

# La Inteligencia Artificial y las Estadísticas

Cómo la Inteligencia Artificial está redefiniendo la  
estadística y la toma de decisiones

**Expo Estadísticas 2025**  
Caguas, Puerto Rico

**Dr. Carlos Ortiz Reyes**  
Recinto de Ciencias Médicas



“La inteligencia artificial es una de las tecnologías más poderosa que jamás hayamos creado. Tiene el potencial de revolucionar todos los aspectos de nuestras vidas, desde la forma en que trabajamos hasta la forma en que nos relacionamos con los demás.”

– Fei-Fei Li (Professor, Stanford University)



# El Renacer de los Datos

*Vivimos un momento histórico sin precedentes donde la explosión de datos converge con los avances tecnológicos, creando una demanda creciente por profesionales capacitados en las ciencias de datos y en la AI.*

## 01

### Explosión de Datos

Cada día se generan **2.5 quintillones de bytes** de datos. El 90% de los datos mundiales fueron creados en los últimos dos años. Esta explosión sin precedentes requiere métodos rigurosos para extraer conocimiento valioso.

## 02

### Convergencia de Disciplinas

La **estadística**, la **inteligencia artificial** y la **computación** se fusionan para crear un campo interdisciplinario poderoso. Ya no son segmentos separados, sino herramientas complementarias que se potencian mutuamente.

## 03

### Demanda Profesional

El mercado global de IA alcanzará **\$4.8 billones para 2033**. Los empleos de Estadísticos y Data Scientists, se espera que crezcan un **34% (2034)**. La demanda de profesionales que dominen ambos mundos es exponencial.

*Este contexto crea una oportunidad única para los estadísticos. La oportunidad de liderar la revolución de datos.*

# Definiciones Fundamentales

Antes de explorar su convergencia, es esencial comprender qué son la Estadística y la Inteligencia Artificial, y cómo se relacionan.

## 01

### Estadística

La **estadística** es la ciencia de recolectar, analizar, interpretar y presentar datos. Proporciona métodos rigurosos para extraer conclusiones válidas de la información, cuantificar incertidumbre y establecer relaciones causales entre variables.

**Elementos clave:** Inferencia estadística, diseño experimental, pruebas de hipótesis, modelado probabilístico, análisis de varianza, regresión, y muestreo.

## 02

### Inteligencia Artificial

La **Inteligencia Artificial** es el campo de la computación dedicado a crear sistemas que pueden realizar tareas que normalmente requieren inteligencia humana: reconocimiento de patrones, toma de decisiones, aprendizaje y predicción.

**Elementos clave:** Aprendizaje automático (Machine Learning), redes neuronales, procesamiento de lenguaje natural, visión por computadora y los agentes de IA.

### La Relación Fundamental

La IA moderna se construye sobre **fundamentos estadísticos**. Los algoritmos de aprendizaje automático son, en esencia, **métodos estadísticos aplicados a escala computacional**.

La estadística proporciona el rigor teórico; la IA proporciona la capacidad de implementación a gran escala.



# El Fundamento: La Estadística

La estadística proporciona los cimientos teóricos y metodológicos que garantizan que las conclusiones derivadas de datos sean válidas, confiables y científicamente rigurosas.

## El Fundamento Teórico

### Teoría de la Probabilidad

Base matemática que permite modelar la incertidumbre y calcular la probabilidad de eventos futuros basándose en datos observados.

### Inferencia Estadística

Proceso de extraer conclusiones sobre poblaciones completas a partir de muestras representativas, cuantificando siempre la incertidumbre asociada.

### Diseño Experimental

Metodología rigurosa para planificar estudios que permitan establecer relaciones causales y controlar variables de confusión.

### Cuantificación de Incertidumbre

Capacidad única de medir y comunicar el grado de confianza en nuestras conclusiones mediante intervalos y pruebas de hipótesis.

## El Valor Indispensable

### Rigor Científico

La estadística proporciona el marco metodológico que distingue la ciencia de la especulación. Garantiza que nuestras afirmaciones estén respaldadas por evidencia cuantificable y reproducible.

### Validez Metodológica

Asegura que los métodos utilizados sean apropiados para el tipo de datos y preguntas de investigación, evitando conclusiones erróneas por análisis inadecuados.

### Decisiones Basadas en Evidencia

Permite a organizaciones y gobiernos tomar decisiones informadas, minimizando riesgos y maximizando resultados mediante análisis objetivo de datos.

### Comunicación de Confianza

Los intervalos de confianza y valores que permiten comunicar no solo resultados, sino también el nivel de certeza que debemos tener en ellos.

# El Fundamento: La Inteligencia Artificial

La Inteligencia Artificial se construye sobre tres pilares fundamentales que le permiten simular capacidades cognitivas humanas y resolver problemas complejos de manera autónoma.

- **Algoritmos de Aprendizaje Automático**

Sistemas que aprenden patrones a partir de datos sin ser explícitamente programados para cada tarea específica.

- **Redes Neuronales Profundas**

Ofrece una Infraestructura distribuida que permite analizar datasets masivos (terabytes, petabytes) que exceden las capacidades de métodos tradicionales, democratizando el análisis de Big Data.

