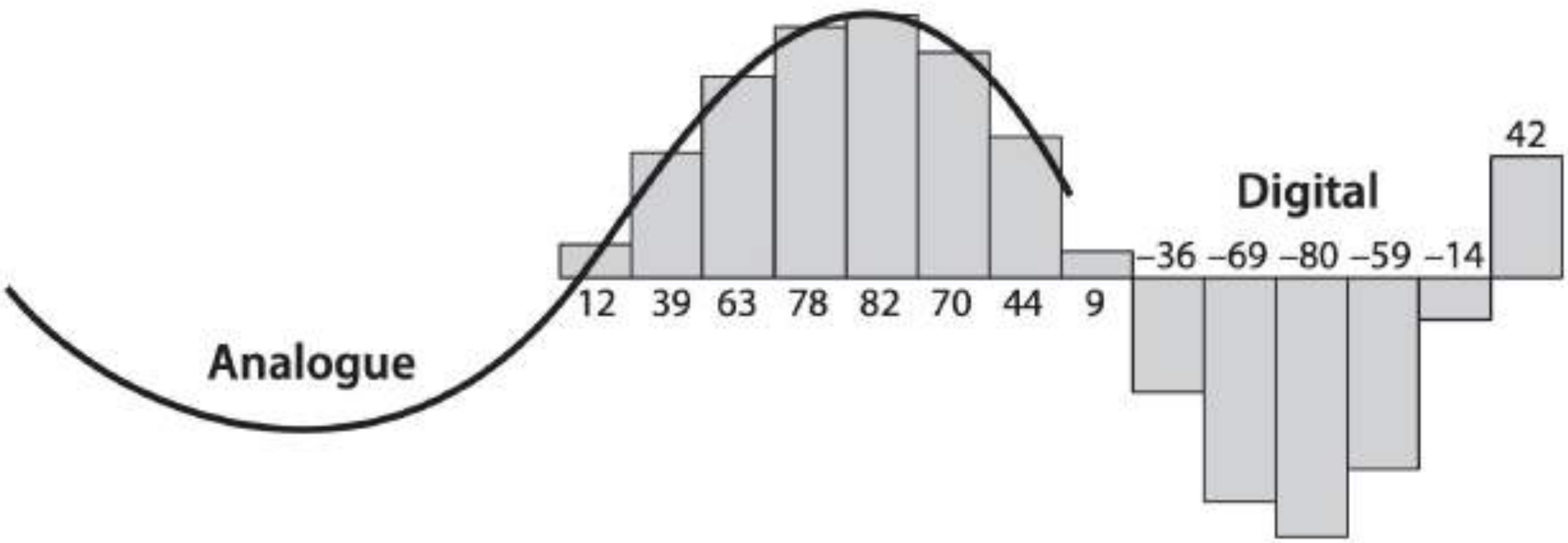
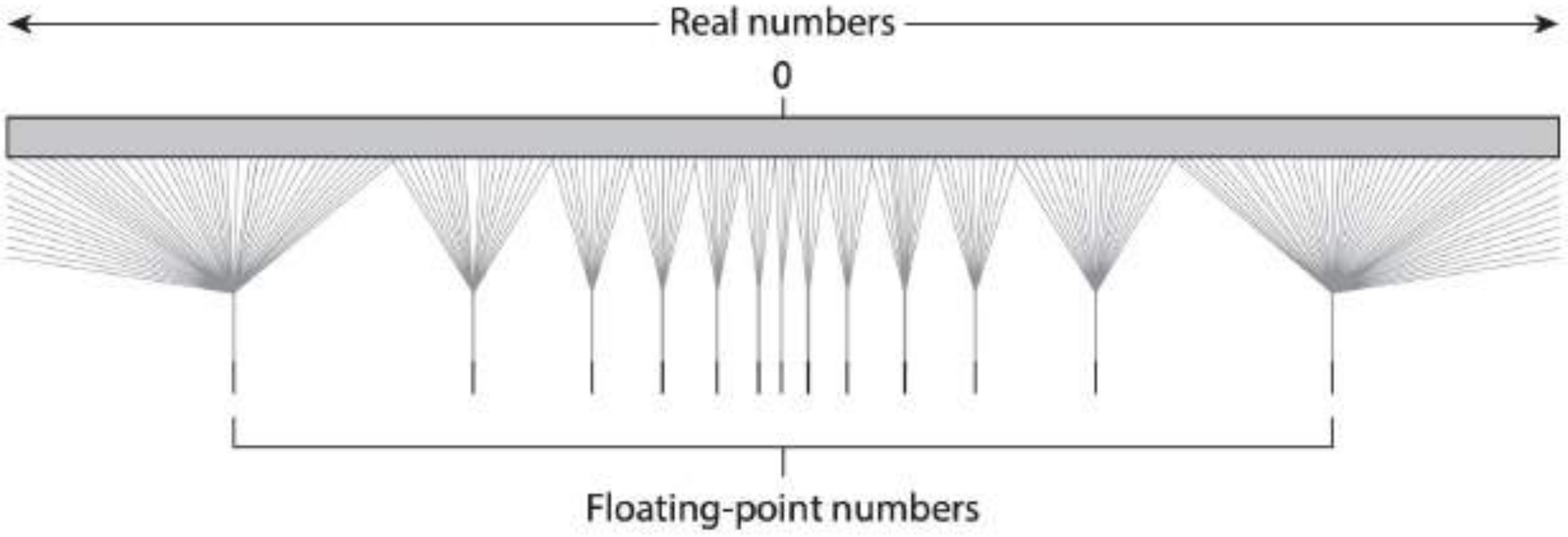
# Computing for Medicine

Monsoon 2025
Lecture 2

# Is Computation Ready for Capturing Human Health?



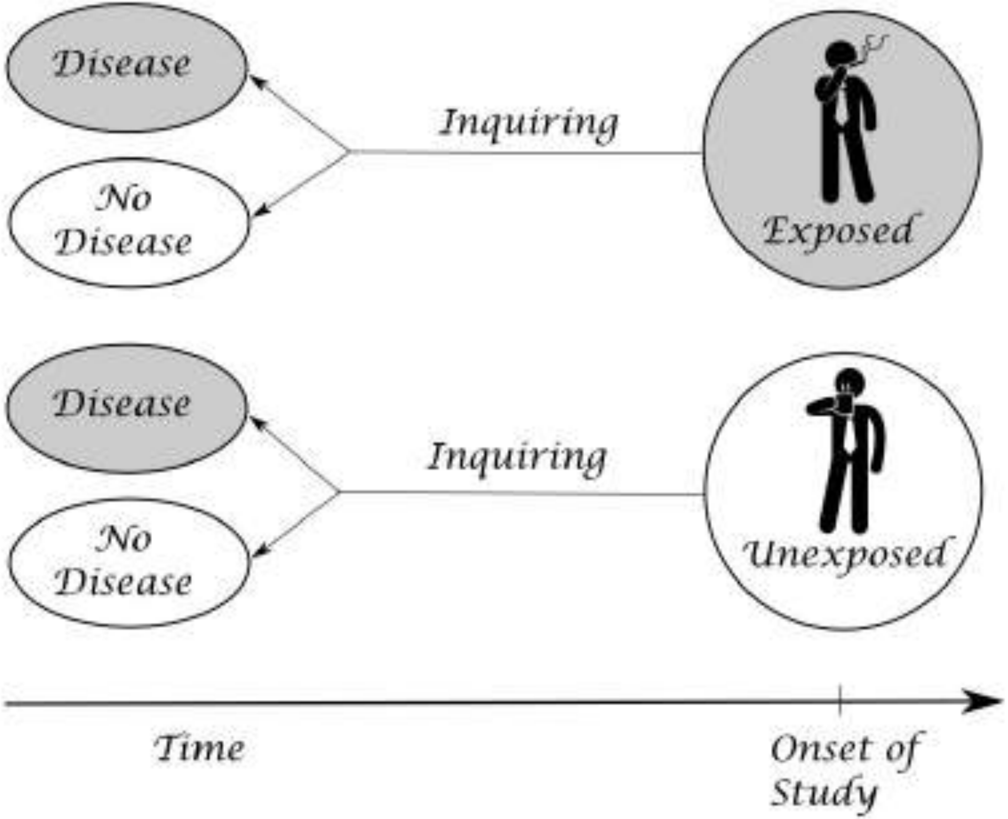
# Is Computation Ready for Capturing Human Health?



