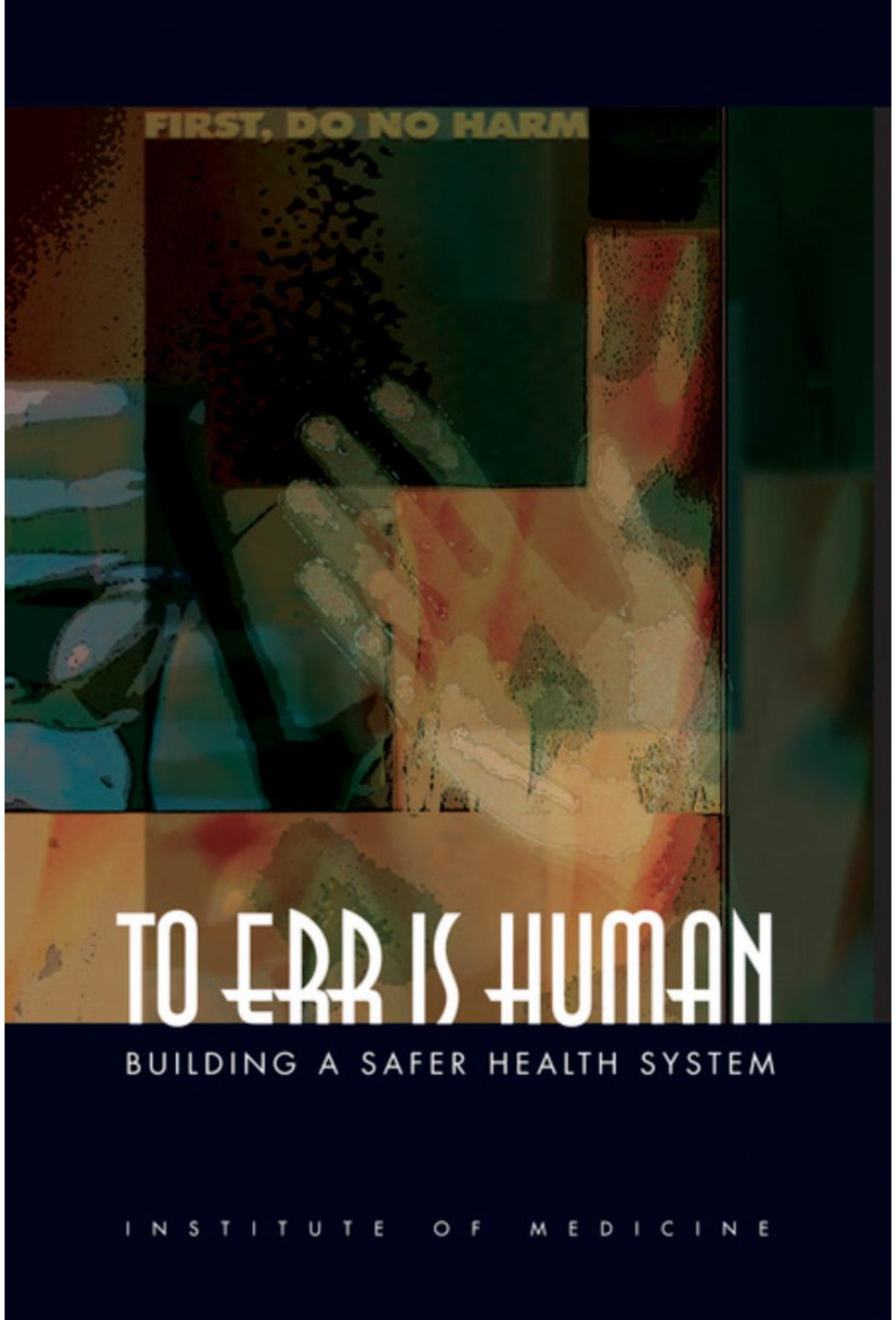
From AI Bias to AI by Us, For All of Us

Leo Anthony Celi

Massachusetts Institute of Technology

Harvard Medical School

Harvard T.H. Chan School of Public Health


FIRST, DO NO HARM
TO ERR IS HUMAN
BUILDING A SAFER HEALTH SYSTEM
INSTITUTE OF MEDICINE

The Perfect Storm

Introduced in House (01/07/2025)

Healthy Technology Act of 2025

This bill establishes that artificial intelligence (AI) or machine learning technology may be eligible to prescribe drugs.

Currently, certain drugs may be dispensed only upon a prescription provided by a practitioner licensed by law to administer the drug. Under this bill, an AI or machine learning technology may qualify as such a prescribing practitioner if the technology is (1) authorized by state law to prescribe the drug involved; and (2) approved, cleared, or authorized under certain federal provisions pertaining to medical devices and products.

CONGRESS.GOV

STAT+ | EXCLUSIVE

FDA's head of device safety among employees terminated in 'challenging' weekend

The New York Times

Updates | Tracking Major Moves | Lawsuits Tracker | In Trump's Cross Hairs | Federal W

F.D.A. Firings Decimated Teams Reviewing A.I. and Food Safety

Staff units evaluating high-tech surgical robots and insulin-delivery systems were gutted by Trump layoffs even though industry fees, not taxpayers, financed the employee salaries.

St. Petersburg plans to introduce ethnicity recognition to monitor migrants

Feb 21, 2025, 10:44 am EST | Masha Borak

CATEGORIES Biometrics News | Facial Recognition | Surveillance

The Russian city of St. Petersburg plans to equip 8,000 cameras with ethnicity recognition capabilities to monitor its migrant population, city authorities announced this week.

