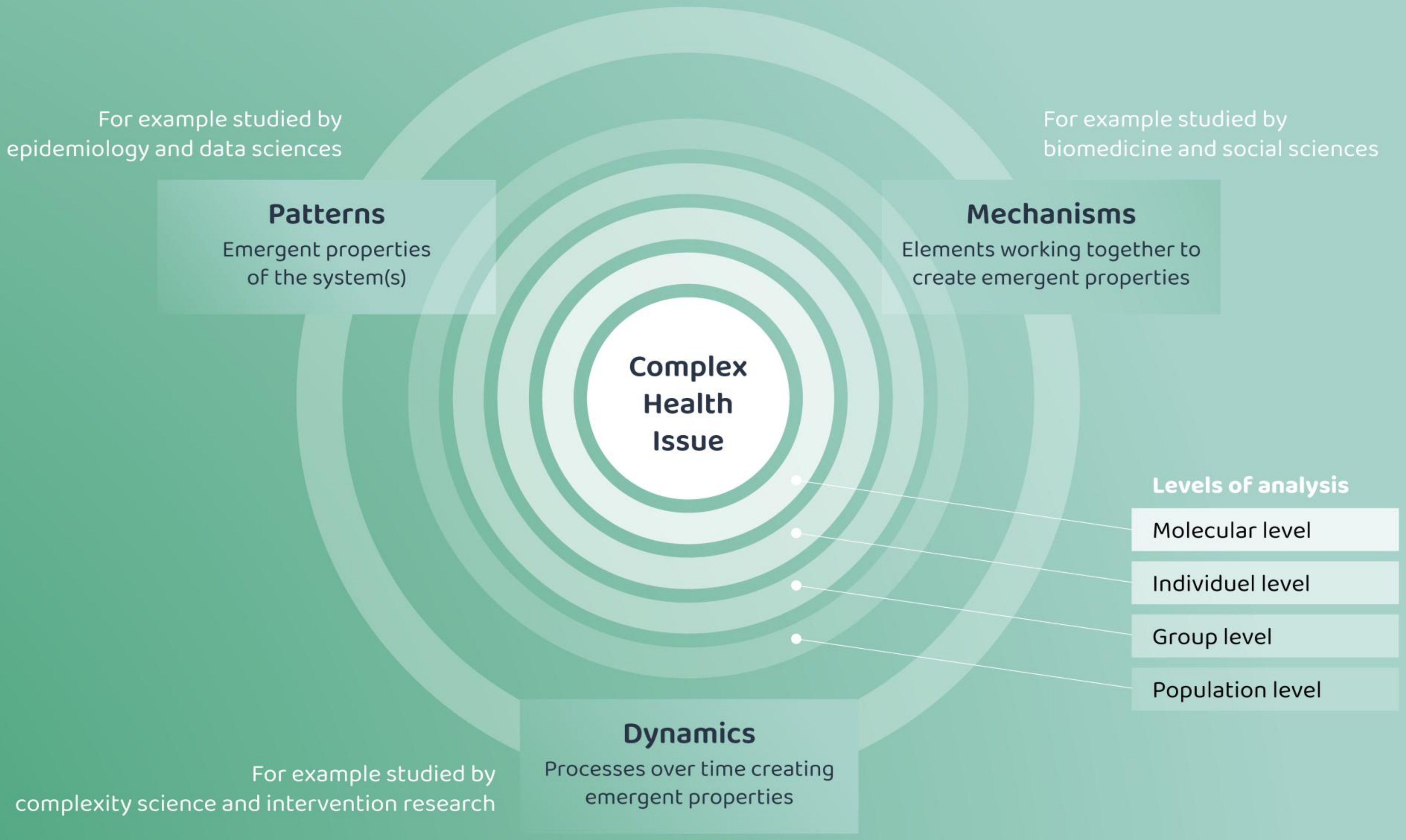
Figure 5.2: The full obesity system map with thematic clusters (see main text 5.1.2 for discussion)[17,18] Variables are represented by boxes, positive causal relationships are represented by solid arrows and negative relationships by dotted lines. The central engine is highlighted in orange at the centre of the map.



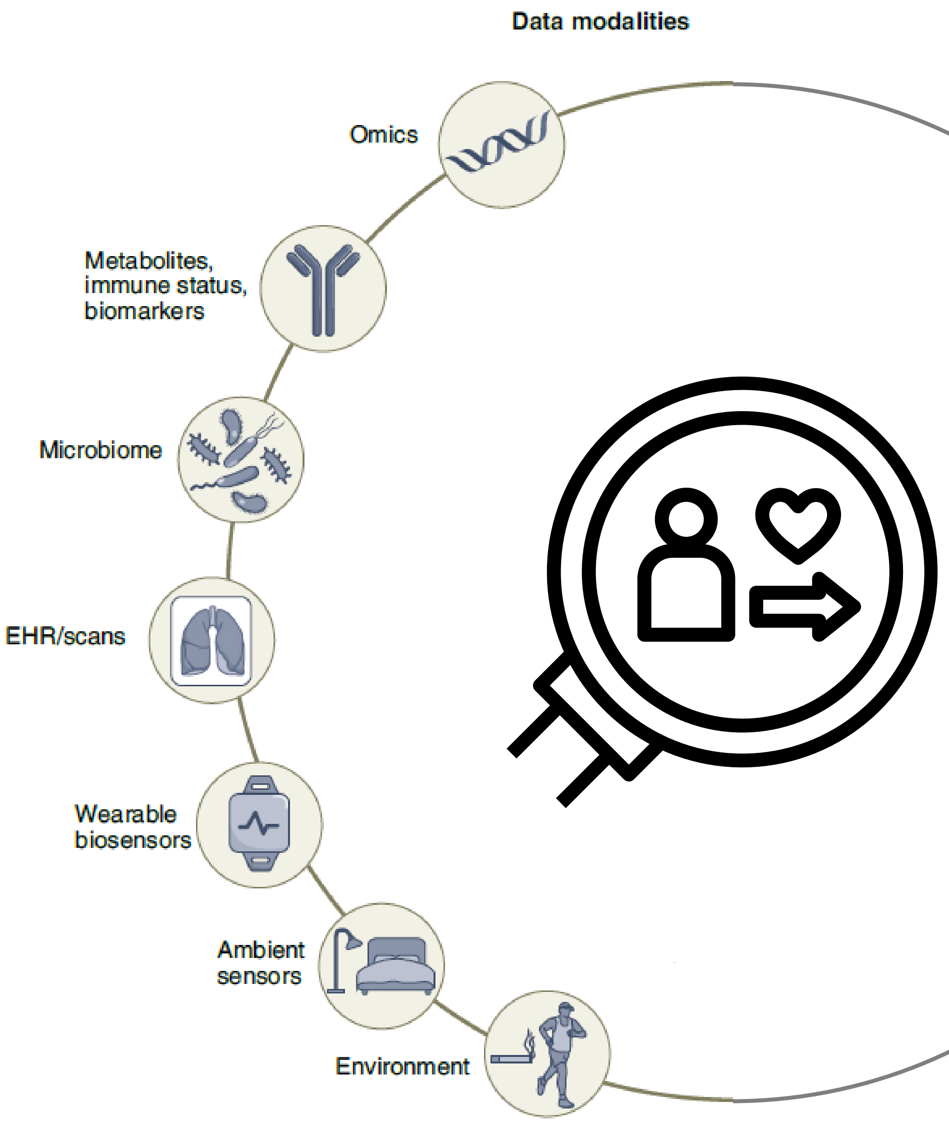


(Jebb et al., 2007)