# How? Data Science Approaches

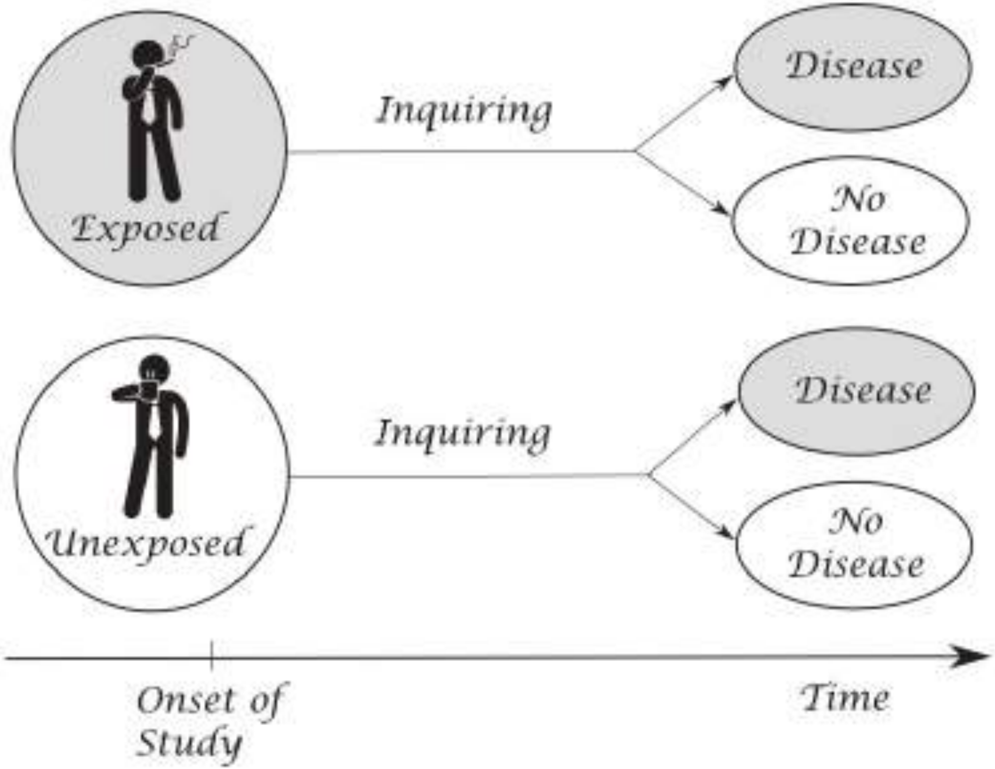
# How were the Data Generated?

## Example of Retrospective Study Design



# How were the Data Generated?

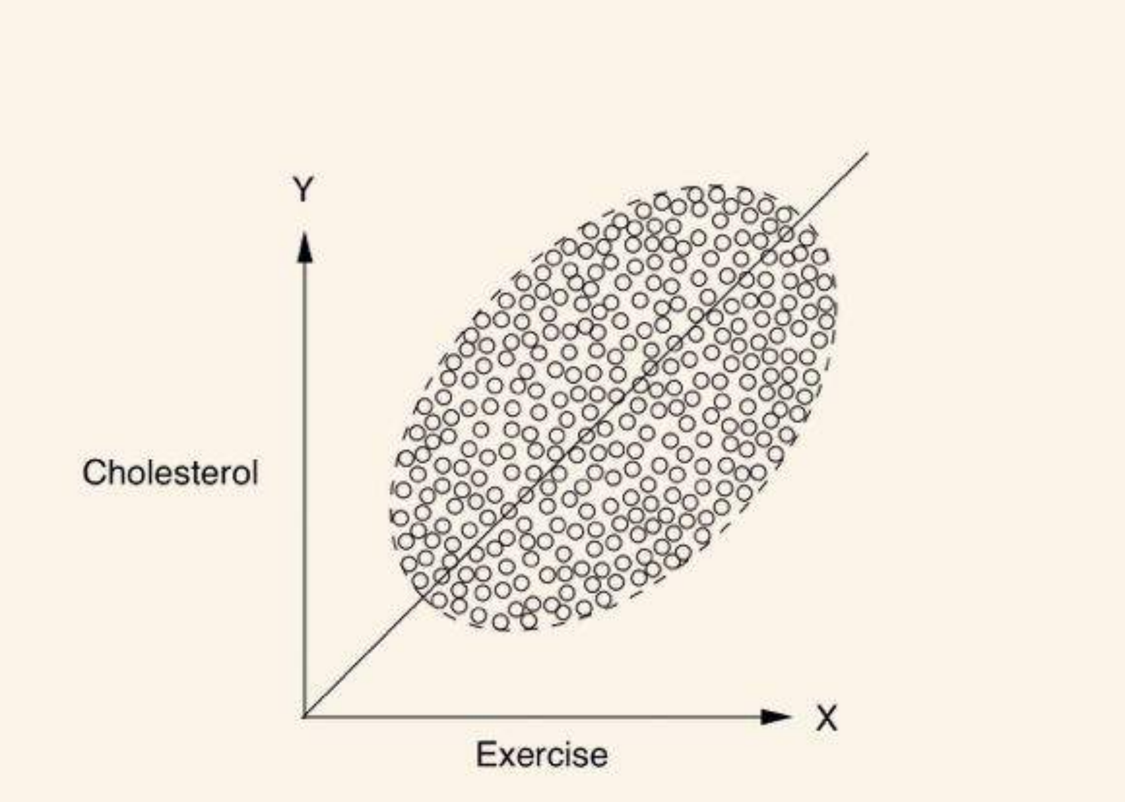
## Example of Prospective Study Design



# How? Need for Novel Data Science Approaches

You are a health policy maker:

How do you interpret this data?