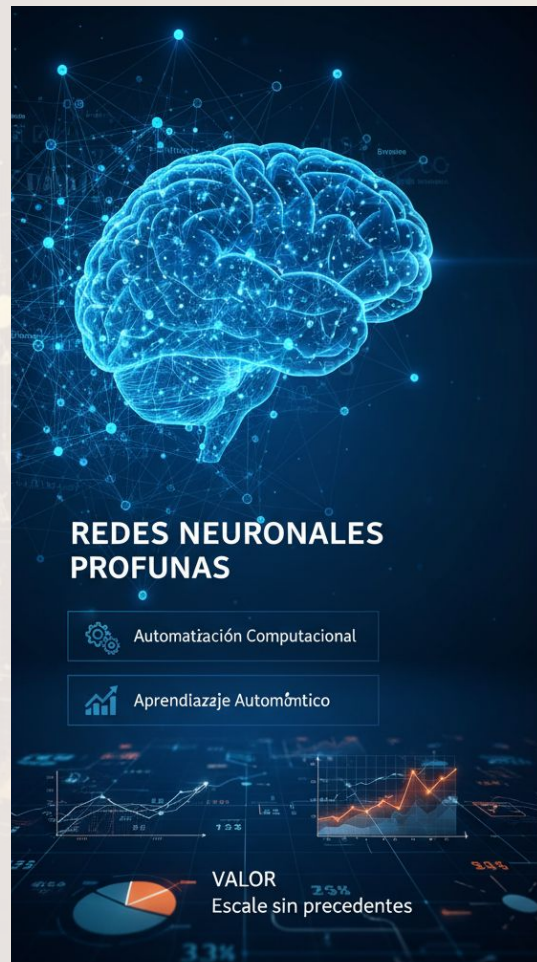
- **Automatización Computacional**

Automatizar procesos repetitivos de limpieza de datos, ingeniería de características y optimización de hiperparámetros, liberando al estadístico para otras tareas de mayor valor.

## El Valor

**Escala sin precedentes:** Procesa volúmenes masivos de datos en tiempo real.

**Detección de patrones complejos:** Identifica relaciones no lineales imposibles de detectar manualmente.



# La Convergencia de Dos Mundos

## Evolución Natural de la Inferencia

La Inteligencia Artificial no es una disciplina ajena a la estadística, sino su evolución natural. Los algoritmos de aprendizaje automático tienen sus raíces en métodos estadísticos clásicos como la regresión, clasificación bayesiana y análisis de componentes principales. La IA amplifica estas técnicas con capacidad computacional masiva y arquitecturas más complejas.

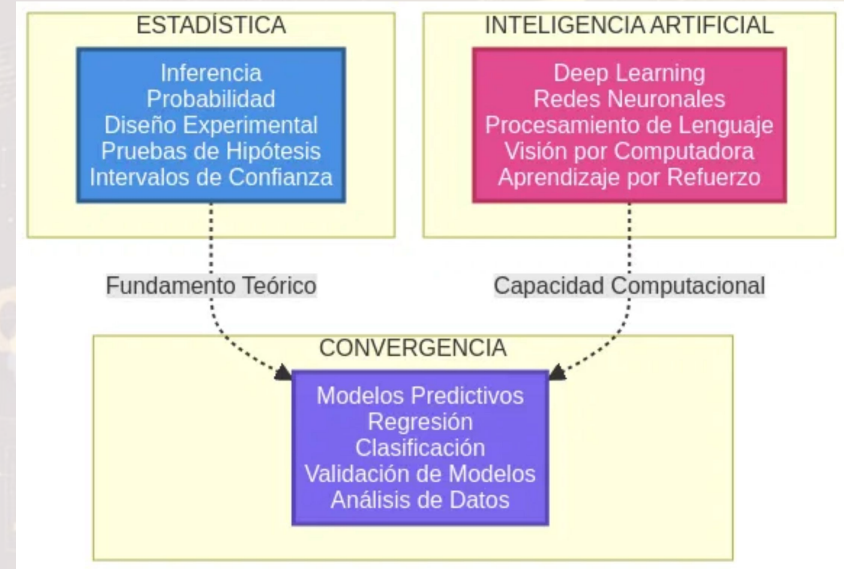
## Interpretabilidad y Rigor Científico

Los modelos estadísticos tradicionales proporcionan interpretabilidad transparente: cada coeficiente tiene un significado, cada prueba de hipótesis tiene un nivel de confianza cuantificable. Esta claridad es esencial para la validación científica, la reproducibilidad de resultados y la toma de decisiones informadas en contextos críticos como medicina, justicia y políticas públicas.

## Capacidad Predictiva y Escalabilidad

Por su parte, los modelos de IA aportan una capacidad predictiva sin precedentes. Las redes neuronales profundas pueden capturar relaciones no lineales extremadamente complejas en datos de alta dimensionalidad. La escalabilidad computacional permite procesar millones de observaciones en tiempo real, abriendo posibilidades que eran impensables hace una década.