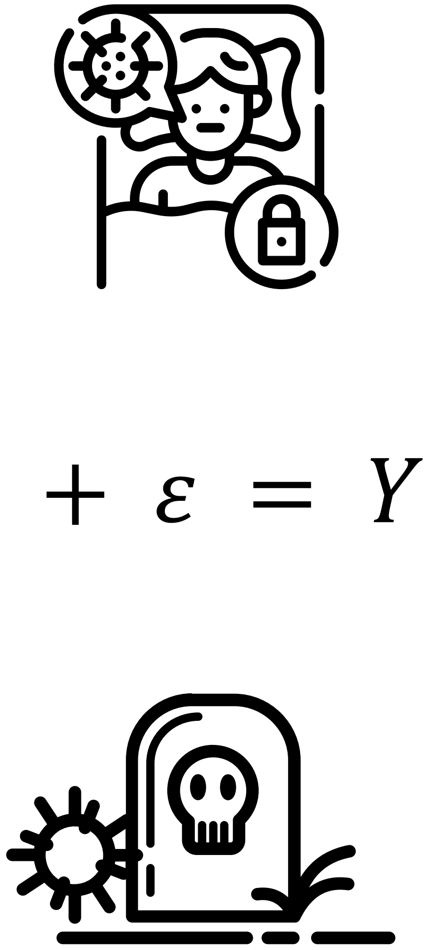
HEALTH COMPLEXITY FRAMEWORK



Rod NH, Broadbent A, Rod MH, Russo F, Arah O, Stronks K.
Complexity in Epidemiology and Public Health. Epidemiology 2023



$$f(X) + \varepsilon = Y$$



Adapted from (Acosta et al. 2022)

$$f(\boldsymbol{X}) + \varepsilon = Y$$

$$X \longrightarrow \boxed{\quad f \quad} \longrightarrow Y$$

Inspired by ”Statistical Modeling: The Two Cultures”, Breiman 2001

$$f(\boldsymbol{X}) + \varepsilon = Y$$

$$X \longrightarrow \boxed{f} \longrightarrow Y$$

Modelaje de Datos

X → *Linear regression*
Logistic regression
Cox model → Y

Inferencia

- Definir relaciones entre X e Y
- Función definida bajo supuestos estadísticos

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i + \varepsilon$$

Inspired by "Statistical Modeling: The Two Cultures", Breiman 2001

$$f(\boldsymbol{X}) + \varepsilon = Y$$

$$X \longrightarrow \boxed{f} \longrightarrow Y$$

Modelaje de Datos

X → *Linear regression* / *Logistic regression* / *Cox model* → Y

Modelaje algorítmico

X → ■ → Y

Decision trees
Neural networks

Inferencia

- Definir relaciones entre X e Y
- Función definida bajo supuestos estadísticos

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_i X_i + \varepsilon$$

Predicción

- Predicciones de Y en base a X
- Función indefinida y aprendida por algoritmos

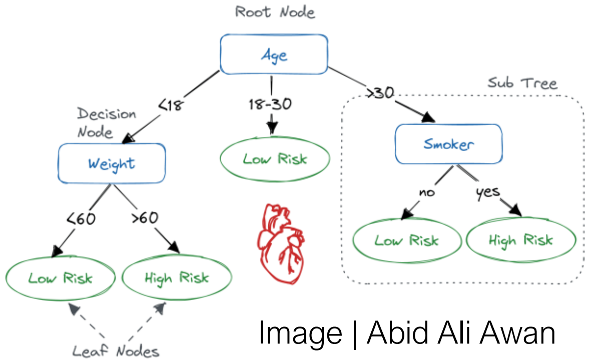
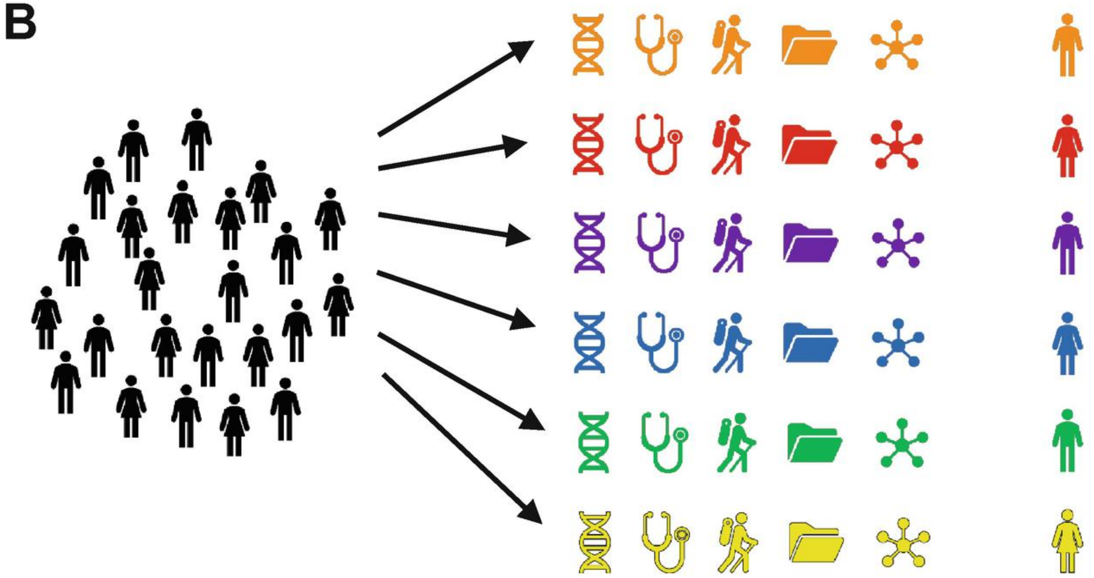


Image | Abid Ali Awan

Inspired by "Statistical Modeling: The Two Cultures", Breiman 2001

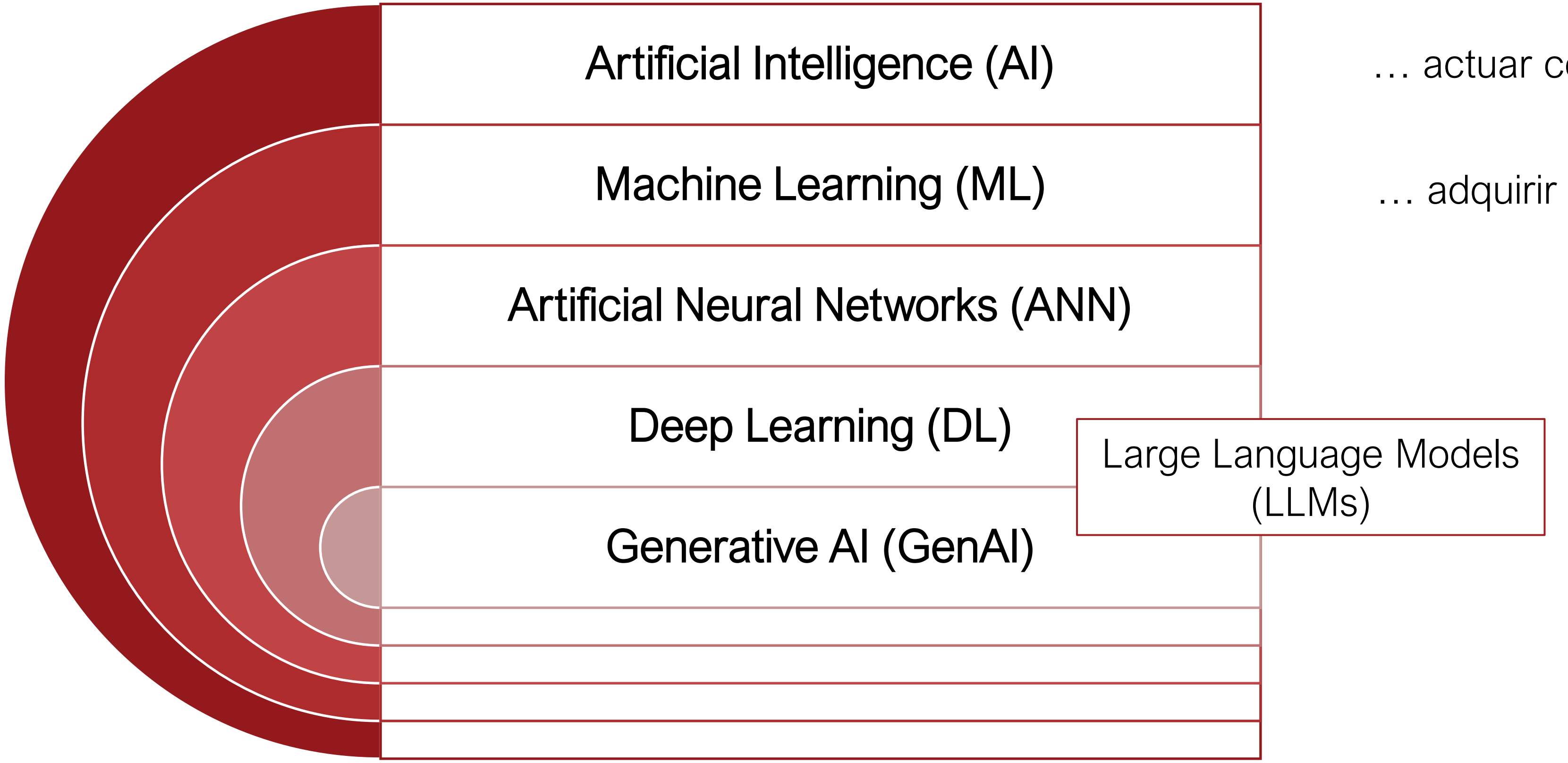


Estratificación de individuos agrupándolos por sexo, edad, genética, etnicidad, biomarcadores, etc que se puedan tratar uniformemente