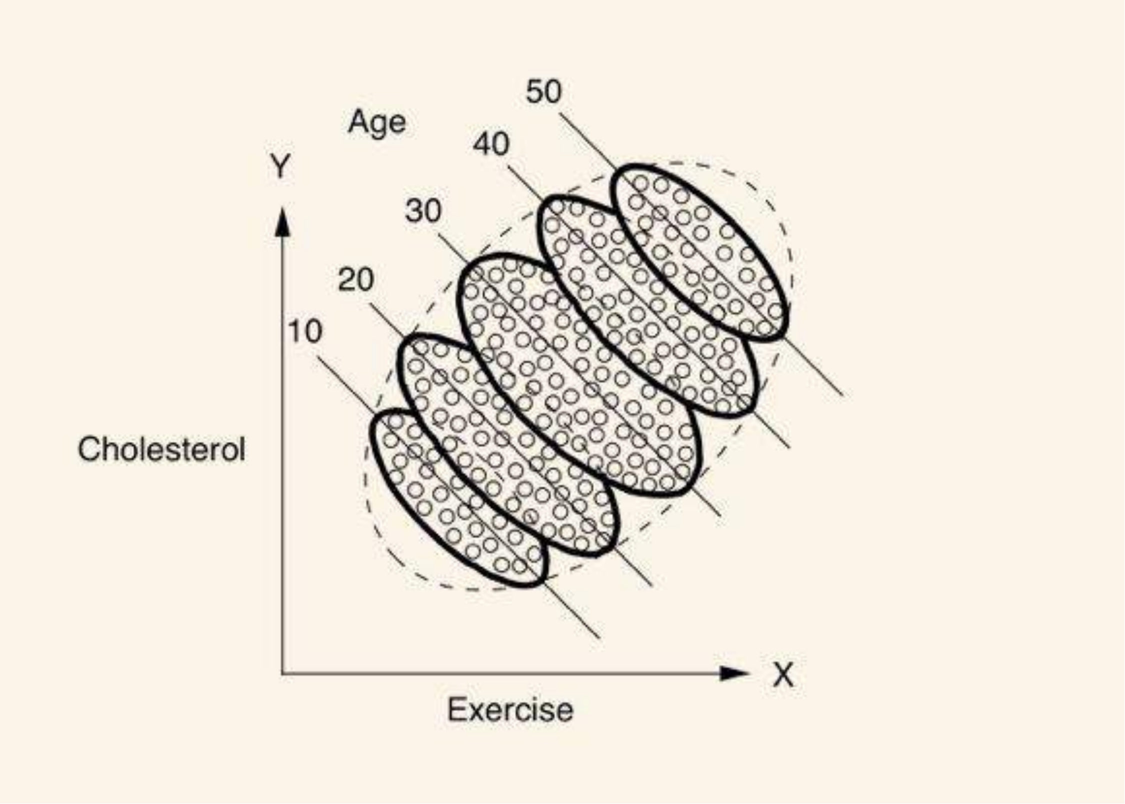
http://bayes.cs.ucla.edu/PRIMER/

# How? Need for Novel Data Science Approaches

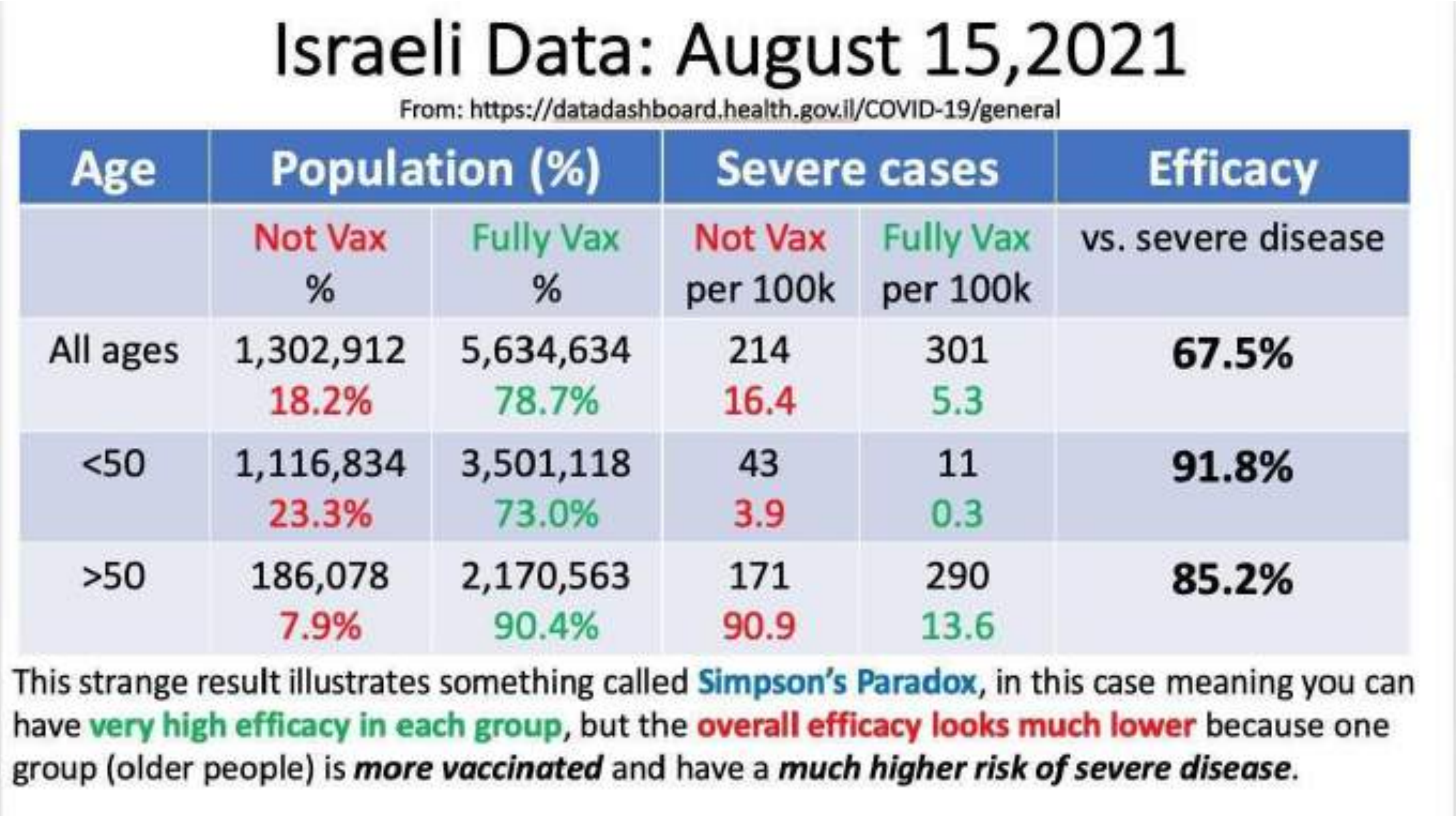
Findings on sub-groups of data may be opposite to that on the whole data!

Q: What is this phenomenon called?

**Simpson's Paradox**



http://bayes.cs.ucla.edu/PRIMER/



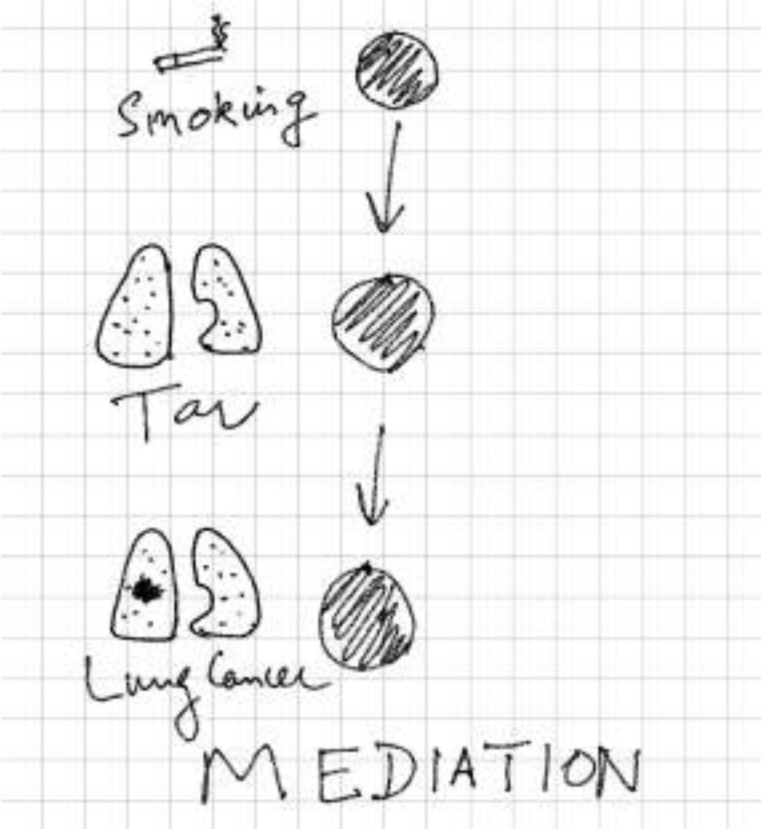
# Israeli Data: August 15,2021

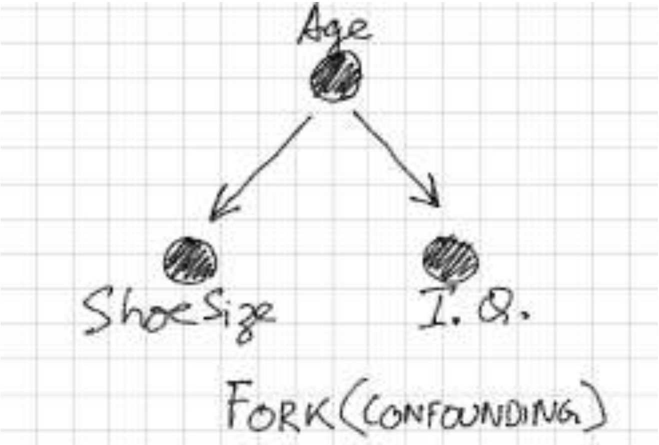
From: https://datadashboard.health.gov.il/COVID-19/general

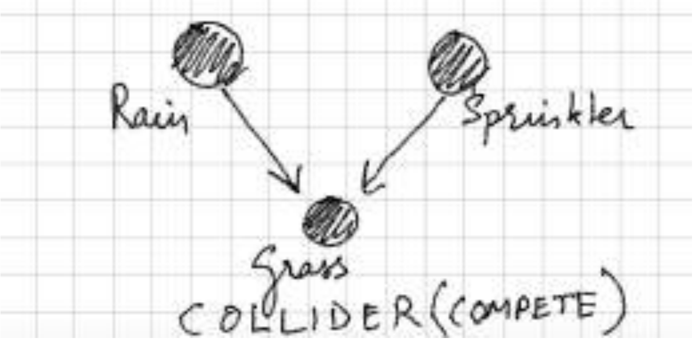
| Age | Population (%) | | Severe cases | | Efficacy |
|---|---|---|---|---|---|
| | Not Vax<br>% | Fully Vax<br>% | Not Vax<br>per 100k | Fully Vax<br>per 100k | vs. severe disease |
| All ages | 1,302,912<br>18.2% | 5,634,634<br>78.7% | 214<br>16.4 | 301<br>5.3 | **67.5%** |
| <50 | 1,116,834<br>23.3% | 3,501,118<br>73.0% | 43<br>3.9 | 11<br>0.3 | **91.8%** |
| >50 | 186,078<br>7.9% | 2,170,563<br>90.4% | 171<br>90.9 | 290<br>13.6 | **85.2%** |

This strange result illustrates something called **Simpson's Paradox**, in this case meaning you can have **very high efficacy in each group**, but the **overall efficacy looks much lower** because one group (older people) is ***more vaccinated*** and have a ***much higher risk of severe disease***.