*En la convergencia nace la Ciencia de Datos moderna: combina el pensamiento estadístico riguroso con las herramientas computacionales de IA para resolver problemas reales.*



La intersección entre Estadística e IA crea el campo del Data Science moderno



# Caso de Uso: Predicción de Riesgo en Salud Pública

Modelos de predicción de enfermedades crónicas mediante la sinergia entre Estadística e IA

## El Rol de la Estadística

### Diseño del Estudio

Define la población objetivo, determina el tamaño de muestra necesario, establece criterios de inclusión/exclusión, y diseña estrategias de muestreo que garanticen representatividad y minimicen sesgos de selección.

### Validación de Variables

Identifica factores de riesgo estadísticamente significativos mediante análisis de regresión, controla variables de confusión y establece relaciones causales usando métodos de inferencia causal como "propensity score matching".

### Cuantificación de Incertidumbre

Calcula intervalos de confianza para las predicciones, realiza pruebas de hipótesis para validar el modelo, y proporciona métricas de precisión con sus límites estadísticos claramente definidos.

## El Rol de la Inteligencia Artificial

### Procesamiento Masivo de Datos

Analiza millones de historiales clínicos electrónicos, integra datos de múltiples fuentes (laboratorio, imágenes, genómica, comportamiento), y procesa información en tiempo real para actualizar predicciones continuamente.

### Detección de Patrones Complejos

Identifica interacciones no lineales entre cientos de variables, descubre biomarcadores predictivos no evidentes y captura relaciones temporales complejas en la evolución de la enfermedad.

### Predicciones Personalizadas

Genera perfiles de riesgo individualizados para cada paciente, adapta recomendaciones según características específicas, y actualiza predicciones dinámicamente conforme nueva información está disponible.

## Resultado: Medicina de Precisión con Rigor Científico

La combinación de ambas disciplinas produce sistemas de predicción que son simultáneamente **precisos** (gracias a la IA) y **confiables** (gracias a la estadística).



# La IA como Dialecto de la Estadística (Metáfora)

Un dialecto es una variante o forma particular de una lengua, que comparte la estructura básica del idioma, pero presenta diferencias en vocabulario o expresiones.

## El Idioma Base: Estadística

Nos provee las reglas, fundamentos, estructura

## El Dialecto: IA/ML

La IA es la forma evolucionada y aplicada del mismo lenguaje

## EL Resultado

Una manera más dinámica y automatizada de “hablar” con los datos.

## Por qué esta metáfora importa

Cuando entendemos que la IA es un **dialecto** y no un idioma nuevo, dejamos de verla como amenaza y la reconocemos como **evolución natural**. No necesitamos abandonar la estadística para adoptar IA; necesitamos **expandir nuestro vocabulario** manteniendo nuestros fundamentos teóricos intactos. Un estadístico que aprende IA no está cambiando de profesión, está **actualizando su forma de expresarse** en el mundo moderno de datos.

La escala, velocidad y complejidad han  
cambiado, pero la esencia de buscar  
la verdad en los  
datos sigue siendo nuestro  
dominio.

# El Rol del Estadístico en la Era de IA

El estadístico moderno no es simplemente un analista de datos. Su rol es multidimensional para asegurar que la IA sea efectiva, confinable y ética. Es el nuevo **Arquitecto de Sistemas Inteligentes**.

## Conocimiento Clave

Cada rol requiere tanto peritaje estadístico clásico como comprensión profunda de herramientas modernas de IA. No son roles separados, sino facetas integradas de una misma profesión.

Los estadísticos son el Puente entre la capacidad computacional de la IA y la necesidad humana de **comprensión y confianza**

## Diseñador de Modelos

Selecciona variables relevantes, valida supuestos estadísticos, diseña experimentos controlados, y construye “pipelines” analíticos que combinan rigor metodológico con capacidad predictiva.

## Auditor de Algoritmos

Detecta sesgos en datos y modelos, evalúa robustez ante cambios, implementa métricas de equidad y garantiza que los algoritmos no perpetúen discriminación sistemática.

## Intérprete de Resultados

Comunica incertidumbre de manera clara, contextualiza hallazgos para audiencias no técnicas, y transforma complejidad algorítmica en conocimiento accionables mediante “storytelling” con datos.

## Guardián Ético

Protege la privacidad de datos sensibles, exige transparencia metodológica, y asegura que los sistemas de IA sirvan al bien común con responsabilidad social.

# La Pregunta Fundamental: ¿La IA Reemplaza al Estadístico?

Los datos del mercado laboral responden esta pregunta de manera contundente: **NO**. La demanda de profesionales que dominan estadística e IA está en crecimiento exponencial.