Integrar toda la información posible para desarrollar predicciones de riesgo y estrategias de tratamiento específicas para cada individuo

Currie G, Delles C. Adv Exp Med Biol. 2018

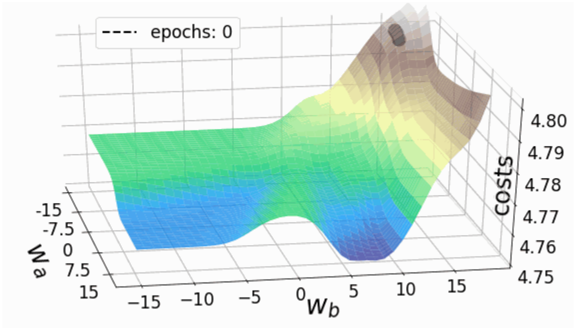
Taxonomia de la Inteligencia Artificial (IA)



Habilidad de los sistemas computacionales para:

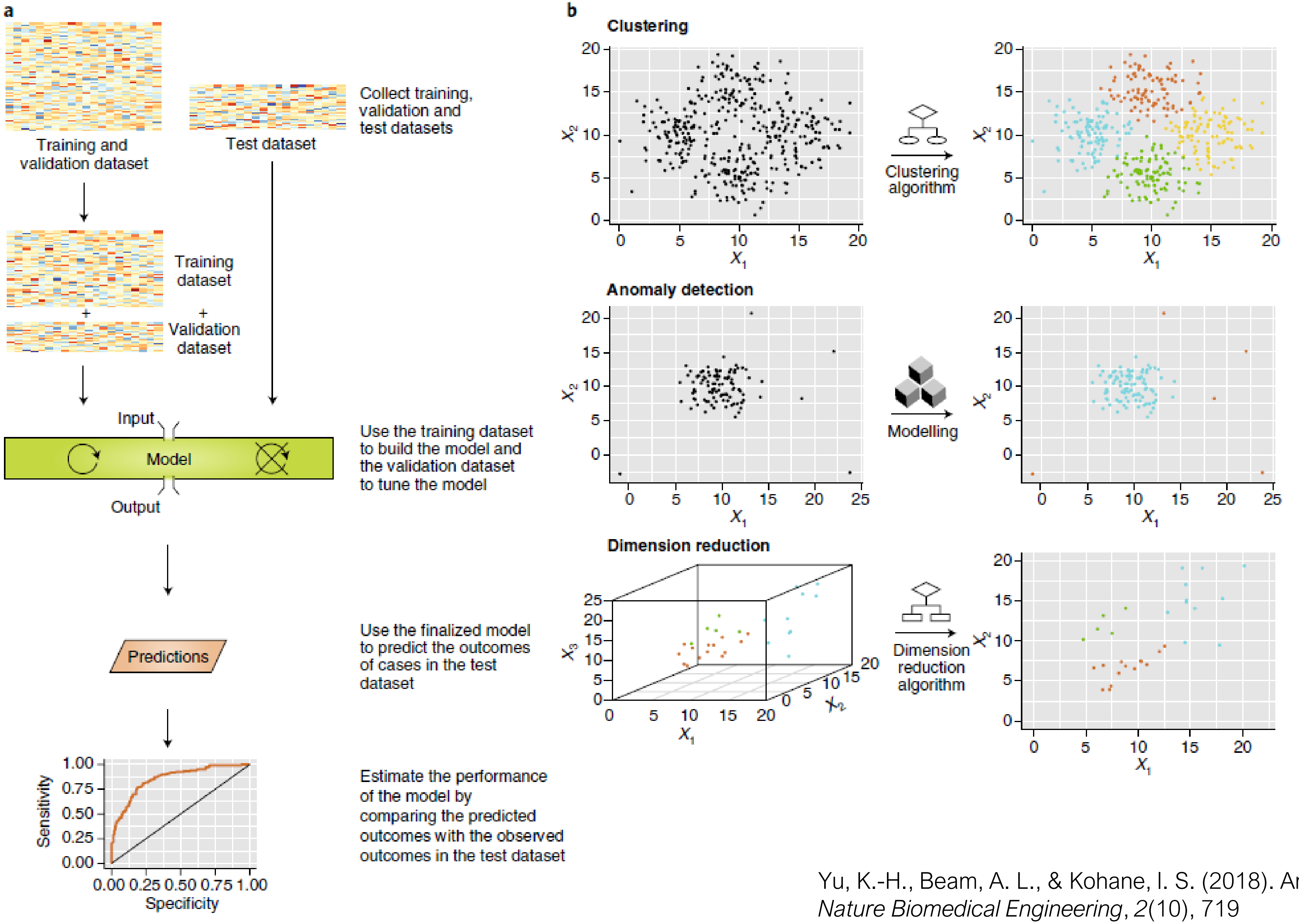
… actuar conocimiento programado por humanos

… adquirir conocimiento



Usar datos para aprender reglas complejas por minimización de errores

Machine learning: Supervised vs unsupervised



Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719

Aprendizaje automático como herramienta científica

Interacciones

Efectos no-lineales

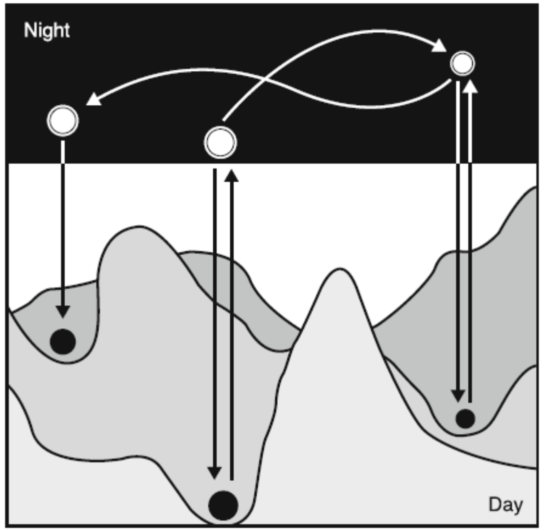
Aprendizaje automático como herramienta científica

Interacciones

Efectos no-lineales

Incluir tantos datos como sea posible, midiendo la incertidumbre en lugar de ignorarla

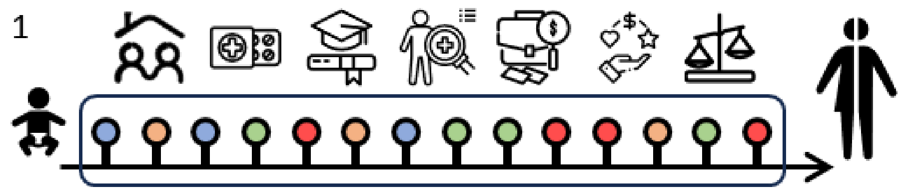
Explorar la estructura en los datos que sugieran nuevas hipótesis



"Ciencia nocturna": explorar el campo no estructurado de posibles hipótesis e ideas no investigadas

(Yanai & Lercher, 2019)