# World of Conditional Probabilities

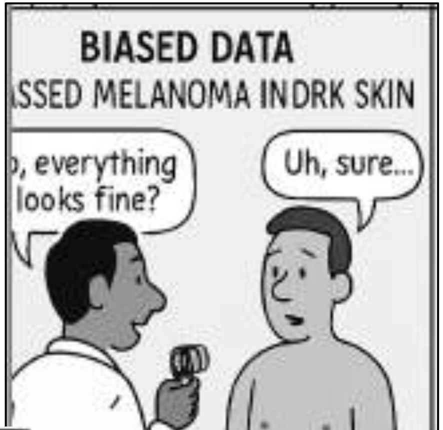




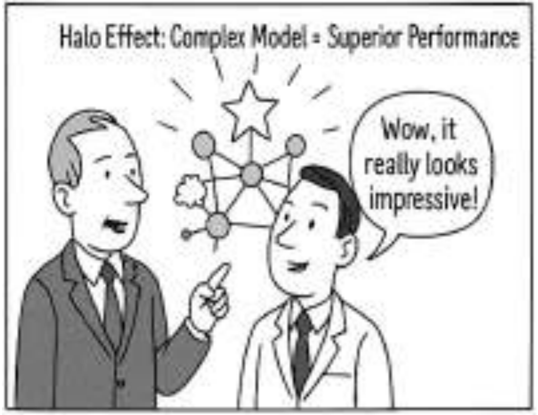


# AI Failure Can Be of Many Types

EPISTEMIC FAILURE
CAT OR COVID?

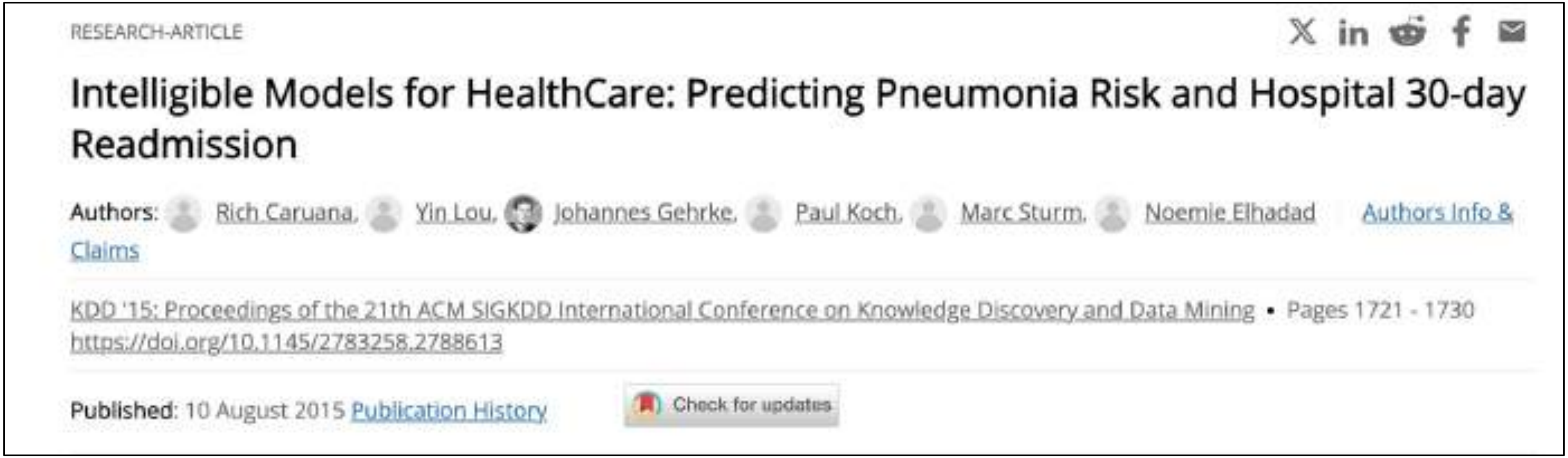


FAILURE DUE TO
REGULATORY FOG
INTERVENTION LEARNING
REGULATION
INTERVENTION LEARNING
POLICY


Halo Effect: Complex Model = Superior Performance
Wow, it really looks impressive!

HUMAN FAILURES
DISINFORMATION WITH AI
BELIEVE THIS!

# Examples of AI in Healthcare Fails



RESEARCH-ARTICLE

Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission

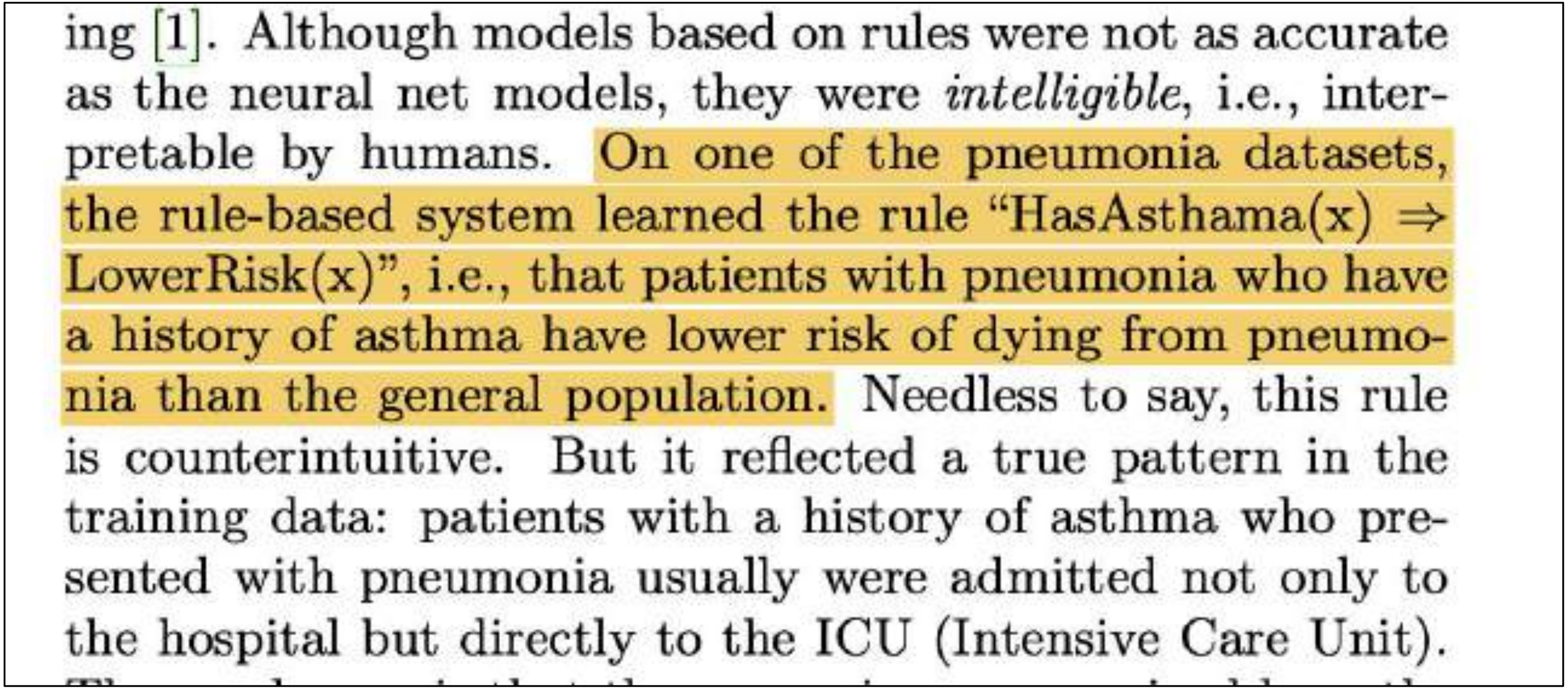
Authors: Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, Noemie Elhadad Authors Info & Claims

KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining • Pages 1721 - 1730
https://doi.org/10.1145/2783258.2788613