## +34%

Crecimiento proyectado en empleos de Data Scientists (2024-2034)

## \$112,590

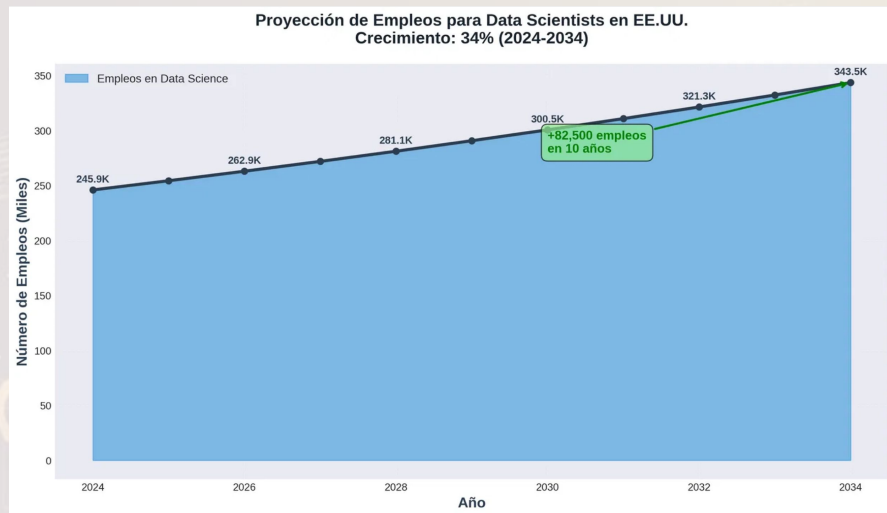
Salario mediano anual de Data Scientists en EE.UU. (2023), con percentil 90 superando \$184,090

## +82,500

Nuevas posiciones de Data Scientists proyectadas para 2034, impulsadas por adopción masiva de IA

## \$4.8 Billones

Mercado global de IA proyectado para 2033, desde \$196.6 mil millones en 2023



## Conclusión Clave

La IA **no** reemplaza al estadístico. Lo hace **más indispensable que nunca**. Cada organización que adopta IA necesita profesionales que garanticen rigor, equidad y confiabilidad.

# Más Allá del Algoritmo

El valor del estadístico no se limita a ejecutar análisis. Reside en capacidades humanas que ningún algoritmo puede replicar.

01

## Pensamiento Crítico

Cuestionar supuestos implícitos, identificar falacias lógicas, y reconocer cuándo un modelo estadísticamente significativo carece de significancia práctica. La capacidad de preguntar **"¿Tiene esto sentido?"** antes de confiar en un resultado.

02

## Comprensión Contextual

Interpretar datos dentro del contexto del dominio específico. Un algoritmo ve correlaciones; el estadístico entiende **por qué existen**, cuándo son falsos, y qué variables podrían estar distorsionando la relación.

03

## Juicio Ético

Evaluar implicaciones morales de decisiones algorítmicas. Reconocer cuándo la optimización matemática entra en conflicto con valores humanos, y **priorizar equidad** sobre precisión. Cuando la justicia social lo demanda.

## El Valor del Estadístico es Irreemplazable

Los algoritmos ejecutan instrucciones con precisión perfecta. Pero solo el estadístico puede determinar **qué preguntas vale la pena responder, qué supuestos son razonables, qué limitaciones deben comunicarse y cuándo decir "no"** a un análisis que, aunque técnicamente correcto, sería éticamente irresponsable. Este juicio humano informado es el núcleo irreemplazable de nuestra profesión.



# Métodos Estadísticos vs Aprendizaje Automático

Dos enfoques complementarios para extraer conocimiento de los datos

	MÉTODOS ESTADÍSTICOS	APRENDIZAJE AUTOMÁTICO
Objetivo Principal	<b>Modelos Explicativos:</b> Entender relaciones causales y cuantificar incertidumbre	<b>Modelos Predictivos:</b> Maximizar precisión en predicciones futuras
Volumen de Datos	<b>Datos Limitados:</b> Funciona eficientemente con muestras pequeñas bien diseñadas	<b>Datos Masivos:</b> Requiere grandes volúmenes para alcanzar su máximo potencial
Interpretabilidad	<b>Transparencia Total:</b> Cada parámetro tiene significado claro y cuantificable	<b>Caja Negra:</b> Patrones complejos difíciles de interpretar humanamente (Explicabilidad)
Tipo de Inferencia	<b>Inferencia Causal:</b> Establece relaciones de causa-efecto con control de confusores	<b>Correlaciones:</b> Identifica asociaciones sin necesariamente implicar causalidad
Cuantificación de Confianza	<b>Intervalos de Confianza:</b> Rangos probabilísticos con nivel de certeza conocido	<b>Métricas de Precisión:</b> Accuracy, F1-score, AUC-ROC en datos de validación
Complejidad de Patrones	<b>Relaciones Simples:</b> Lineales o con transformaciones específicas predefinidas	<b>Relaciones Complejas:</b> No lineales, interacciones de alto orden automáticas
Pregunta Fundamental	¿POR QUÉ?	¿QUÉ?

La estadística pregunta **POR QUÉ** , la IA predice **QUÉ** .



# Del Dato al Conocimiento

La transformación de datos en conocimiento accionable no es automática. Requiere una jerarquía de procesamiento donde cada nivel agrega valor y comprensión.

## DATO → INFORMACIÓN

Los datos brutos se limpian, estructuran y organizan. El estadístico valida la calidad, detecta valores atípicos y asegura que los datos sean confiables antes de cualquier análisis.