Usando los registros de salud e inteligencia artificial

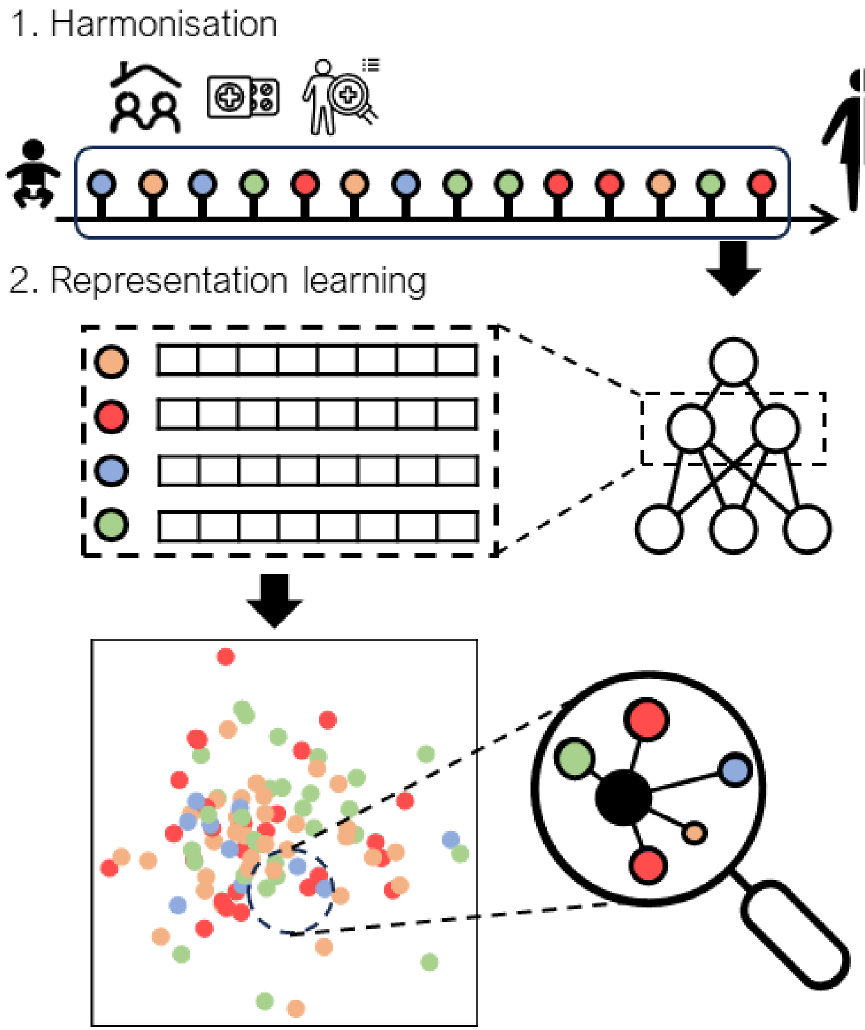


- Secuencia de eventos (palabras) en la vida de un individuo (frases)

- Generar representaciones numéricas (embeddings) de los eventos

- Calcular distancias entre eventos para ver estructuras y relaciones (patrones)

Ejemplo de diabetes en registros nacionales daneses

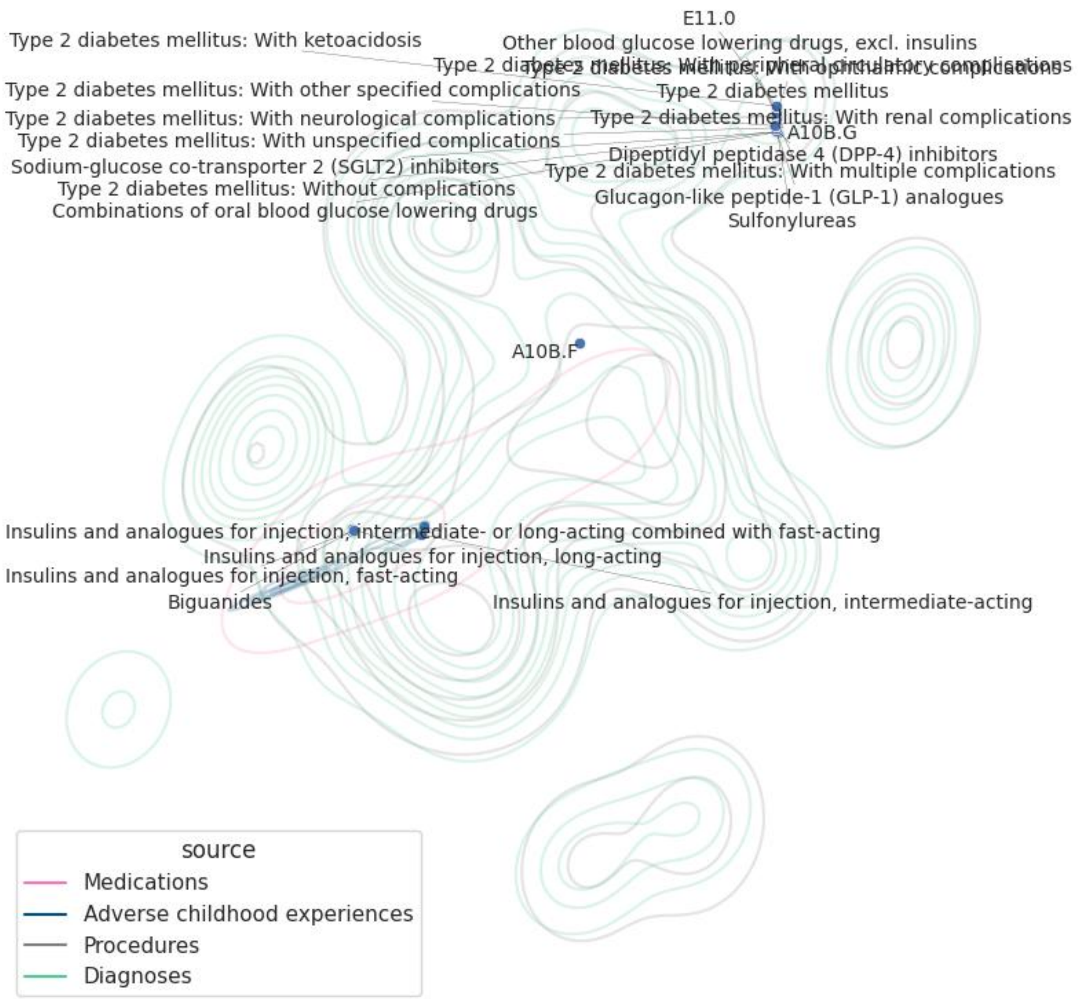


- Datos de 2.2 millones de individuos:
  - DANLIFE, nacidos entre 1980-2015
- Búsqueda de términos clinicos:
  - Med: A10 - DRUGS USED IN DIABETES
  - Diag: E11 - Type 2 diabetes mellitus

❖ Dimensionality reduction

❖ Hierarchical clustering of neighboring terms

Ejemplo (en desarrollo)



Explainable Machine Learning

$$X \longrightarrow \boxed{f?} \longrightarrow Y$$

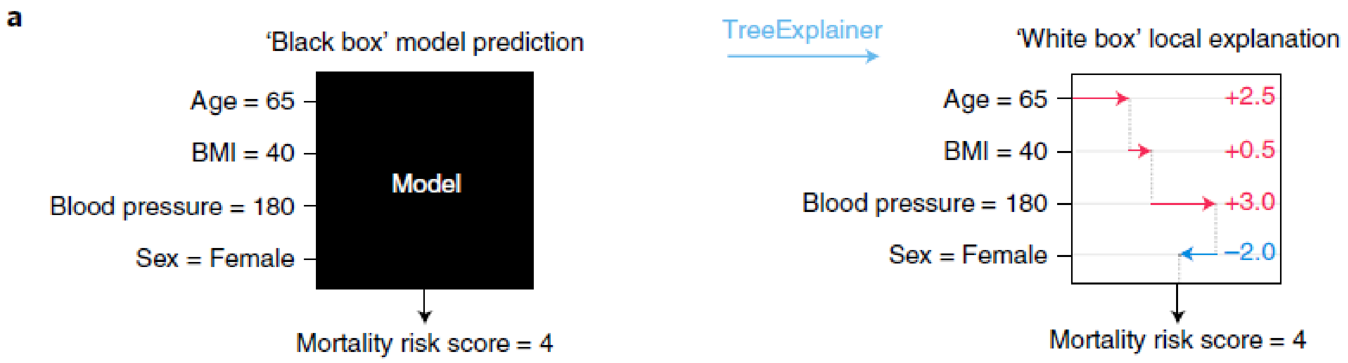
~~A “black box”?~~

Explainable Machine Learning

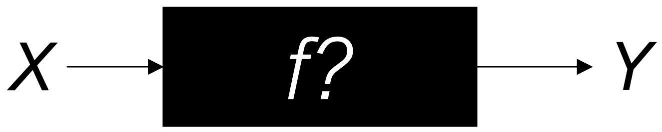
~~A "black box"?~~

SHapley Additive exPlanations (SHAP)