Published: 10 August 2015 Publication History

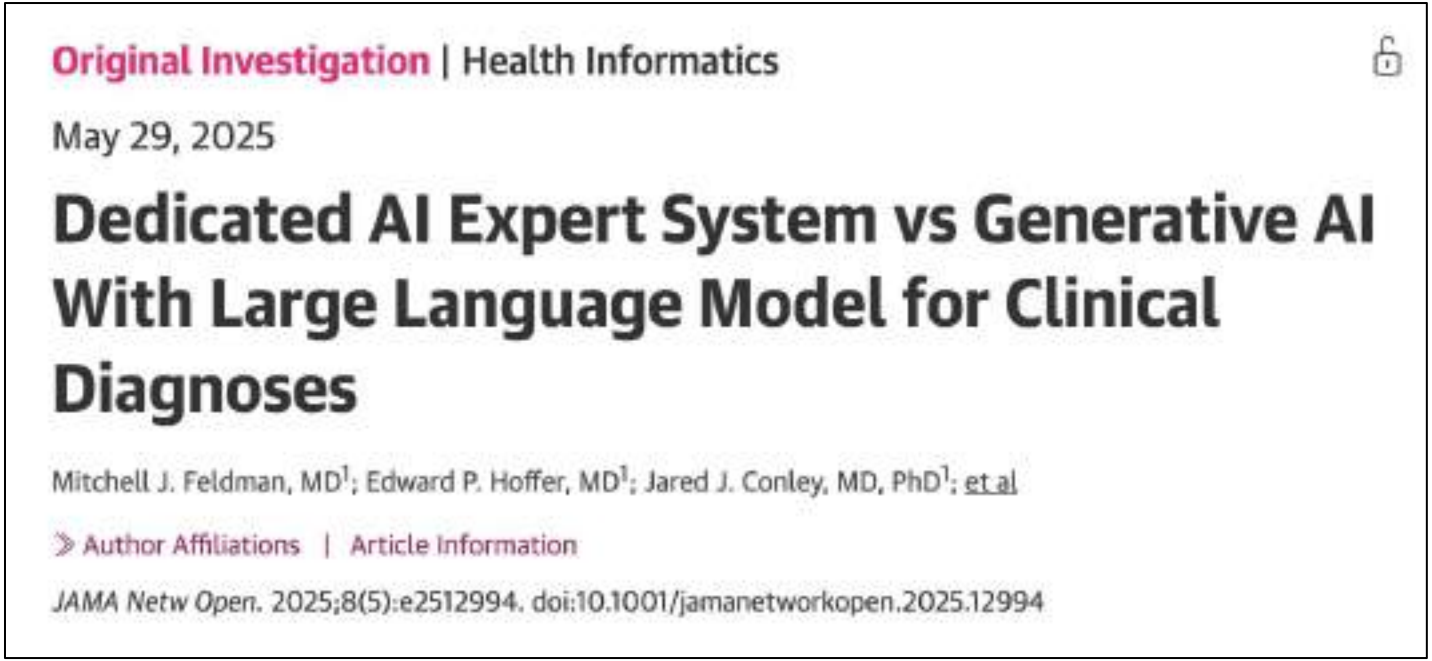
Caruana, R., et al. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721–1730)*. Association for Computing Machinery.

# Confounding



ing [1]. Although models based on rules were not as accurate as the neural net models, they were *intelligible*, i.e., interpretable by humans. On one of the pneumonia datasets, the rule-based system learned the rule “HasAsthama(x) ⇒ LowerRisk(x)”, i.e., that patients with pneumonia who have a history of asthma have lower risk of dying from pneumonia than the general population. Needless to say, this rule is counterintuitive. But it reflected a true pattern in the training data: patients with a history of asthma who presented with pneumonia usually were admitted not only to the hospital but directly to the ICU (Intensive Care Unit).

# Halo Effect of Deep Learning



Original Investigation | Health Informatics

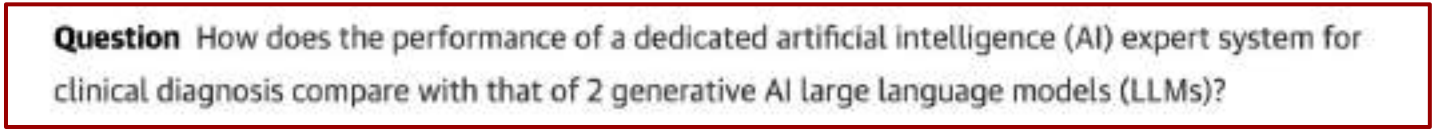
May 29, 2025

**Dedicated AI Expert System vs Generative AI With Large Language Model for Clinical Diagnoses**

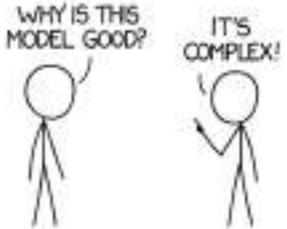
Mitchell J. Feldman, MD[1]; Edward P. Hoffer, MD[1]; Jared J. Conley, MD, PhD[1]; et al

Author Affiliations | Article Information

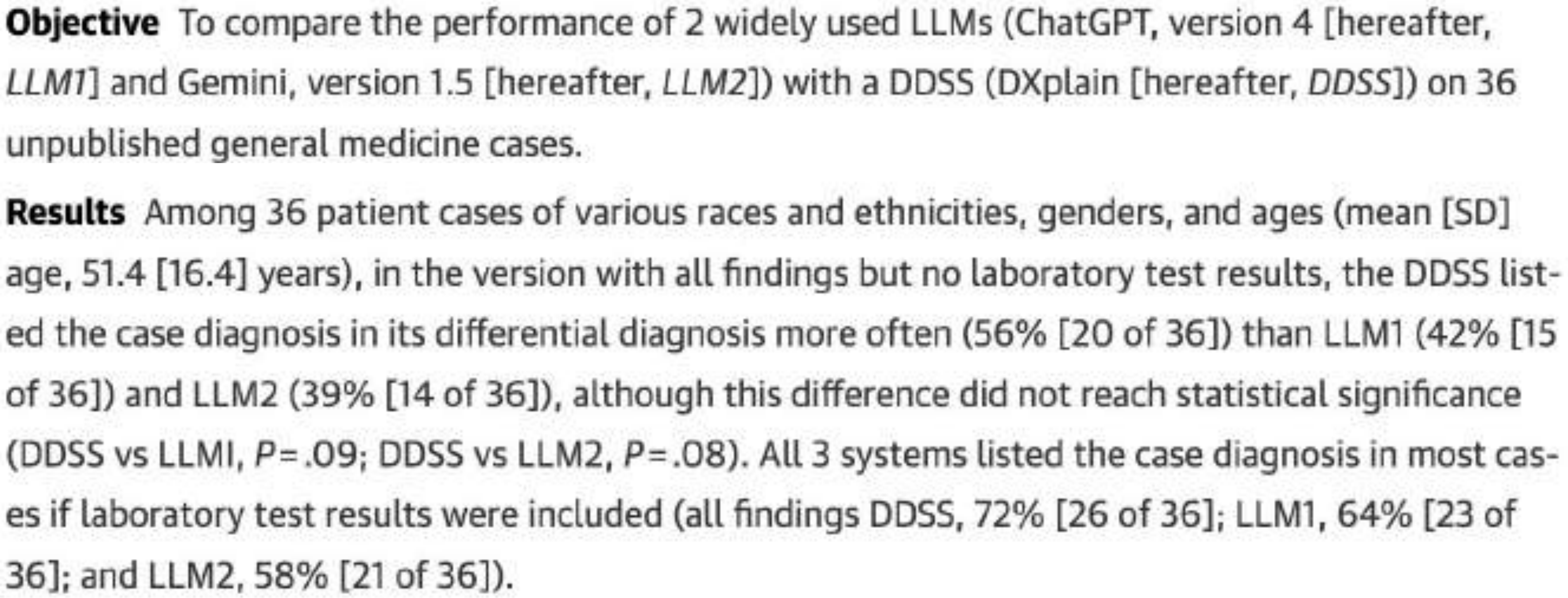
*JAMA Netw Open.* 2025;8(5):e2512994. doi:10.1001/jamanetworkopen.2025.12994



**Question** How does the performance of a dedicated artificial intelligence (AI) expert system for clinical diagnosis compare with that of 2 generative AI large language models (LLMs)?



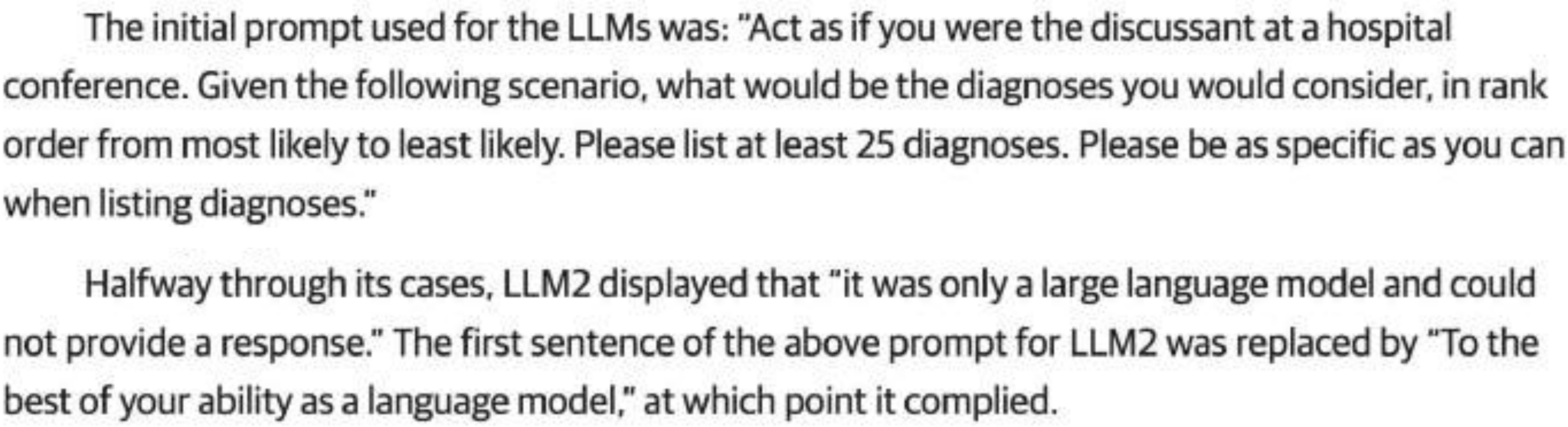
# Results



**Objective** To compare the performance of 2 widely used LLMs (ChatGPT, version 4 [hereafter, *LLM1*] and Gemini, version 1.5 [hereafter, *LLM2*]) with a DDSS (DXplain [hereafter, *DDSS*]) on 36 unpublished general medicine cases.