## INFORMACIÓN → ANÁLISIS

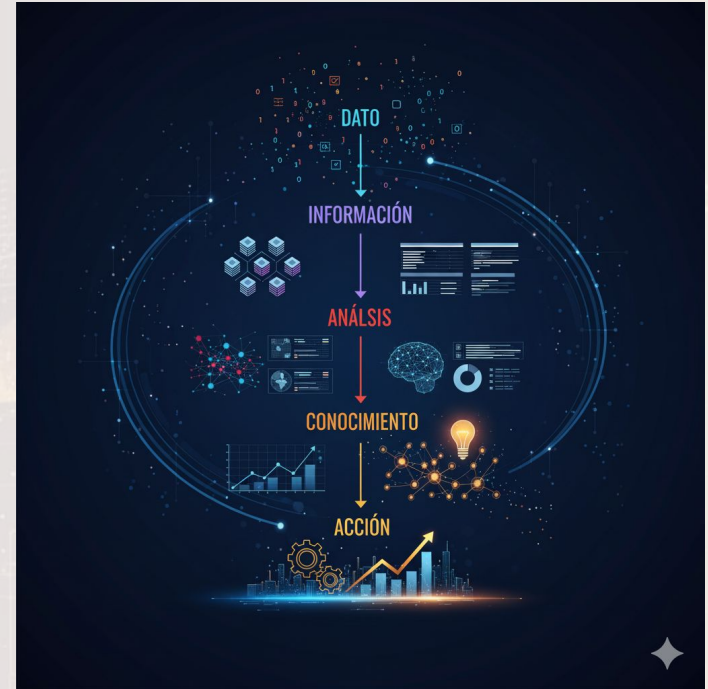
La información estructurada se somete a análisis estadístico y modelos de ML. La IA procesa grandes volúmenes rápidamente, mientras la estadística garantiza validez metodológica.

## ANÁLISIS → CONOCIMIENTO

Los resultados analíticos se interpretan y contextualizan. El estadístico cuantifica la incertidumbre, evalúa la significancia y transforma números en comprensión profunda del fenómeno estudiado.

## CONOCIMIENTO → ACCIÓN

El conocimiento validado se convierte en decisiones estratégicas y acciones concretas. Este es el impacto final: cambios reales basados en evidencia sólida y análisis riguroso.



La IA acelera el ciclo y mejora la precisión, pero el estadístico asegura que cada transacción sea válida, confiable y ética

# Caso Real: Sesgo Algorítmico en Salud

Un algoritmo ampliamente usado en hospitales de EE.UU. para identificar pacientes de alto riesgo contenía sesgo racial sistemático que perjudicaba a pacientes afroamericanos.

## El Problema Fundamental

El algoritmo era técnicamente preciso en predecir costos futuros, pero **estadísticamente inválido** como medida de necesidad de salud. Pacientes afroamericanos tienen menos acceso a atención médica debido a barreras socioeconómicas, generando menos costos históricos.

## Rol del Estadístico

Un estadístico habría identificado la **variable de sesgo** (acceso desigual) y exigido validación por grupo demográfico antes del despliegue. La IA optimizó la métrica equivocada.

## Lección Aprendida

La precisión técnica no garantiza validez estadística ni equidad ética. El estadístico es el guardián que previene que la IA automatice injusticias.

### RESEARCH

#### RESEARCH ARTICLE

##### ECONOMICS

## Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer<sup>1,2,\*</sup>, Brian Powers<sup>3</sup>, Christine Vogt<sup>4</sup>, Sendhil Mullainathan<sup>5,†</sup>

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

There is growing concern that algorithms may reproduce racial and gender disparities via the people building them or through the data used to train them (1–3). Empirical work is increasingly lending support to these concerns. For example, job search ads for highly paid positions are less likely to be presented to women (4), searches for distinctively Black-sounding names are more likely to trigger ads for arrest records (5), and image searches for professions such as CEO produce fewer images of women (6). Facial recognition systems increasingly used in law enforcement perform worse on recognizing faces of women and Black individuals (7, 8), and natural language processing algorithms encode language in gendered ways (9). Empirical investigations of algorithmic bias, though, have been hindered by a key constraint: Algorithms deployed on large scales are typically proprietary, making it difficult for independent researchers to dissect them. Instead, researchers must work “from the outside,” often with great ingenuity, and resort to clever workarounds such as audit studies. Such efforts can document disparities, but understanding how and why they arise—much less figuring out what to do about them—is difficult without greater access to the algorithms themselves. Our understanding of a mechanism therefore typically relies on theory or exercises with

researcher-created algorithms (10–12). Without an algorithm’s training data, objective function, and prediction methodology, we can only guess as to the actual mechanisms for the important algorithmic disparities that arise.

In this study, we exploit a rich dataset that provides insight into a live, scaled algorithm deployed nationwide today. It is one of the largest and most typical examples of a class of commercial risk-prediction tools that, by a model system, are applied to roughly 200 million people in the United States each year. Large health systems and payers rely on this algorithm to target patients for “high-risk care management” programs. These programs seek to improve the care of patients with complex health needs by providing additional resources, including greater attention from trained providers, to help ensure that care is well coordinated. Most health systems use these programs as the cornerstone of population health management efforts, and they are widely considered effective at improving outcomes and satisfaction while reducing costs (14–17). Because the programs are themselves expensive—with costs going toward teams of dedicated nurses, extra primary care appointment slots, and other scarce resources—health systems rely extensively on algorithms to identify patients who will benefit the most (18, 19).