- Basado en teoría del juego
- Contribución de cada variable a los valores predichos por un modelo



Explainable Machine Learning

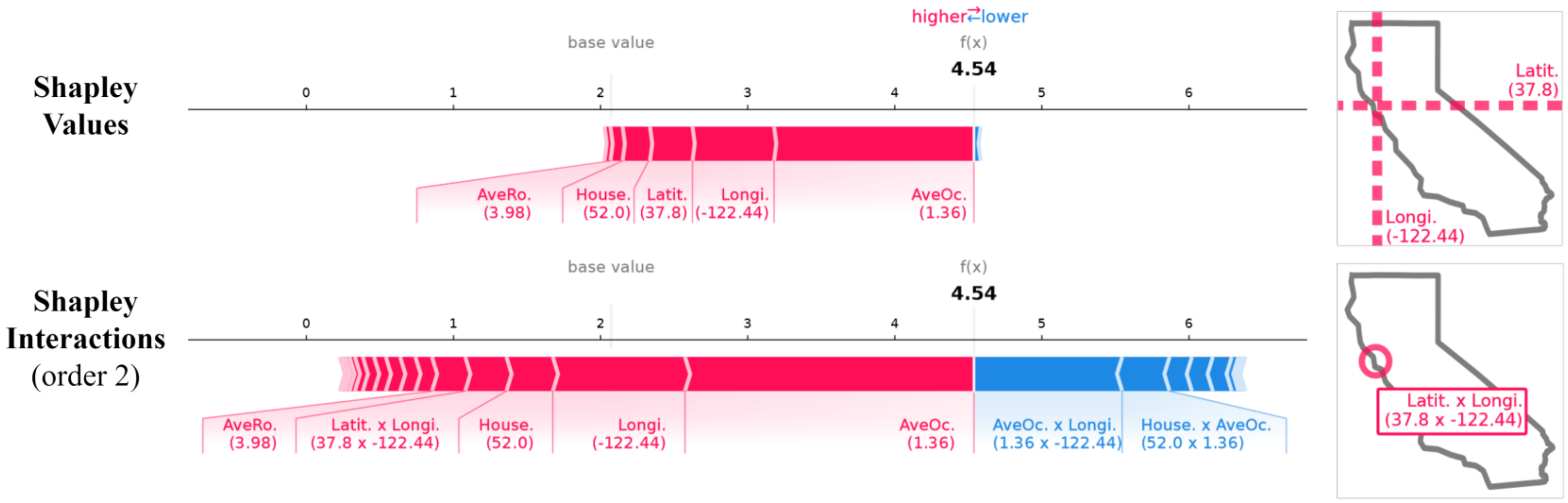


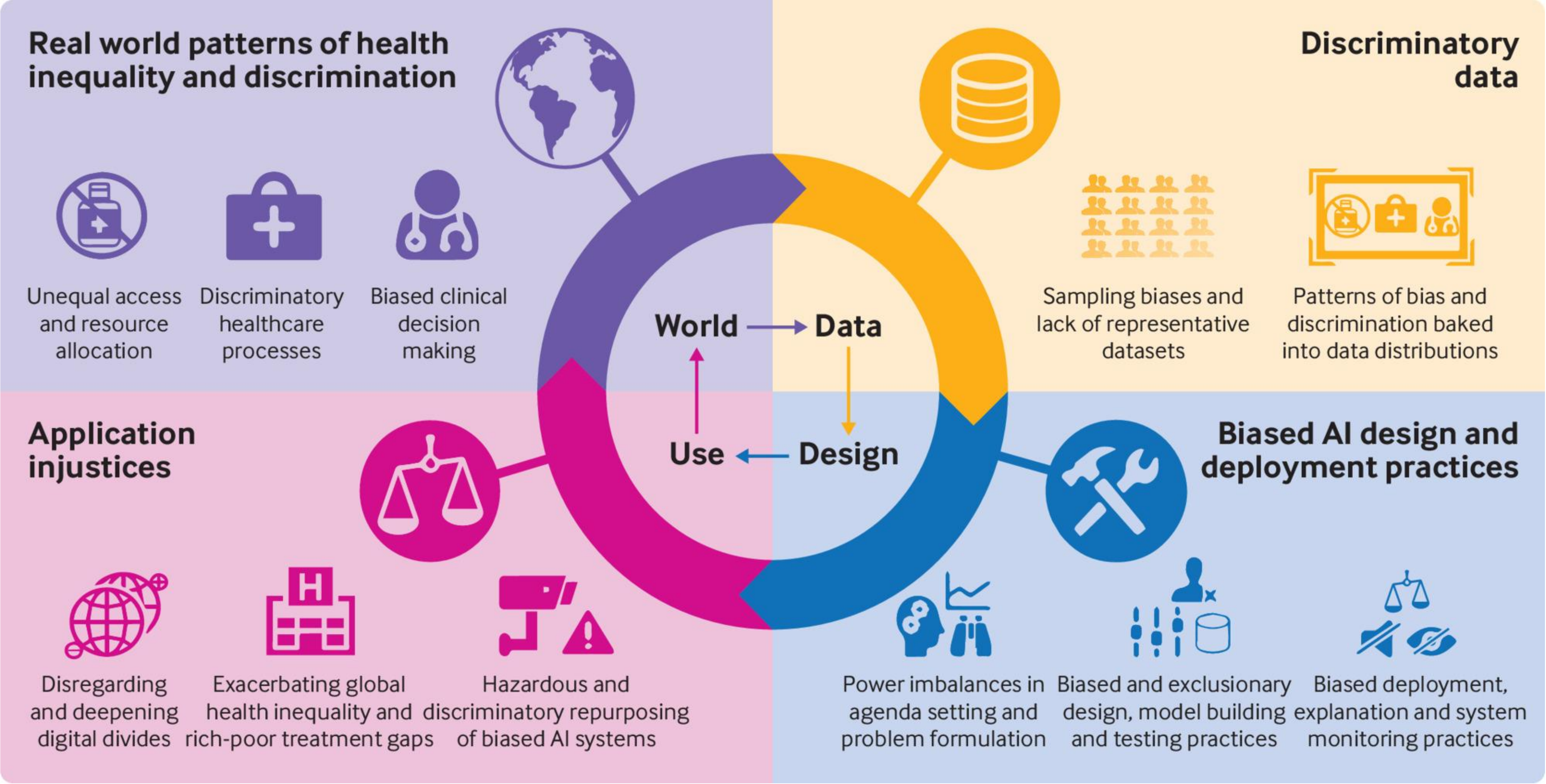
~~A “black box”?~~

SHapley Additive exPlanations (SHAP)

- Basado en teoría del juego
- Contribución de cada variable a los valores predichos por un modelo
- Interacciones Shapley