TECHNOLOGY

AI IS TAKING WATER FROM THE DESERT

New data centers are springing up every week. Can the Earth sustain them?

By Karen Hao

ARTIFICIAL INTELLIGENCE

Training a single AI model can emit as much carbon as five cars in their lifetimes

Deep learning has a terrible carbon footprint.

By Karen Hao

June 6, 2019

BOTCHED OPERATION

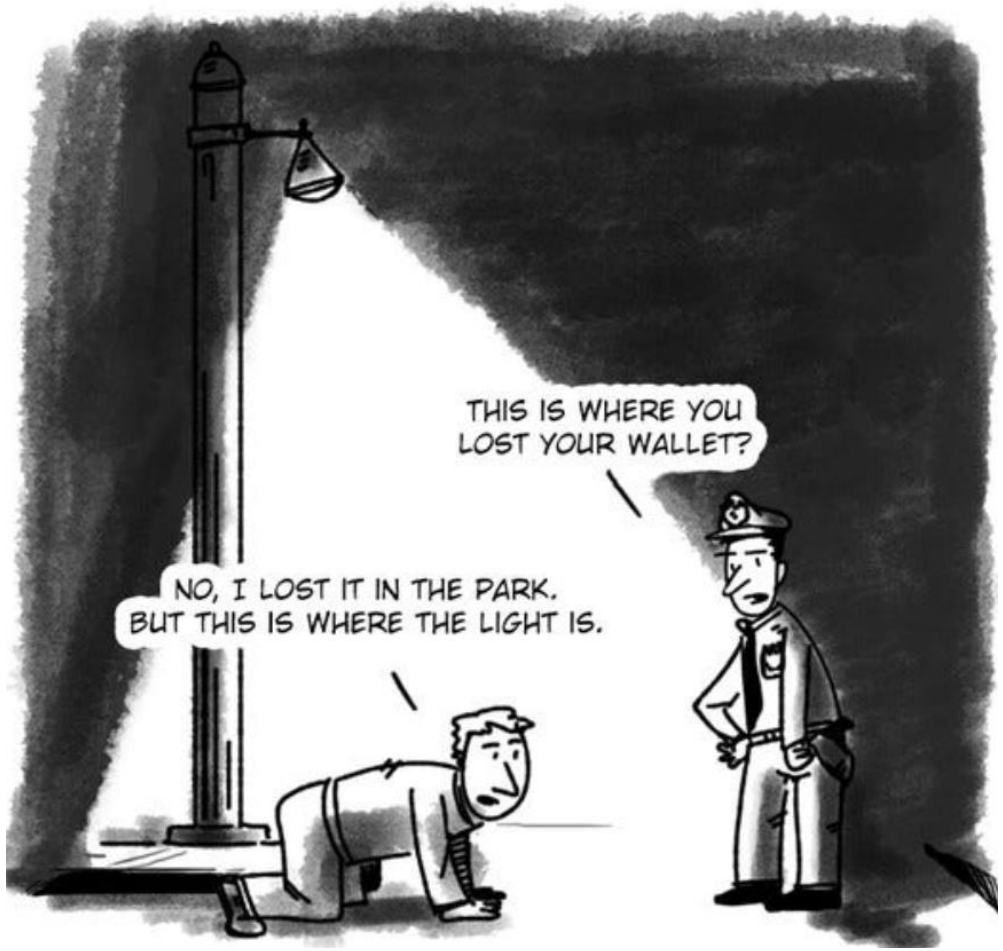
Death By 1,000 Clicks: Where Electronic Health Records Went Wrong

The U.S. government claimed that turning American medical charts into electronic records would make health care better, safer and cheaper. Ten years and $36 billion later, the system is an unholy mess. Inside a digital revolution that took a bad turn.

By **Fred Schulte** and **Erika Fry, Fortune** • MARCH 18, 2019

(The Voorhes for Fortune)

The “Backstory” of the Data:
Social Patterning of the Data Generation Process



Critical Care Clinics
Available online 27 March 2023
In Press, Corrected Proof What's this?