Identifying patients who will derive the greatest benefit from these programs is a challenging causal inference problem that requires estimation of individual treatment effects. To solve this problem, health systems make a key assumption: Those with the greatest care needs will benefit the most from the program. Under this assumption, the targeting problem becomes a pure prediction policy problem (20). Developers then build algorithms

that rely on past data to build a predictor of future health care needs.

Our dataset describes one such typical algorithm. It contains both the algorithm’s predictions as well as the data needed to understand its inner workings: that is, the underlying ingredients used to form the algorithm (data, objective function, etc.) and links to a rich set of outcome data. Because we have the inputs, outputs, and eventual outcomes, our data allow us a rare opportunity to quantify racial disparities in algorithms and isolate the mechanisms by which they arise. It should be emphasized that this algorithm is not unique. Rather, it is emblematic of a generalized approach to risk prediction in the health sector, widely adopted by a range of for- and non-profit medical centers and governmental agencies (21).

Our analysis has implications beyond what we learn about this particular algorithm. First, the specific problem solved by this algorithm has analogies in many other sectors: The predicted risk of some future outcome (in our case, health care needs) is widely used to target policy interventions under the assumption that the treatment effect is monotonic in that risk, and the methods used to build the algorithm are standard. Mechanisms of bias uncovered in this study likely operate elsewhere. Second, even beyond our particular finding, we hope that this exercise illustrates the importance, and the large opportunity, of studying algorithmic bias in health care, not just as a model system but also in its own right. By any standard—e.g., number of lives affected, life-and-death consequences of the decision—health is one of the most important and widespread social sectors in which algorithms are already used at scale today, unbeknownst to many.

### Data and analytic strategy

Working with a large academic hospital, we identified all primary care patients enrolled in risk-based contracts from 2013 to 2015. Our primary interest was in studying differences between White and Black patients. We formed race categories by using hospital records, which are based on patient self-reporting. Any patient who identified as Black was considered to be Black for the purposes of this analysis. Of the remaining patients, those who self-identified as races other than White (e.g., Hispanic) were so considered (data on these patients are presented in table S1 and fig. S1 in the supplementary materials). We considered all remaining patients to be White. This approach allowed us to study one particular racial difference of social and historical interest between patients who self-identified as Black and patients who self-identified as White without another race or ethnicity; it has the disadvantage of not allowing for the study of intersectional racial

<sup>1</sup>School of Public Health, University of California, Berkeley, Berkeley, CA, USA. <sup>2</sup>Department of Emergency Medicine, Brigham and Women’s Hospital, Boston, MA, USA. <sup>3</sup>Department of Medicine, Brigham and Women’s Hospital, Boston, MA, USA. <sup>4</sup>Morgan Institute Health Policy Center, Massachusetts General Hospital, Boston, MA, USA. <sup>5</sup>Booth School of Business, University of Chicago, Chicago, IL, USA. <sup>\*</sup>These authors contributed equally to this work. <sup>†</sup>Corresponding author. Email: sendhil.mullainathan@chicagobooth.edu

# Estadísticos: Guardianes de la Verdad Algorítmica

En la era de la IA, el estadístico asume una responsabilidad crítica: ser el guardián que asegura que los algoritmos sirvan a la verdad, no a la conveniencia.

## 01

### Guardián de la Validez

Verificar que los modelos cumplan supuestos estadísticos, que las muestras sean representativas, y que las conclusiones estén respaldadas por evidencia sólida. **Decir "no"** cuando un análisis es técnicamente correcto pero estadísticamente engañoso.

## 02

### Guardián de la Equidad

Auditar modelos para detectar sesgos discriminatorios, exigir validación por grupo demográfico, y asegurar que los sistemas de IA **no perpetúen injusticias** sistémicas. La precisión sin equidad es opresión automatizada.

## 03

### Guardián de la Transparencia

Comunicar limitaciones honestamente, cuantificar incertidumbre y traducir resultados técnicos en lenguaje accesible. **Resistir la tentación** de presentar predicciones probabilísticas como certezas absolutas.

## Nuestro Compromiso Profesional

Cada vez que validamos un modelo, auditamos un algoritmo o comunicamos resultados, estamos ejerciendo un acto de **responsabilidad social**. No somos simplemente técnicos que ejecutan análisis; somos **profesionales éticos** que garantizan que la revolución de la IA beneficie a toda la humanidad, no solo a quienes tienen poder. Este es nuestro llamado, nuestra misión y nuestro legado.



# Desafíos Éticos en IA

La adopción masiva de IA plantea desafíos éticos complejos que requieren la expertise del estadístico para identificar, cuantificar y mitigar.

Los desafíos éticos no son problemas técnicos con soluciones algorítmicas. Son dilemas que requieren juicio humano informado por rigor estadístico.

## 1. Sesgo y Discriminación Algorítmica

Modelos que perpetúan o amplifican sesgos históricos contra grupos marginados. Requiere auditoría estadística rigurosa por grupo demográfico y análisis de equidad.