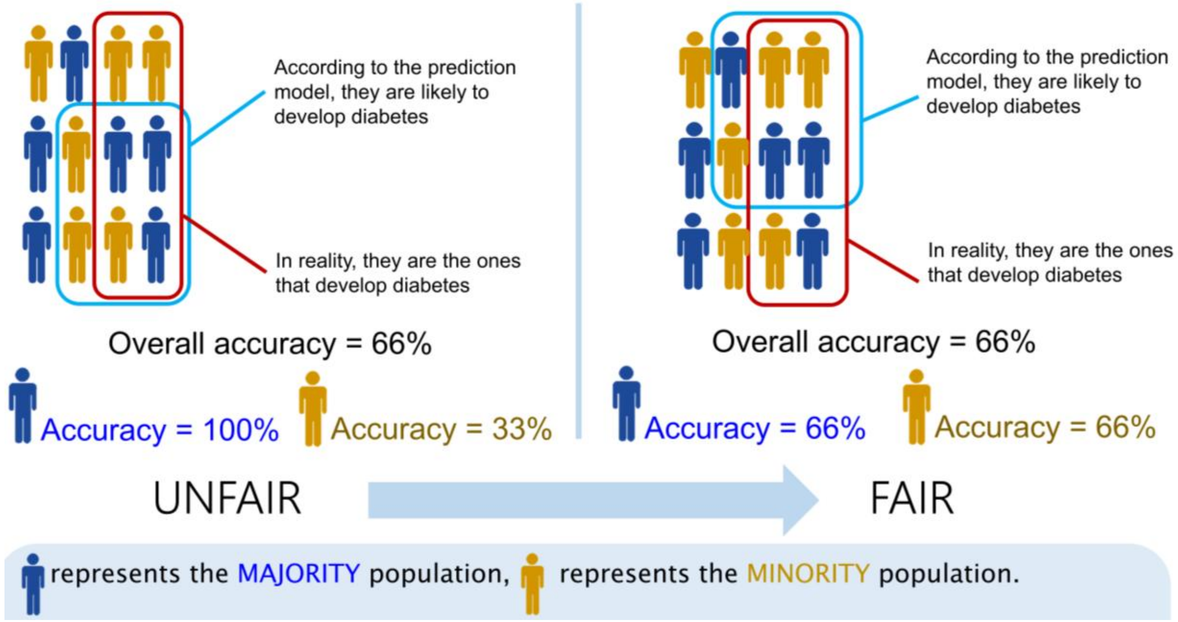
Potential Question: "Does the *location* of my property affect its price $\hat{y} = 454\,000$?"

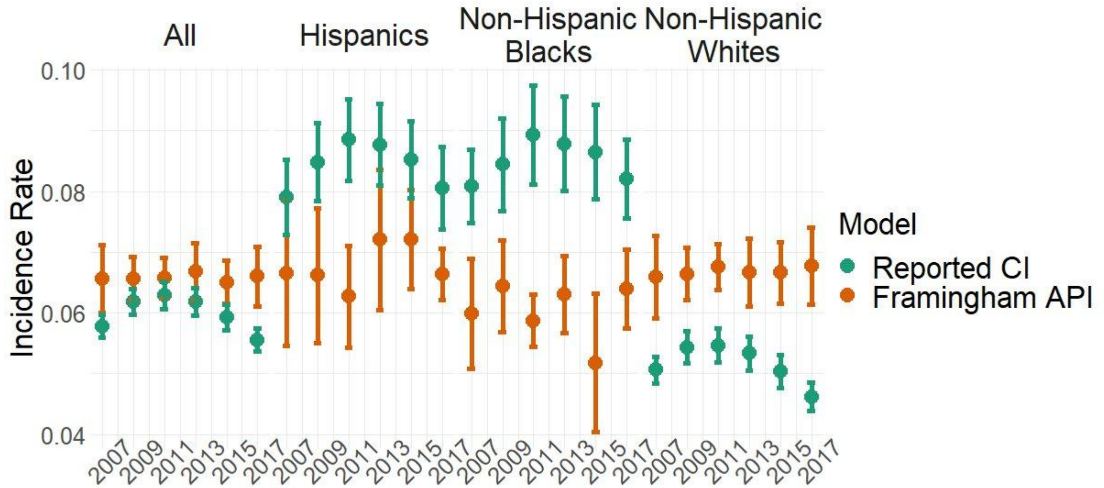




(Leslie et al., 2021)

AlfaDiab - Algorithmic Fairness in Diabetes Prediction





Comparison of predicted incidences by the Framingham Offspring Risk Score type 2 diabetes prediction algorithm (red), and real life incidences (green). This landmark risk score performs reasonably well on average, but underestimates risk for Hispanics and Blacks, and overestimates risk for Whites.

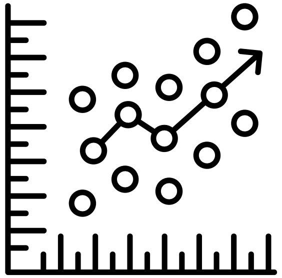
(Cronjé et al., 2023)

Imágenes cortesía de Dr. Tibor V. Varga

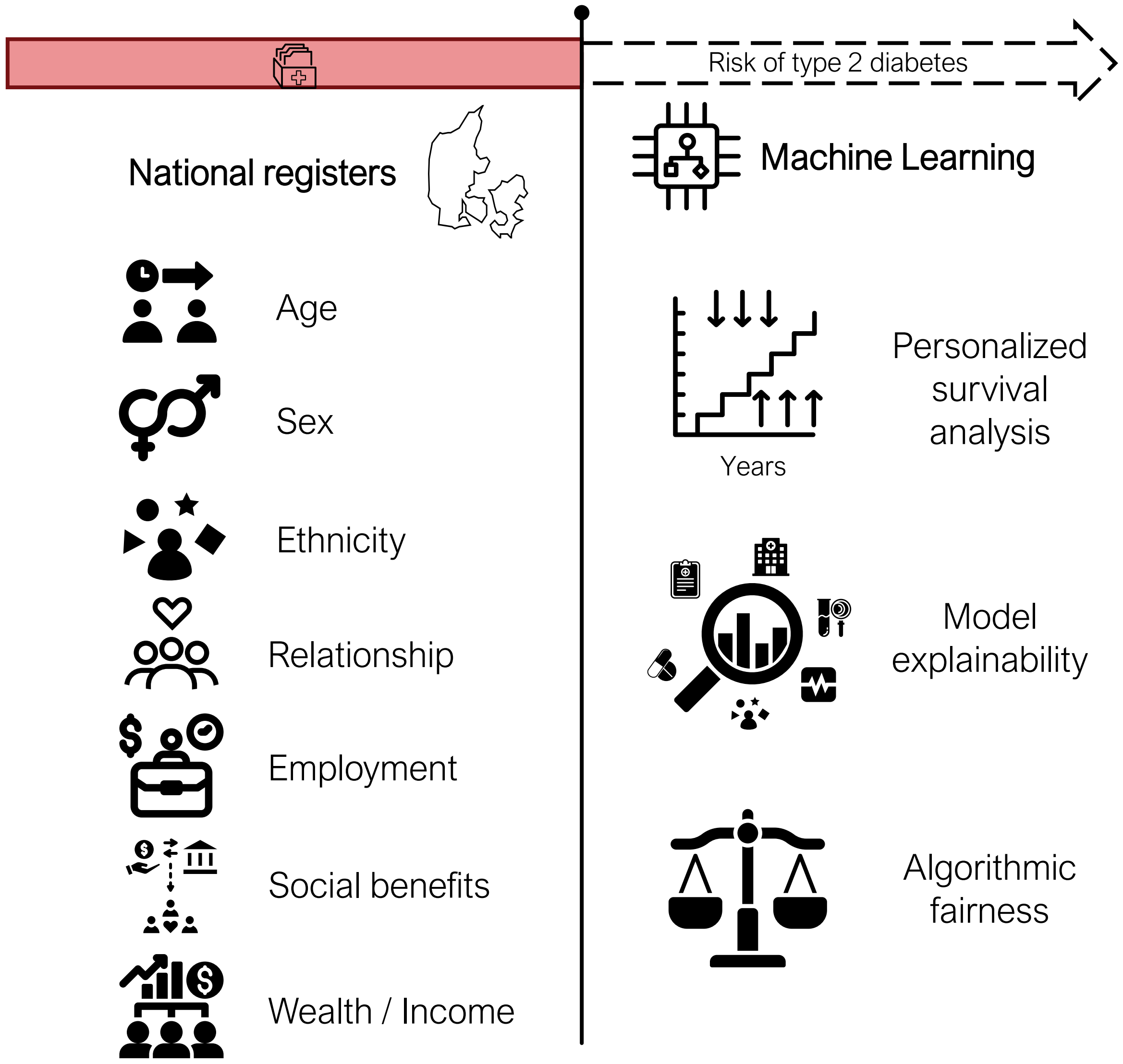
AlfaDiab - Algorithmic Fairness in Diabetes Prediction