Critical Bias in Critical Care Devices

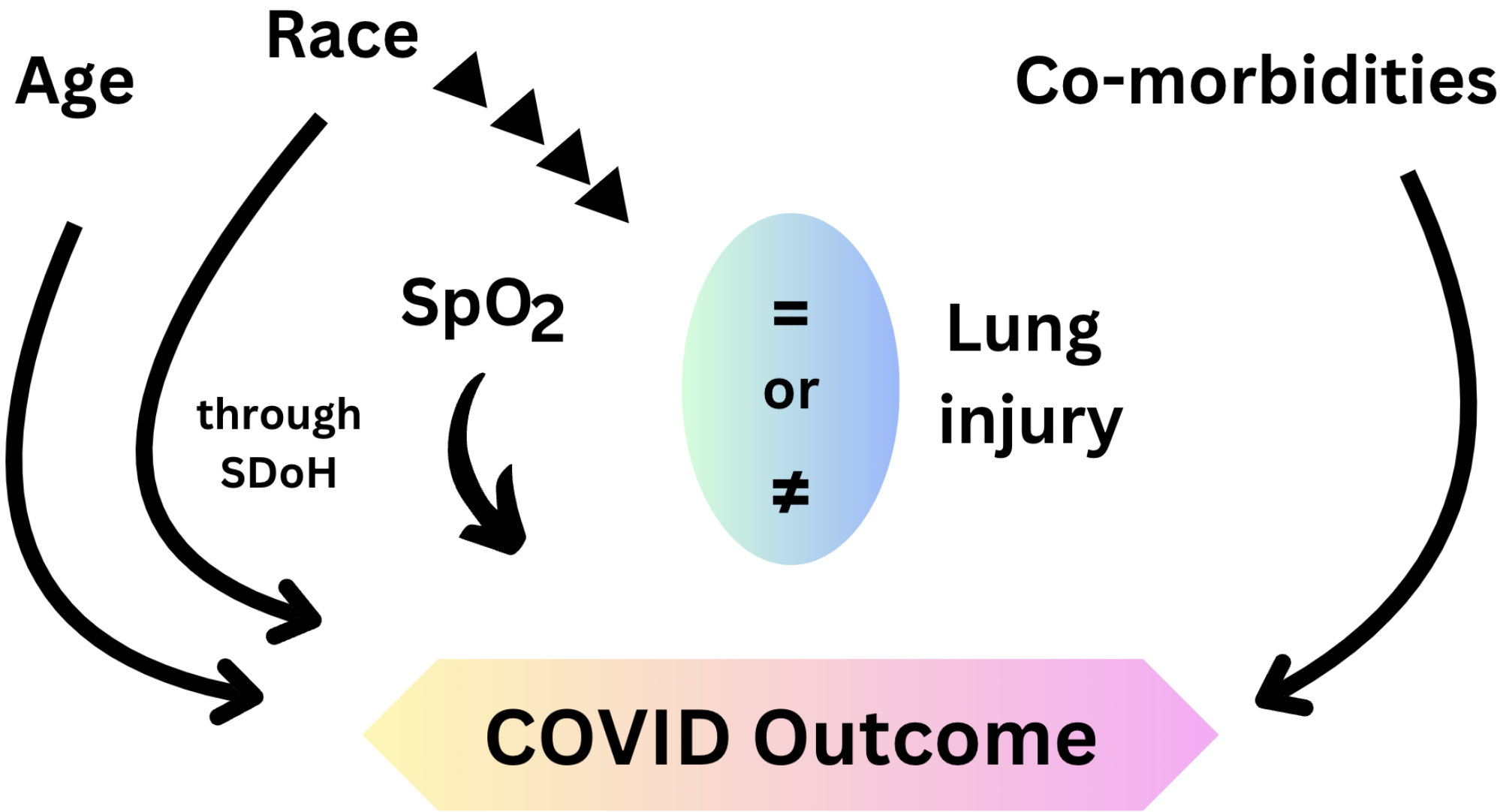
Marie-Laure Charpignon MSc [a], Joseph Byers [b], Stephanie Cabral MD [c], Leo Anthony Celi MD, MSc, MPH [d e f], Chrystinne Fernandes PhD [d], Jack Gallifant [g], Mary E. Lough PhD, RN, CCNS, FCCM, FAHA, FCNS, FAAN [h], Donald Mlombwa [i j k], Lama Moukheiber MSc [l], Bradley Ashley Ong MD [m], Anupol Panitchote MD [r], An-Kwok Ian Wong MD, PhD [p], Lama Nazer PharmD, BCPS, FCCM [q]

Journal of Biomedical Informatics
Volume 153, May 2024, 104643

Special Communication

Variation in monitoring: Glucose measurement in the ICU as a case study to preempt spurious correlations

Khushboo Teotia [a 1], Yueran Jia [a 1], Naira Link Woite [b], Leo Anthony Celi [a b c], João Matos [a d e 2], Tristan Struja [a f 2]


Age
Race
Co-morbidities
through SDoH
SpO_2
=
or
≠
Lung injury
COVID Outcome

Sampling Selection: Who is in the database?

JAMA Network

JAMA Network Open

JAMA Network Open

Enter Search Term

This Issue | Views 23,508 | Citations 1 | Altmetric 1323 | Comments 1

Download PDF | More | CME & MOC | Cite This | Permissions

Original Investigation | Public Health

January 19, 2021

Association of Intensive Care Unit Patient Load and Demand With Mortality Rates in US Department of Veterans Affairs Hospitals During the COVID-19 Pandemic

Dawn M. Bravata, MD[1,2,3,4,5,6]; Anthony J. Perkins, MS[1,7]; Laura J. Myers, PhD[1,2,4,6]; et al

Sampling Selection: Who is in the database?

Table 3. Proportional Hazard Results From Admission to 30 Days Postdischarge or Death

Characteristic	Overall		General ward only		ICU	
	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value
Age, y						
<65	1 [Reference]	<.001	1 [Reference]	<.001	1 [Reference]	<.001
65-74	2.10 (1.75-2.52)		2.35 (1.55-3.57)		2.08 (1.70-2.55)	
75-84	3.04 (2.50-3.69)		4.79 (3.15-7.29)		2.73 (2.18-3.42)	
≥85	6.85 (5.59-8.39)		11.15 (7.33-16.97)		5.39 (4.20-6.92)	
History of smoking						
Never	1 [Reference]	<.001	1 [Reference]	.001	1 [Reference]	<.001
Current	0.70 (0.56-0.87)		0.73 (0.47-1.14)		0.74 (0.57-0.96)	
Former	0.93 (0.81-1.06)		0.98 (0.77-1.26)		0.92 (0.78-1.08)	
Unknown	1.55 (1.30-1.84)		1.59 (1.19-2.12)		1.59 (1.27-1.98)	
BMI						
<25	1 [Reference]	.001	1 [Reference]	.01	1 [Reference]	.12
25-29	0.83 (0.72-0.95)		0.75 (0.59-0.94)		0.88 (0.74-1.05)	
30-34	0.80 (0.68-0.94)		0.72 (0.53-0.97)		0.84 (0.69-1.02)	
≥35	0.94 (0.79-1.11)		0.96 (0.68-1.36)		0.98 (0.79-1.20)	
Missing data	1.38 (0.97-1.98)		1.65 (0.92-2.93)		1.39 (0.87-2.22)	