## 2. Privacidad y Protección de Datos

Balance entre utilidad de datos y derechos individuales. Técnicas como privacidad diferencial requieren comprensión estadística profunda de trade-offs entre precisión y anonimización.

## 3. Transparencia e Interpretabilidad

Sistemas "caja negra" que toman decisiones sin explicación. El estadístico debe exigir modelos interpretables o desarrollar métodos "post-hoc" para explicar predicciones.

## 4. Responsabilidad y Rendición de Cuentas

¿Quién es responsable cuando un algoritmo causa daño? El estadístico debe documentar supuestos, limitaciones, y validaciones para establecer cadenas de responsabilidad claras.

## 5. Impacto Social y Desplazamiento Laboral

Automatización que elimina empleos sin crear alternativas. Requiere análisis estadístico de impactos por sector, región y nivel educativo.

## 6. Uso Dual y Aplicaciones Maliciosas

Tecnologías que pueden usarse para vigilancia masiva o manipulación. El estadístico debe considerar implicaciones éticas antes de participar en proyectos cuestionables.

# Herramientas del Estadístico Moderno

El arsenal tecnológico para trabajar en la era de la Inteligencia Artificial

## Lenguajes de Programación

- **Python**

Ecosistema completo para ML y análisis de datos

- **R**

Análisis estadístico avanzado y visualización

- **SQL**

Manipulación de bases de datos relacionales

## Frameworks de ML

- **scikit-learn**

Algoritmos clásicos de ML

- **TensorFlow / PyTorch**

Deep learning y redes neuronales

- **XGBoost / LightGBM**

Gradient boosting de alto rendimiento

## Análisis y Visualización

- **Pandas / NumPy**

Manipulación y análisis de datos

- **Matplotlib / Seaborn**

Visualización estadística

- **Tableau / Power BI**

Dashboards interactivos

## Big Data y Cloud

- **Spark / PySpark**

Procesamiento distribuido de datos masivos

- **AWS / Azure / GCP**

Infraestructura cloud para ML

- **Docker / Kubernetes**

Orquestación

## Control de Versiones

- **Git / GitHub**

Versionado de código y colaboración

- **MLflow / DVC**

Versiones de modelos y datos

## Estadística Clásica

- **SAS / SPSS / Stata**

Software estadístico tradicional

- **JAGS / Stan**

Inferencia bayesiana

## Notebooks y IDEs

- **Jupyter / JupyterLab**

Análisis interactivo y documentación

- **VS Code / PyCharm / RStudio**

Entornos de desarrollo integrados

# Herramientas de IA

## Automatización

- Julius
- Coefficient

## Ampliación y Generación de Datos

- Mostly.AI
- COPILOT

## Preparación de Datos

- CHATCSV
- TOMAT AI

## Data Insights

- Data Chat
- PowerDrill

## Visualización de Datos

- ChartGPT
- GraphMaker
- HighCharts

## Storytelling

- DataCamp
- Secoda



# El Nuevo Perfil del Estadístico

Competencias esenciales para liderar en la era de la Inteligencia Artificial

## 1. Competencias Técnicas

- Programación (Python, R, SQL)
- Machine Learning y Deep Learning
- Big Data y computación distribuida
- Cloud computing (AWS, Azure, GCP)
- Control de versiones (Git)
- Bases de datos relacionales y NoSQL
- APIs y servicios web

## 2. Competencias Estadísticas

- Diseño experimental y muestreo
- Inferencia causal y análisis de variables confusoras
- Modelos lineales generalizados y mixtos
- Análisis multivariado y reducción de dimensionalidad
- Series temporales y pronósticos
- Inferencia bayesiana
- Análisis de supervivencia y datos censurados

## 3. Competencias de Comunicación

- Storytelling con datos
- Visualización efectiva de información compleja
- Traducción de resultados técnicos a lenguaje ejecutivo
- Presentaciones persuasivas
- Documentación clara de metodologías
- Colaboración interdisciplinaria
- Mentoría y liderazgo de equipos

## 4. Competencias Éticas y de Negocio

- Auditoría de sesgos algorítmicos
- Privacidad y protección de datos (GDPR, CCPA)
- Transparencia e interpretabilidad de modelos
- Comprensión del contexto de negocio
- Evaluación de ROI de proyectos de IA
- Gestión de stakeholders
- Pensamiento crítico y cuestionamiento de supuestos



# Puerto Rico en la Era de IA y Estadísticas

Posicionamiento estratégico de la isla en la revolución de datos

## Nuestra Posición Única

Puerto Rico cuenta con una **comunidad académica y profesional robusta en estadística**, instituciones de investigación de clase mundial, y una posición geográfica estratégica que nos permite **liderar la intersección de IA y estadística en el Caribe y Latinoamérica**. Tenemos el talento, la infraestructura educativa, y la necesidad urgente de soluciones basadas en datos para enfrentar desafíos complejos en salud, economía, y resiliencia ante desastres.

## OPORTUNIDADES

### Sector Salud

Análisis de datos de salud pública para prevención de enfermedades crónicas, optimización de recursos hospitalarios, y sistemas de alerta temprana para epidemias. Puerto Rico puede ser laboratorio de innovación en salud basada en datos.