Risk of type 2 diabetes

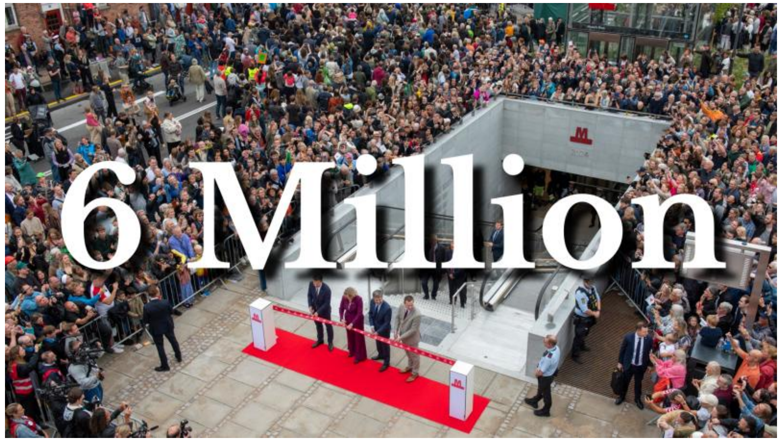
- Age
- Sex
- Ethnicity
- Laboratory tests
- Family history
- Medications



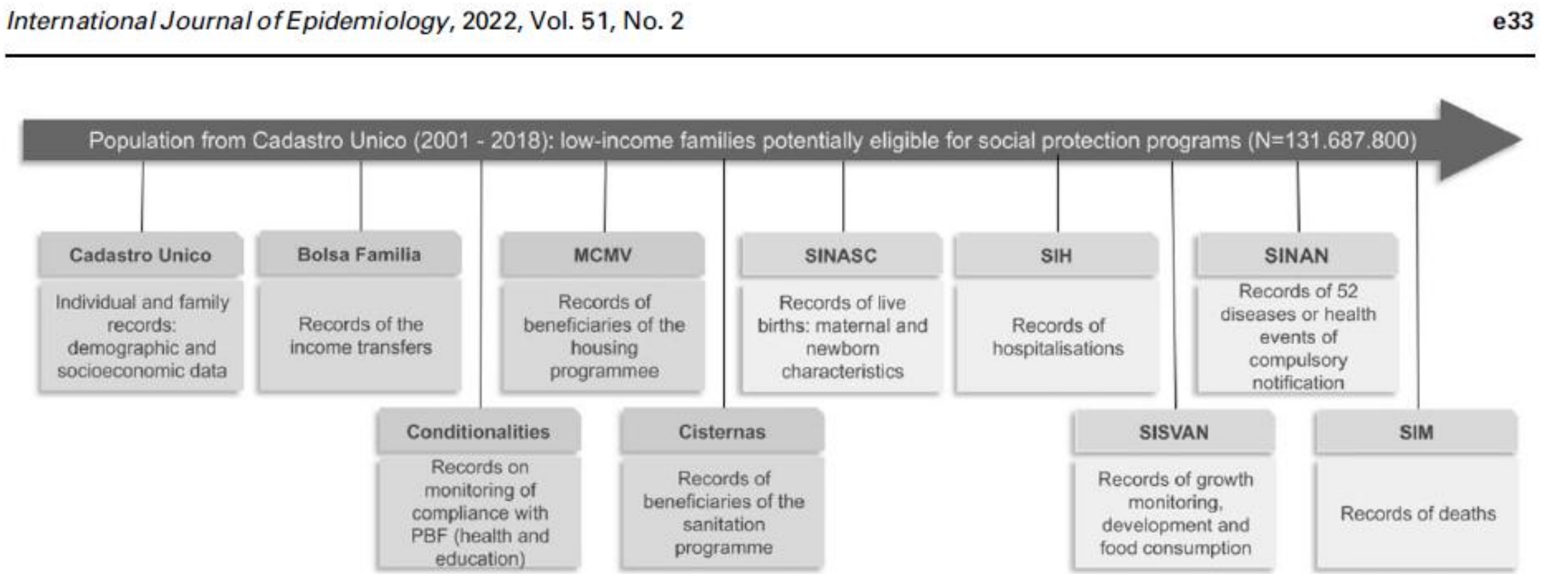
AlfaDiab - Algorithmic Fairness in Diabetes Prediction



Cohortes en LATAM?

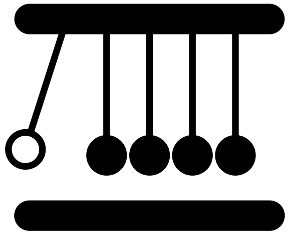


- Población en Dinamarca: 6 millones
  - Larga tradición de registros de salud
  - Diagnósticos a nivel de hospitales registrados
  - Sociedad bastante homogénea
- Brasil:
  - The 100 Million Brazilian Cohort



Perspectivas a futuro

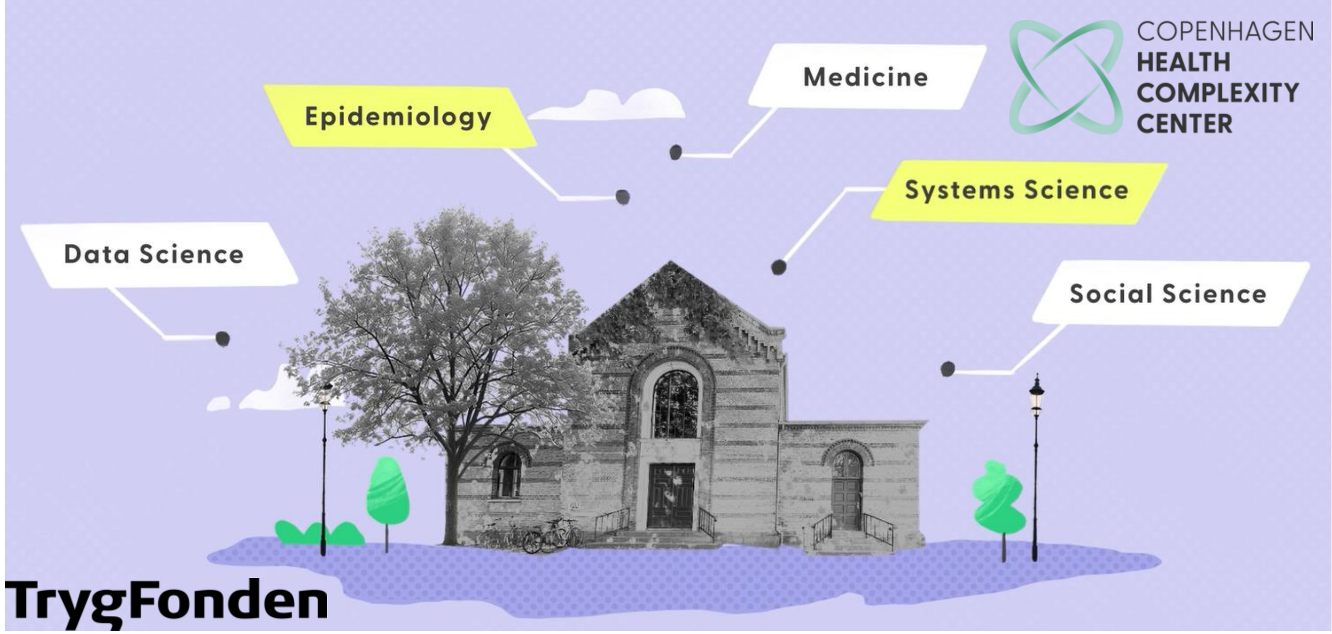
- **Métodos mixtos y modelos causales**
  - Integración de datos cualitativos con herramientas cuantitativas
  - Uso de principios de inferencia causal ("causal inference")