The Lancet Digital Health
Available online 11 May 2022
In Press, Corrected Proof

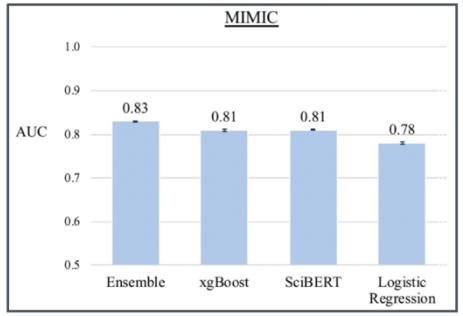
Articles

AI recognition of patient race in medical imaging: a modelling study

Judy Wawira Gichoya MD [a], Imon Banerjee PhD [c], Ananth Reddy Bhimireddy MS [a], John L Burns MS [d], Leo Anthony Celi MD [e, g], Li-Ching Chen BS [h], Ramon Correa BS [c], Natalie Dullerud MS [i], Marzyeh Ghassemi PhD [e, f], Shih-Cheng Huang [j], Po-Chih Kuo PhD [h], Matthew P Lungren MD [j], Lyle J Palmer PhD [k, l], Brandon J Price MD [n] Saptarshi Purkayastha PhD [d], Ayis T Pyrros MD [n], Lauren Oakden-Rayner MD [k], Chima Okechukwu MS [o] … Haoran Zhang MS [i]

Nursing Progress Note

NEURO: sedated with propofol gtt 85mcg/kg
RESP: remains intubated with IMV 12/750/5peep/5psv/40%fio2
GU: inc large amt foul smelling urine foley placed with UO ~50cc/hr, **dialysis** to be initiated at 6pm
SKIN: sacral decub w-d dsg changes **wound red** beefy, small amt bloody drainage, heel dsg w-d dsg changed no drainage
ACCESS: left EJ, right groin introducer, left rad aline
PLAN: **dialysis** this eve, wean extubate tomorrow, titrate up po meds for **hypertension**



… for difficult reasons.

Word	% Black Patients	% White Patients
very difficult	4.94%	3.87%
difficult to understand	4.06%	3.19%
difficult to assess	3.99%	3.46%
is difficult	3.53%	2.36%
and difficult	3.32%	2.90%
very difficult to	3.28%	2.27%
difficult stick	2.40%	1.09%

An artificial intelligence system predicted the sex of people just by analyzing their eyes

It is a model based on deep learning that was trained with photos of retinas. Applications in the field of medicine

Newsroom Infobae

March 15, 2022

ORIGINAL ARTICLE DATA SCIENCE | VOLUME 19, ISSUE 1, P184-191, JANUARY 01, 2022

Detecting Racial/Ethnic Health Disparities Using Deep Learning From Frontal Chest Radiography

Ayis Pyrros, MD • Jorge Mario Rodríguez-Fernández, MD • Stephen M. Borstelmann, MD • Judy Wawira Gichoya, MD • Jeanne M. Horowitz, MD • Brian Fornelli, MS • Nasir Siddiqui, MD • Yury Velichko, PhD • Oluwasanmi Koyejo, PhD • William Galanter, MD, PhD • Show less

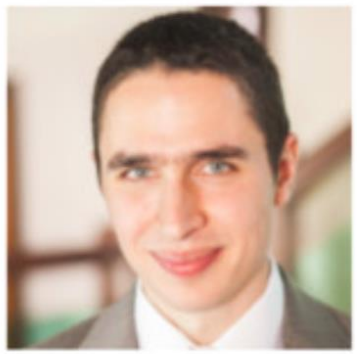
DOI: https://doi.org/10.1016/j.jacr.2021.09.010 • Check for updates