### Sector Financiero y Empresarial

Detección de fraude, análisis de riesgo crediticio, optimización de cadenas de suministro, y modelos predictivos para decisiones de inversión. Oportunidad de atraer empresas de fintech y analytics.

### Resiliencia y Gestión de Desastres

Modelos predictivos de huracanes, optimización de distribución de recursos en emergencias, y análisis de vulnerabilidad de infraestructura crítica usando datos geospaciales y sensores IoT.

## DESAFÍOS

### Fuga de Talento

Profesionales altamente capacitados emigran por mejores oportunidades. Necesitamos crear ecosistema local atractivo con salarios competitivos, proyectos desafiantes, y oportunidades de crecimiento profesional.

### Cultura de Datos Limitada

Muchas organizaciones aún no adoptan decisiones basadas en datos. Necesitamos evangelizar la cultura de analytics y demostrar ROI tangible de proyectos de IA y estadística.

### Colaboración Academia-Industria

Desconexión entre investigación académica y necesidades del sector privado. Fortalecer puentes mediante proyectos colaborativos, pasantías, y transferencia de conocimiento bidireccional.

# Tres Ideas Finales

Lo esencial que deben recordar de esta presentación

1

La IA No Reemplaza al Estadístico, Lo Hace Indispensable

2

El Estadístico es el Guardián de la Verdad Algorítmica

3

El Futuro Pertenece a los Estadísticos que Abrazan la IA





**“Sin estadística no hay IA; sin IA, la estadística no trasciende.”**

*- Dr. Carlos Ortiz*



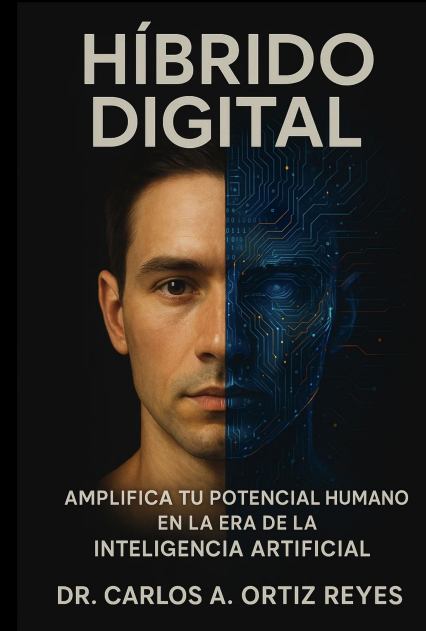
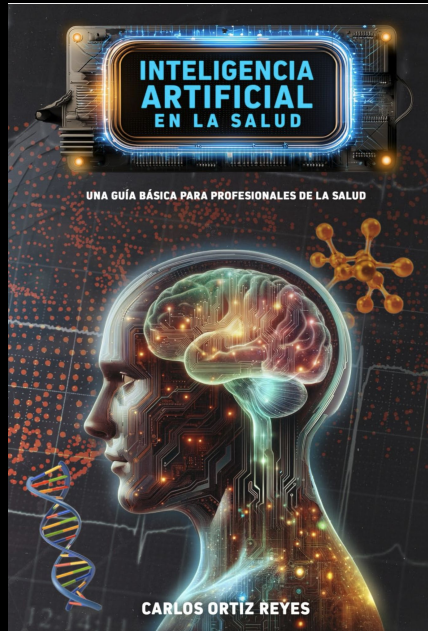
**La estadística siempre ha sido el arte de navegar la incertidumbre.  
La IA nos ha dado un barco más rápido y un mapa más detallado.**

**Pero el capitán, el que elige el destino, el que entiende las corrientes  
y tiene la responsabilidad de la tripulación, sigue siendo el estadístico.**

**¡Gracias por ser parte de esta transformación!**



# Publicaciones: Inteligencia Artificial (IA)



Dr. Carlos A. Ortiz Reyes  
Disponibles en Amazon

# Muchas Gracias

Dr. Carlos Ortiz Reyes

Catedrático Asociado

UPR-Recinto de Ciencias Médicas



[carlos.ortiz33@upr.edu](mailto:carlos.ortiz33@upr.edu)



(787) 685-6323

# Referencias

- [1] **U.S. Bureau of Labor Statistics.** (2024). *Occupational Outlook Handbook: Data Scientists*. Retrieved from <https://www.bls.gov/ooh/math/data-scientists.htm>
- [2] **Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S.** (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453.
- [3] **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*(2nd ed.). Springer.
- [4] **James, G., Witten, D., Hastie, T., & Tibshirani, R.** (2021). *An Introduction to Statistical Learning with Applications in R*(2nd ed.). Springer.
- [5] **Breiman, L.** (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199-231.
- [6] **Pearl, J., & Mackenzie, D.** (2018). *The Book of Why: The New Science of Cause and Effect* Basic Books.
- [7] **Gelman, A., & Hill, J.** (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- [8] **Goodfellow, I., Bengio, Y., & Courville, A.** (2016). *Deep Learning*. MIT Press.
- [9] **Dwork, C., & Roth, A.** (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4), 211-407.
- [10] **Molnar, C.** (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). Retrieved from <https://christophm.github.io/interpretable-ml-book/>