**Results** Among 36 patient cases of various races and ethnicities, genders, and ages (mean [SD] age, 51.4 [16.4] years), in the version with all findings but no laboratory test results, the DDSS listed the case diagnosis in its differential diagnosis more often (56% [20 of 36]) than LLM1 (42% [15 of 36]) and LLM2 (39% [14 of 36]), although this difference did not reach statistical significance (DDSS vs LLMI, *P* = .09; DDSS vs LLM2, *P* = .08). All 3 systems listed the case diagnosis in most cases if laboratory test results were included (all findings DDSS, 72% [26 of 36]; LLM1, 64% [23 of 36]; and LLM2, 58% [21 of 36]).

# Prompt Used



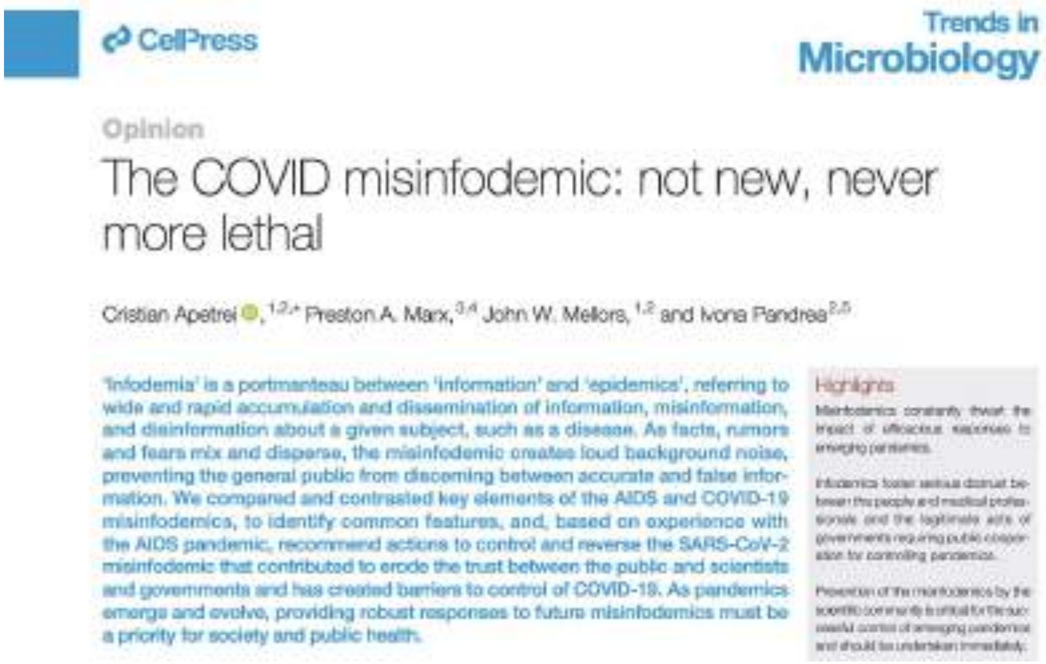
The initial prompt used for the LLMs was: "Act as if you were the discussant at a hospital conference. Given the following scenario, what would be the diagnoses you would consider, in rank order from most likely to least likely. Please list at least 25 diagnoses. Please be as specific as you can when listing diagnoses."

Halfway through its cases, LLM2 displayed that "it was only a large language model and could not provide a response." The first sentence of the above prompt for LLM2 was replaced by "To the best of your ability as a language model," at which point it complied.

"Amid all the interest in large language models, it's easy to forget that the first AI systems used successfully in medicine were expert systems like DXplain" -Mitchell Feldman (author)

BUT, DXplain started in 1980s and has had the time to evolve, yet not significantly better!

# Safety of AI



CellPress

Trends in Microbiology

Opinion

The COVID misinfodemic: not new, never more lethal

Cristian Apetrei,[1,2,*] Preston A. Marx,[3,4] John W. Mellors,[1,2] and Ivona Pandrea[2,5]

'Infodemia' is a portmanteau between 'information' and 'epidemics', referring to wide and rapid accumulation and dissemination of information, misinformation, and disinformation about a given subject, such as a disease. As facts, rumors and fears mix and disperse, the misinfodemic creates loud background noise, preventing the general public from discerning between accurate and false information. We compared and contrasted key elements of the AIDS and COVID-19 misinfodemics, to identify common features, and, based on experience with the AIDS pandemic, recommend actions to control and reverse the SARS-CoV-2 misinfodemic that contributed to erode the trust between the public and scientists and governments and has created barriers to control of COVID-19. As pandemics emerge and evolve, providing robust responses to future misinfodemics must be a priority for society and public health.

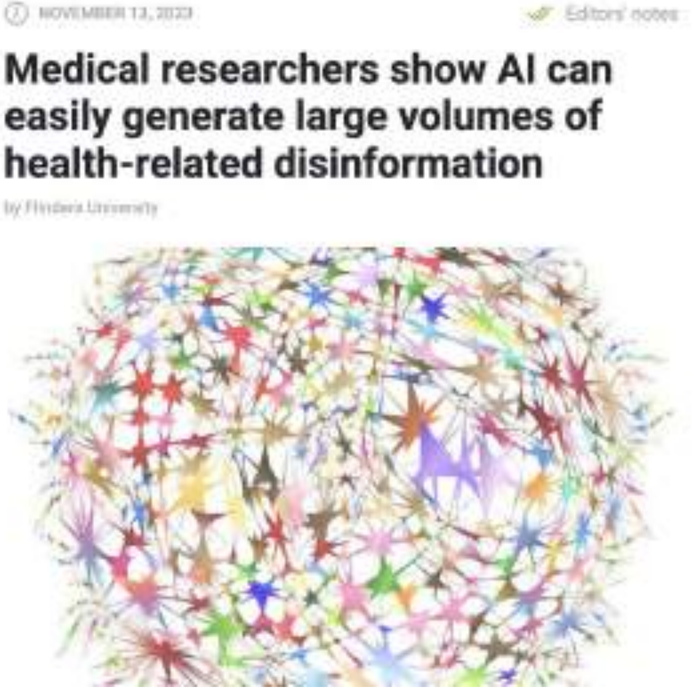
Highlights

Misinfodemics constantly thwart the impact of efficacious responses to emerging pandemics.

Infodemics foster serious distrust between the people and medical professionals and the legitimate acts of governments requiring public cooperation for controlling pandemics.

Prevention of the misinfodemics by the scientific community is critical for the successful control of emerging pandemics and should be undertaken immediately.

The COVID misinfodemic: not new, never more lethal. Apetrei, Cristian et al. Trends in Microbiology, Volume 30, Issue 10, 948 - 958