"Why Should I Trust You?" Explaining the Predictions of Any Classifier

Sameer Singh

UCIrvine

Marco Tulio Ribeiro

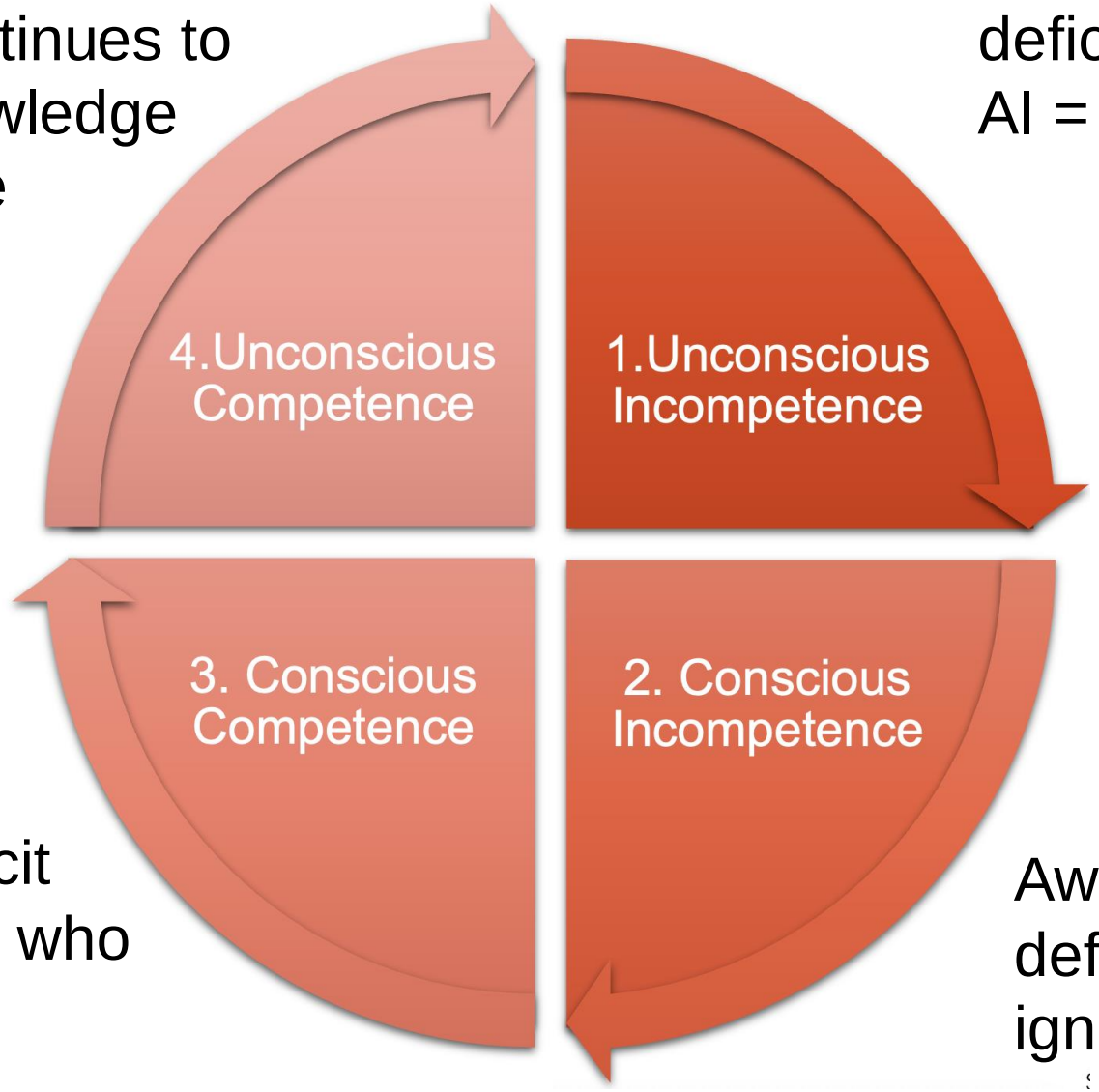


UNIVERSITY of WASHINGTON

Carlos Guestrin

Routinely consults others who don't think like them but continues to acknowledge that knowledge may still be incomplete

Unaware of knowledge deficit, also known as
AI = arrogant & ignorant



Fills knowledge deficit by consulting others who don't think like them

Aware of knowledge deficit but stuck with ignorance

https://www.nnlm.gov/guides/data-glossary/data-provenance

What is Data Provenance?
A documented trail that:

Tracks the **origin** of a piece of data.	Logs its **movements and changes** over time.	Ensures the **credibility** and **transparency** of research data.

This Issue Views 26,161 | Citations 0 | Altmetric 94 | Comments 5

Viewpoint | AI in Medicine

January 29, 2024

AI-Generated Clinical Summaries Require More Than Accuracy

Katherine E. Goodman, JD, PhD[1,2]; Paul H. Yi, MD[3]; Daniel J. Morgan, MD, MS[1,4]

» Author Affiliations | Article Information

JAMA. 2024;331(8):637-638. doi:10.1001/jama.2024.0555

HOSPITALS

STAT+

Hospitals struggle to validate AI-generated clinical summaries. 'It's a bit chaotic'