- **Entender la complejidad de la salud con datos diversos**
  - Inclusión de componentes ambientales y psicosociales
  - Mapeo de sistemas adaptativos complejos en salud pública


Prof. Naja Hulvej Rod


Assoc. Prof. Tibor V. Varga



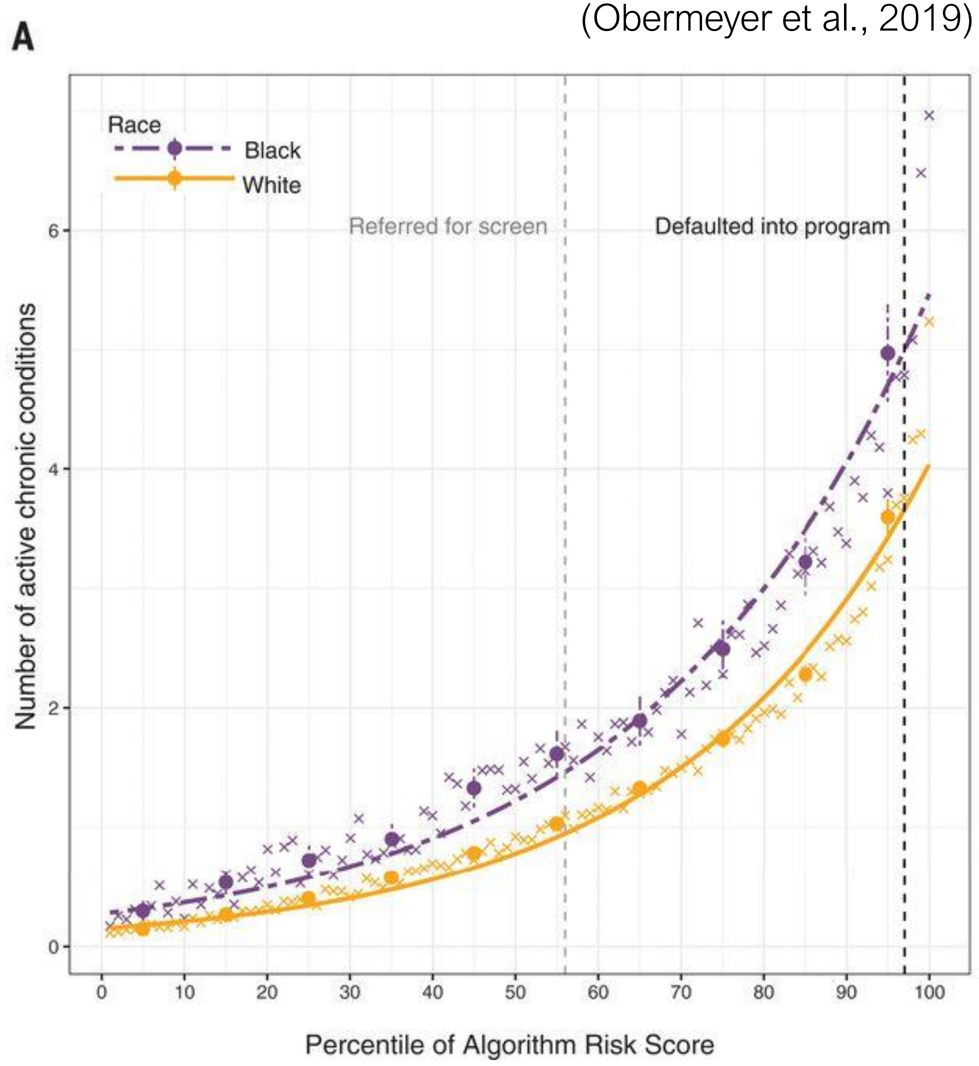
adrigabzu@sund.ku.dk / adrianzucco.com

Diapositivas extra

Perpetuando sesgos sociales

(Obermeyer et al., 2019)

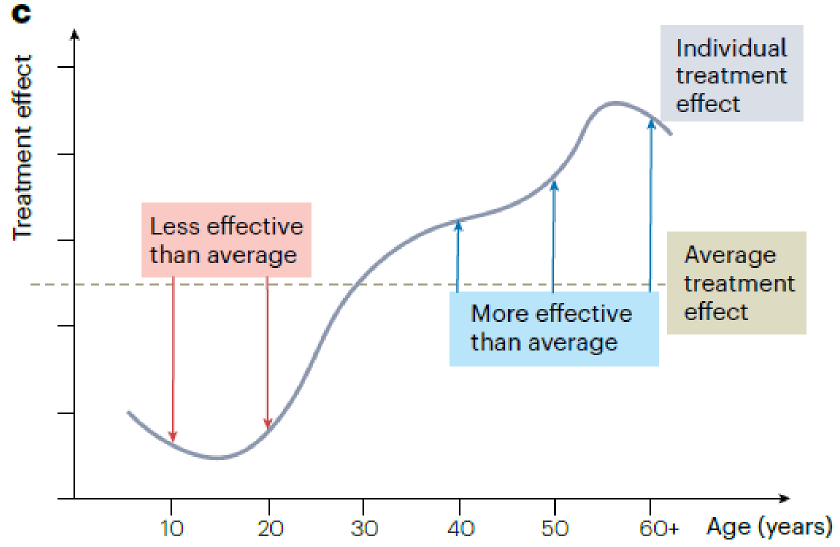
- Healthcare algorithm for screening and preventive intervention
- Applied to roughly 200 million people in the United States each year
- Trained on health costs as a proxy for health needs
  - Less money is spent on Black patients who have the same level of need

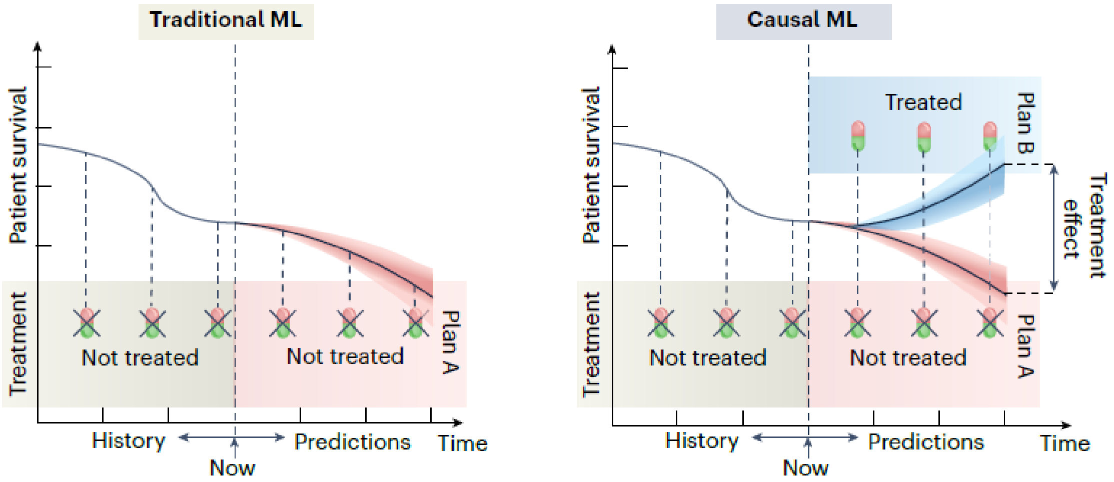


"At the same level of algorithm-predicted risk,
Blacks have significantly more illness burden than Whites."

Causal inference approaches

- Inverse probability weighting
- Propensity scoring:
  - 50% in completely randomized trials with two treatment arms of equal size
- G-computation
- Targeted maximum likelihood estimation
- Causal Machine Learning





(Feuerriegel et al., 2024)

A new perspective on causality

	Layer (Symbolic)	Typical Activity	Typical Question	Example
$\mathcal{L}_1$	Associational $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell us about the disease?
$\mathcal{L}_2$	Interventional $P(y\|do(x),c)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured?
$\mathcal{L}_3$	Counterfactual $P(y_x\|x',y')$	Imagining	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

Table 1.1: Pearl's Causal Hierarchy.

Observational studies

RCTs

(Bareinboim et al., 2022)