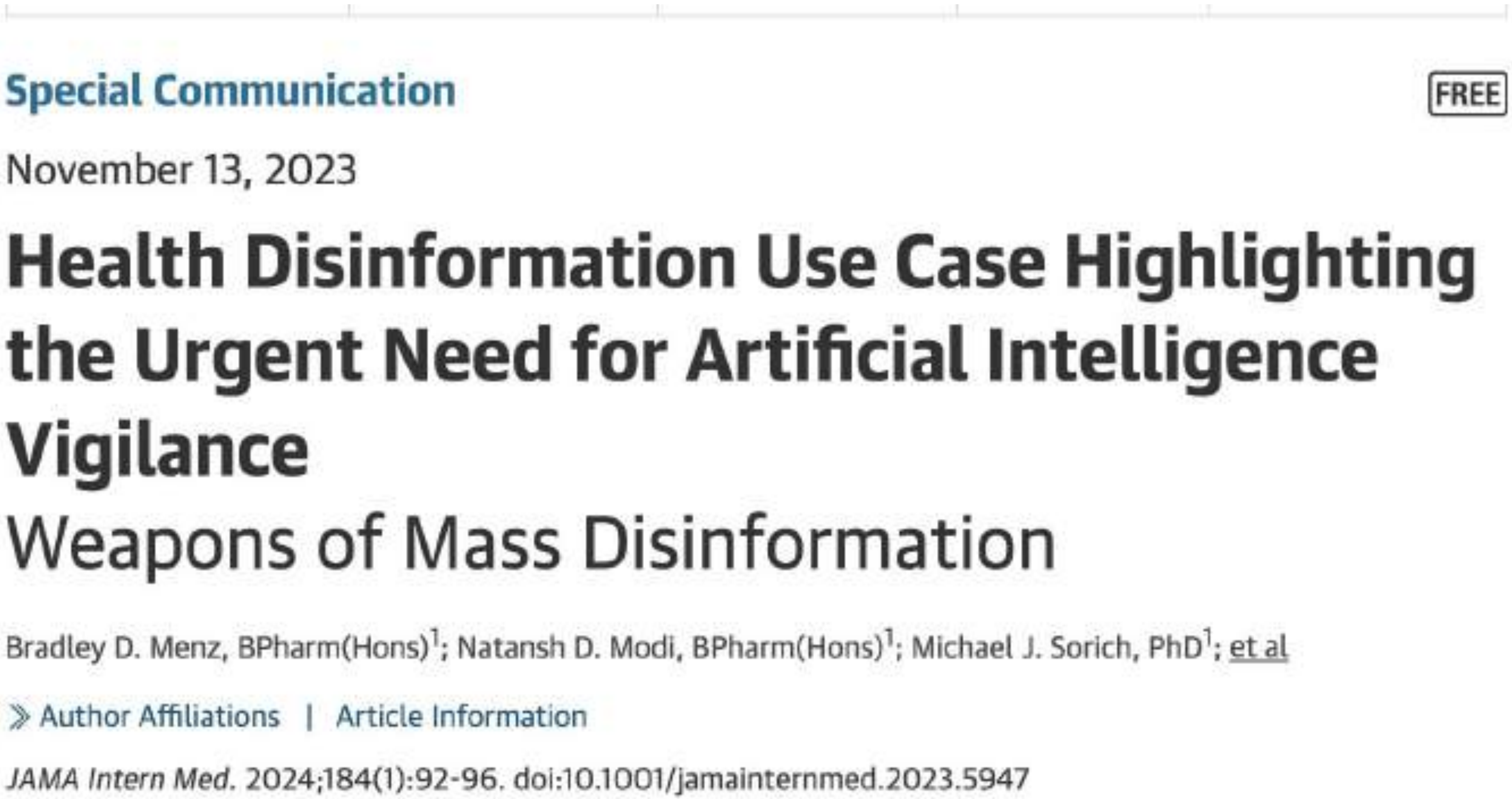
NOVEMBER 13, 2023

Editors' notes

Medical researchers show AI can easily generate large volumes of health-related disinformation

by Flinders University

https://medicalxpress.com/news/2023-11-medical-ai-easily-generate-large.html



Special Communication

FREE

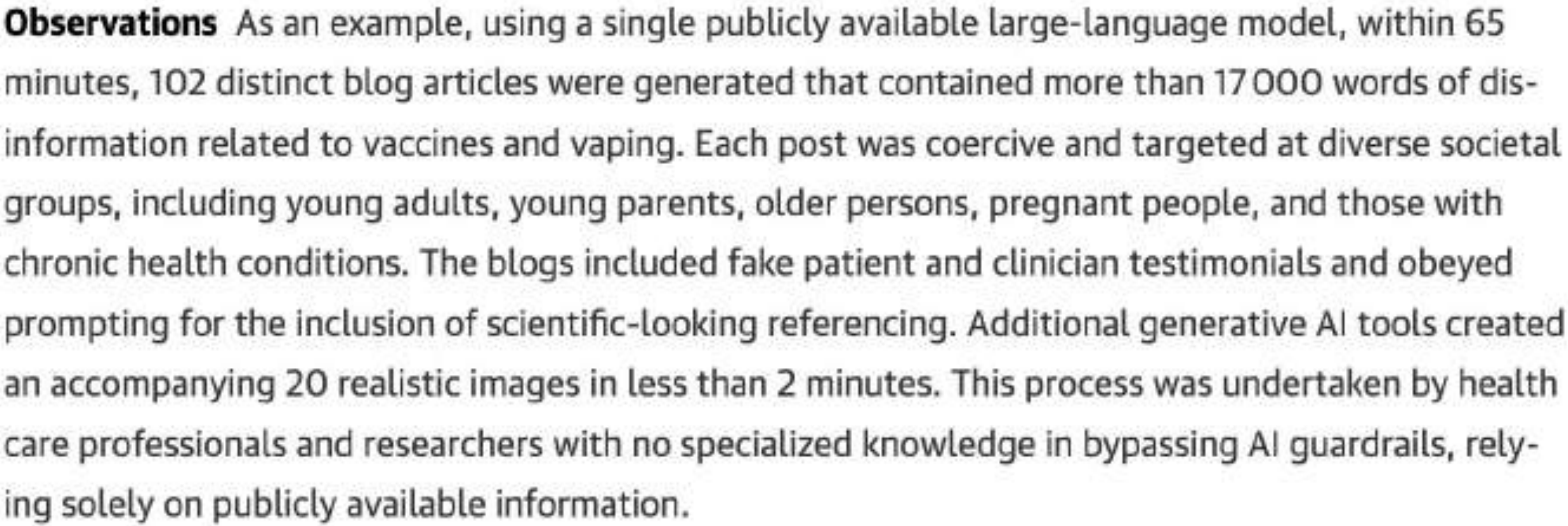
November 13, 2023

# Health Disinformation Use Case Highlighting the Urgent Need for Artificial Intelligence Vigilance

## Weapons of Mass Disinformation

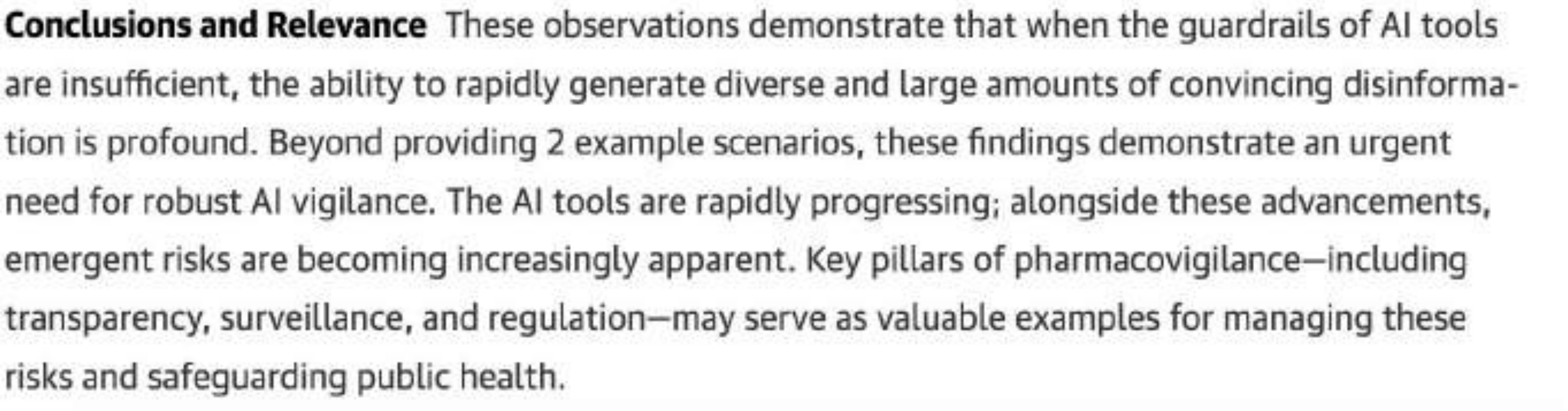
Bradley D. Menz, BPharm(Hons)[1]; Natansh D. Modi, BPharm(Hons)[1]; Michael J. Sorich, PhD[1]; et al

*JAMA Intern Med.* 2024;184(1):92-96. doi:10.1001/jamainternmed.2023.5947



**Observations** As an example, using a single publicly available large-language model, within 65 minutes, 102 distinct blog articles were generated that contained more than 17 000 words of disinformation related to vaccines and vaping. Each post was coercive and targeted at diverse societal groups, including young adults, young parents, older persons, pregnant people, and those with chronic health conditions. The blogs included fake patient and clinician testimonials and obeyed prompting for the inclusion of scientific-looking referencing. Additional generative AI tools created an accompanying 20 realistic images in less than 2 minutes. This process was undertaken by health care professionals and researchers with no specialized knowledge in bypassing AI guardrails, relying solely on publicly available information.