By Katie Palmer and Casey Ross March 13, 2024

Reprints

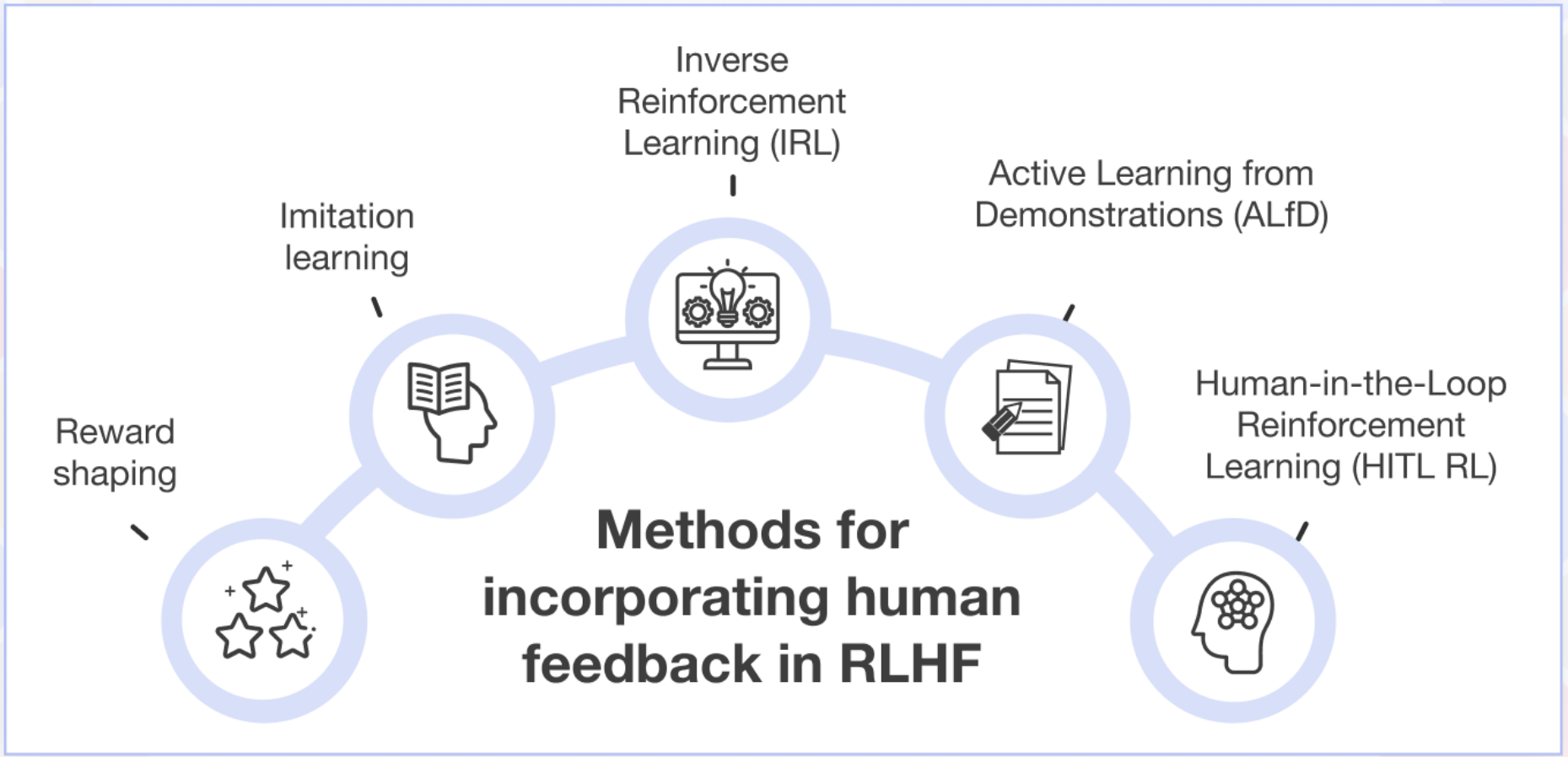
The Lancet Digital Health

Volume 6, Issue 1, January 2024, Pages e12-e22

Articles

Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: a model evaluation study

Travis Zack PhD [a b] †, Eric Lehman MSc [c] †, Mirac Suzgun [e f], Jorge A Rodriguez MD [h], Prof Leo Anthony Celi MD [d j k], Prof Judy Gichoya MD [m], Prof Dan Jurafsky PhD [e g], Prof Peter Szolovits PhD [c], Prof David W Bates MD [h l], Prof Raja-Elie E Abdulnour MD [i n], Prof Atul J Butte MD [a o], Emily Alsentzer PhD [h n]


Inverse Reinforcement Learning (IRL)
Imitation learning
Active Learning from Demonstrations (ALfD)
Reward shaping
Human-in-the-Loop Reinforcement Learning (HITL RL)
Methods for incorporating human feedback in RLHF

SIMFORM

NBC NEWS

TSA Chief Out After Agents Fail 95 Percent of Airport Breach Tests

A DHS investigation found security breaches that allowed undercover investigators to smuggle weapons through checkpoints at dozens of U.S. airports.

seidos
HOLD IT
6
4
1

Human Factors: The Journal of the Human Factors and Ergonomics Society

HFES Systems That Work for Humans

Impact Factor: **2.9** / 5-Year Impact Factor: **3.8**

Jou

Restricted access | Research article | First published online June 25, 2018

Redundant Automation Monitoring: Four Eyes Don't See More Than Two, if Everyone Turns a Blind Eye

Dietlind Helene Cymek ✉ View all authors and affiliations

Volume 60, Issue 7 | https://doi.org/10.1177/0018720818781192

appviewx

POSTED ON AUGUST 1, 2023 BY DEBARATI BISWAS

New EMA Research Report Spotlights SSL/TLS Certificate Management Challenges

TechCrunch

MIT study finds Tesla drivers become inattentive when Autopilot is activated

"I Love Lucy"

JAMA The Journal of the American Medical Association

This Issue | Views 11,587 | Citations 8 | Altmetric 91 | Comments 1

Viewpoint

March 14, 2024

The Limits of Clinician Vigilance as an AI Safety Bulwark

Julia Adler-Milstein, PhD[1]; Donald A. Redelmeier, MD[2,3]; Robert M. Wachter, MD[1]

Author Affiliations | Article Information

[1]Department of Medicine, University of California, San Francisco

[2]Department of Medicine, University of Toronto, Toronto, Ontario, Canada

[3]Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

JAMA. 2024;331(14):1173-1174. doi:10.1001/jama.2024.3620

JAMA The Journal of the American Medical Association

This Issue | Views 23,378 | Citations 35 | Altmetric 392

Download PDF | More | Cite This | Permissions

Original Investigation | AI in Medicine

December 19, 2023

Measuring the Impact of AI in the Diagnosis of Hospitalized Patients

A Randomized Clinical Vignette Survey Study

Sarah Jabbour, MSE[1]; David Fouhey, PhD[1,2,3]; Stephanie Shepard, PhD[1]; et al

» Author Affiliations | Article Information

JAMA. 2023;330(23):2275-2284. doi:10.1001/jama.2023.22295

AJR American Journal of Roentgenology

Search American Journal of Roentgenology

ADVANCED SEARCH

ARTICLES & ISSUES COLLECTIONS INFORMATION AUTHORS REVIEWERS MULTIMEDIA SECTIONS

Original Research | Neuroradiology/Head and Neck Imaging | September 4, 2024

Prospective Evaluation of Artificial Intelligence Triage of Intracranial Hemorrhage on Noncontrast Head CT Examinations

Authors: Cody H. Savage, MD, Manoj Tanwar, MD, Asser Abou Elkassem, MD, Adam Sturdivant, MPH, MS, Omar Hamki, BS, Houman Sotoudeh, MD, Gopi Sirineni, MD, Aparna Singhal, MD, Desmin Milner, MD, Jesse Jones, MD, Dirk Rehder, MD, Mei Li, PhD, Yufeng Li, PhD, Kevin Junck, PhD, Srini Tridandapani, MD, PhD, MBA, Steven A. Rothenberg, MD, and Andrew D. Smith, MD, PhD SHOW FEWER | AUTHOR INFO & AFFILIATIONS

Does AI Help or Hurt Human Radiologists' Performance? It Depends on the Doctor

New research shows radiologists and AI don't always work well together

By EKATERINA PESHEVA | March 19, 2024 | Research

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature medicine › articles › article

Article | Open access | Published: 19 March 2024

Heterogeneity and predictors of the effects of AI assistance on radiologists

Feiyang Yu, Alex Moehring, Oishi Banerjee, Tobias Salz, Nikhil Agarwal & Pranav Rajpurkar ✉

Nature Medicine **30**, 837–849 (2024) | Cite this article

JAMA The Journal of the American Medical Association

This Issue | Views 144,369 | Citations 12 | Altmetric 1182 | Comments 4

Download PDF | More | Cite This | Permissions

Original Investigation | Health Informatics

October 28, 2024

Large Language Model Influence on Diagnostic Reasoning
A Randomized Clinical Trial

Ethan Goh, MBBS, MS[1,2]; Robert Gallo, MD[3]; Jason Hom, MD[4]; Eric Strong, MD[4]; Yingjie Weng, MHS[5]; Hannah Kerman, MD[6,7]; Joséphine A. Cool, MD[6,7]; Zahir Kanjee, MD, MPH[6,7]; Andrew S. Parsons, MD, MPH[8]; Neera Ahuja, MD[4]; Eric Horvitz, MD, PhD[9,10]; Daniel Yang, MD[11]; Arnold Milstein, MD[2]; Andrew P. J. Olson, MD[12]; Adam Rodman, MD, MPH[6,7]; Jonathan H. Chen, MD, PhD[1,2,13]

» Author Affiliations | Article Information

JAMA Netw Open. 2024;7(10):e2440969. doi:10.1001/jamanetworkopen.2024.40969

Research

Electronic health record alerts for acute kidney injury: multicenter, randomized clinical trial

BMJ 2021 ; 372 doi: https://doi.org/10.1136/bmj.m4786 (Published 18 January 2021)
Cite this as: *BMJ* 2021;372:m4786

Opinion

The challenge of minimal risk in e-alert trials

F Perry Wilson, associate professor[1 2], *Melissa Martin*, research associate[1 2], *Yu Yamamoto*, statistician[1 2], *Caitlin Partridge*, data scientist[3], *Erica Moreira*, data scientist[3], *Tanima Arora*, postdoctoral researcher[1 2], *Aditya Biswas*, data scientist[1 2], *Harold Feldman*, professor[4], *Amit X Garg*, professor[5], *Jason H Greenberg*, assistant professor[2 6], *Monique Hinchcliff*, associate professor[7], *Stephen Latham*, professor[8], *Fan Li*, assistant professor[9], *Haiqun Lin*, professor[10], *Sherry G Mansour*, instructor[1 2], *Dennis G Moledina*, assistant professor[1 2], *Paul M Palevsky*, professor[11], *Chirag R Parikh*, professor[12], *Michael Simonov*, instructor[2], *Jeffrey Testani*, associate professor[13], *Ugochukwu Ugwuowo*, postdoctoral researcher[1 2]