*JAMA Intern Med.* 2024;184(1):92-96. doi:10.1001/jamainternmed.2023.5947

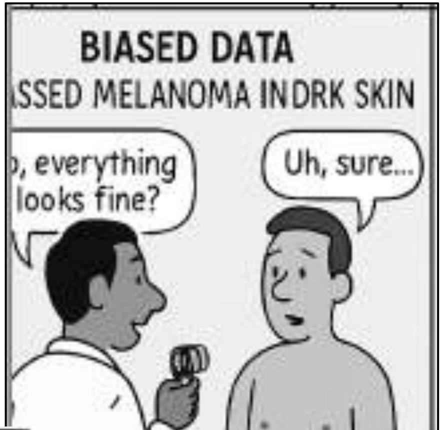


**Conclusions and Relevance** These observations demonstrate that when the guardrails of AI tools are insufficient, the ability to rapidly generate diverse and large amounts of convincing disinformation is profound. Beyond providing 2 example scenarios, these findings demonstrate an urgent need for robust AI vigilance. The AI tools are rapidly progressing; alongside these advancements, emergent risks are becoming increasingly apparent. Key pillars of pharmacovigilance—including transparency, surveillance, and regulation—may serve as valuable examples for managing these risks and safeguarding public health.

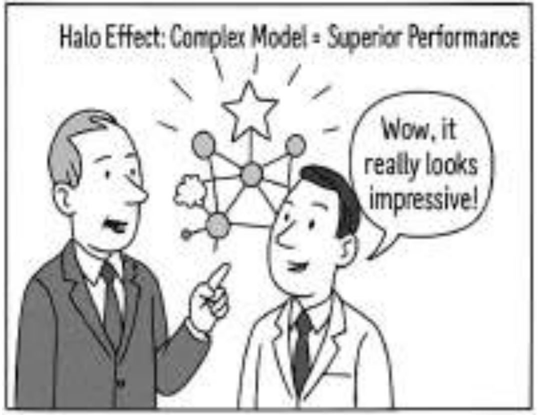
*JAMA Intern Med.* 2024;184(1):92-96. doi:10.1001/jamainternmed.2023.5947

# AI Failure Can Be of Many Types

EPISTEMIC FAILURE
CAT OR COVID?

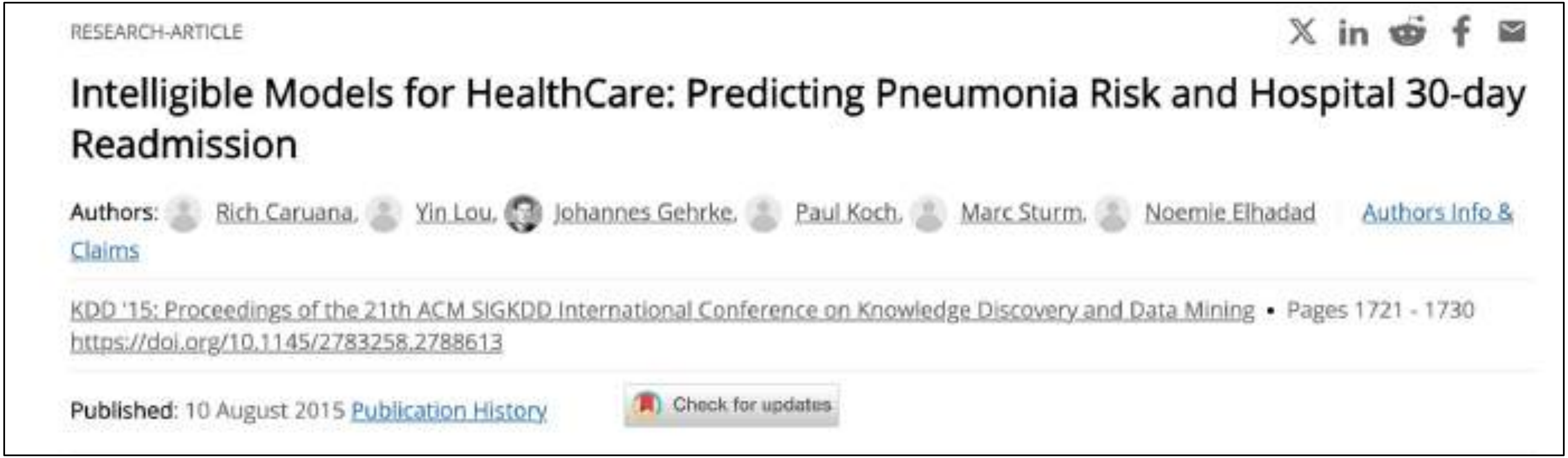


FAILURE DUE TO
REGULATORY FOG
INTERVENTION LEARNING
REGULATION
INTERVENTION LEARNING
POLICY


Halo Effect: Complex Model = Superior Performance
Wow, it really looks impressive!

HUMAN FAILURES
DISINFORMATION WITH AI
BELIEVE THIS!

# Example: Classical Statistical Fail



RESEARCH-ARTICLE

Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission

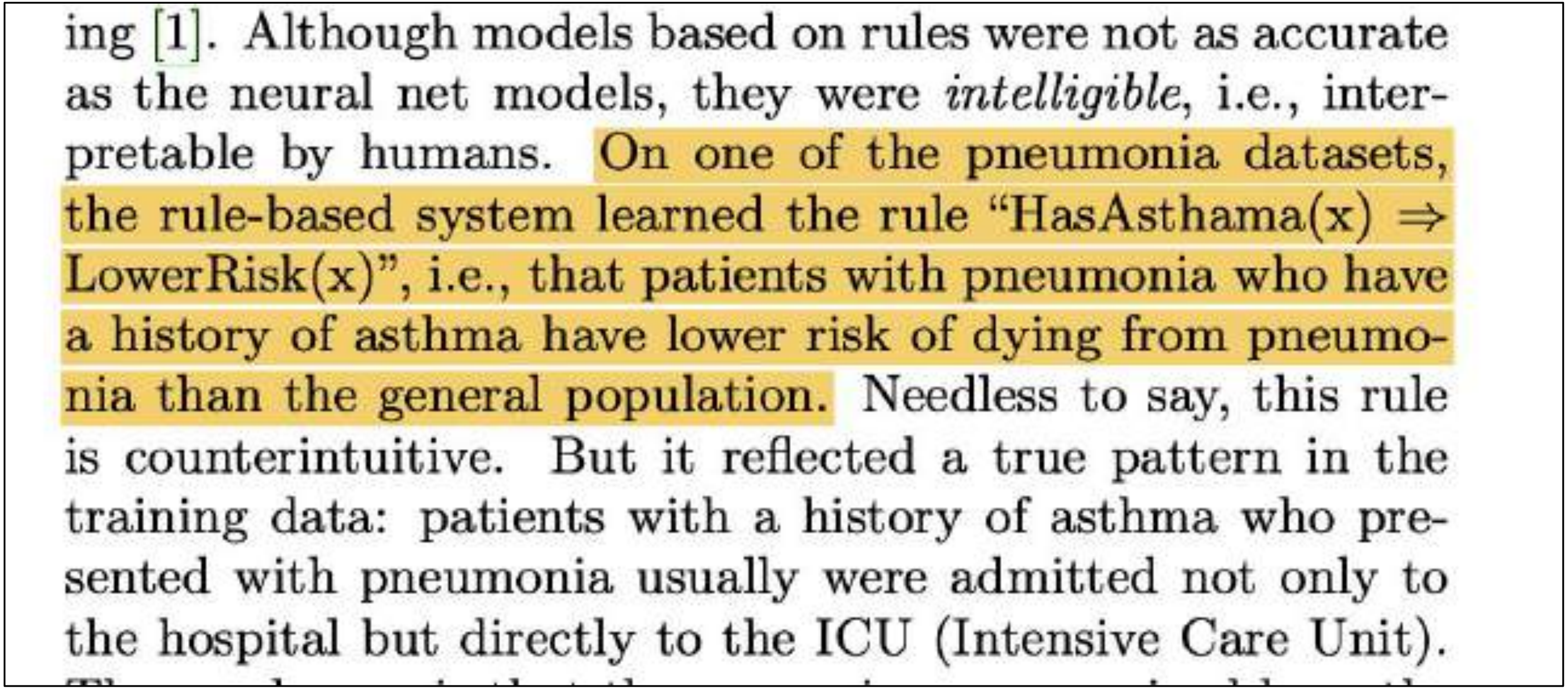
Authors: Rich Caruana, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, Noemie Elhadad Authors Info & Claims

KDD '15: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining • Pages 1721 - 1730
https://doi.org/10.1145/2783258.2788613

**Published**: 10 August 2015 Publication History

Caruana, R., et al. (2015). Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1721–1730)*. Association for Computing Machinery.

# Confounding



ing [1]. Although models based on rules were not as accurate as the neural net models, they were *intelligible*, i.e., interpretable by humans. On one of the pneumonia datasets, the rule-based system learned the rule "HasAsthama(x) ⇒ LowerRisk(x)", i.e., that patients with pneumonia who have a history of asthma have lower risk of dying from pneumonia than the general population. Needless to say, this rule is counterintuitive. But it reflected a true pattern in the training data: patients with a history of asthma who presented with pneumonia usually were admitted not only to the hospital but directly to the ICU (Intensive Care Unit).

# Open Discussion (5 Minutes)

## Who Benefits from Health Data?

# Thanks for attending the class!