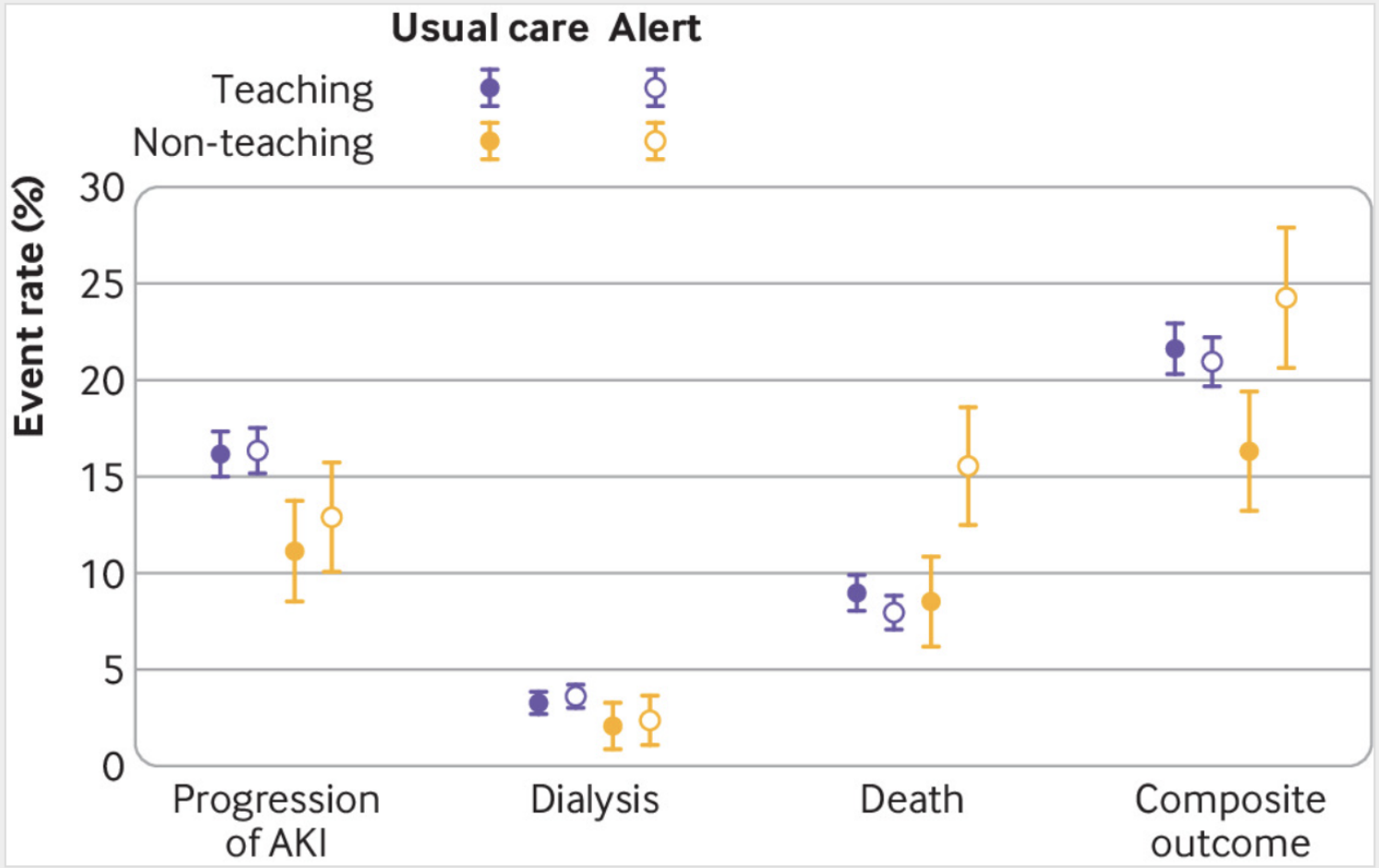


Fig 2

Primary and secondary outcome events, stratified by hospital type. Error bars are 95% confidence intervals of the observed proportion of events. AKI=acute kidney injury

- >2,400 hospitals surveyed in 2023: 65% reported using AI
  - inpatients' health trajectories (92%)
  - spot high-risk outpatients (79%)
  - scheduling (51%)
  - billing support (36%)

- 61% evaluated models using local data
- 44% reported performing bias interrogation
- 4 in 5 used models from EHR vendor
- Critical access and rural hospitals buy "off the shelf" AI

Research Methods & Reporting

TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods

BMJ 2024 ; 385 doi: https://doi.org/10.1136/bmj-2023-078378 (Published 16 April 2024)

Gary S Collins [1], *professor* [1], *Karel G M Moons* [2], *professor* [2], *Paula Dhiman* [1], *senior researcher in medical statistics* [1], *Richard D Riley* [3 4], *professor* [3 4], *Andrew L Beam* [5], *assistant professor* [5], *Ben Van Calster* [6 7], *associate professor* [6 7], *Marzyeh Ghassemi* [8], *assistant professor* [8], *Xiaoxuan Liu* [9 10], *senior clinician scientist* [9 10], *Johannes B Reitsma* [2], *associate professor* [2], *Maarten van Smeden* [2], *associate professor* [2], *Anne-Laure Boulesteix* [11], *professor* [11], *Jennifer Catherine Camaradou* [12 13], *Leo Anthony Celi* [14 15 16], *principal research scientist* [14 15 16], *Spiros Denaxas* [17 18], *professor* [17 18], *Alastair K Denniston* [4 9], *professor* [4 9], *Ben Glocker* [19], *professor* [19], *Robert M Golub* [20], *professor* [20], *Hugh Harvey* [21], *managing director* [21], *Georg Heinze* [22], *associate professor* [22], *Michael M Hoffman* [23 24 25 26], *associate professor* [23 24 25 26], *André Pascal Kengne* [27], *professor* [27], *Emily Lam* [12], *Naomi Lee* [28], *head of organisational transformation* [28], *Elizabeth W Loder* [29 30], *professor* [29 30], *Lena Maier-Hein* [31], *professor* [31], *Bilal A Mateen* [17 32 33], *associate professor* [17 32 33], *Melissa D McCradden* [34 35], *assistant professor* [34 35], *Lauren Oakden-Rayner* [36], *director of research* [36], *Johan Ordish* [37], *deputy director* [37], *Richard Parnell* [12], *Sherri Rose* [38], *professor* [38], *Karandeep Singh* [39], *associate professor* [39], *Laure Wynants* [40], *assistant professor* [40], *Patricia Logullo* [1], *EQUATOR researcher* [1]

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature medicine › consensus statements › article

Consensus Statement | Published: 08 January 2025

The TRIPOD-LLM reporting guideline for studies using large language models

Jack Gallifant, Majid Afshar, Saleem Ameen, Yindalon Aphinyanaphongs, Shan Chen, Giovanni Cacciamani, Dina Demner-Fushman, Dmitriy Dligach, Roxana Daneshjou, Chrystinne Fernandes, Lasse Hyldig Hansen, Adam Landman, Lisa Lehmann, Liam G. McCoy, Timothy Miller, Amy Moreno, Nikolaj Munch, David Restrepo, Guergana Savova, Renato Umeton, Judy Wawira Gichoya, Gary S. Collins, Karel G. M. Moons, Leo A. Celi & Danielle S. Bitterman ✉

What Auto Safety Teaches Us About AI Safety

"A complex system that works is invariably found to have evolved from a simple system that worked." - John Gall, Systemantics

ASIMOV'S ADDENDUM AND TIM O'REILLY

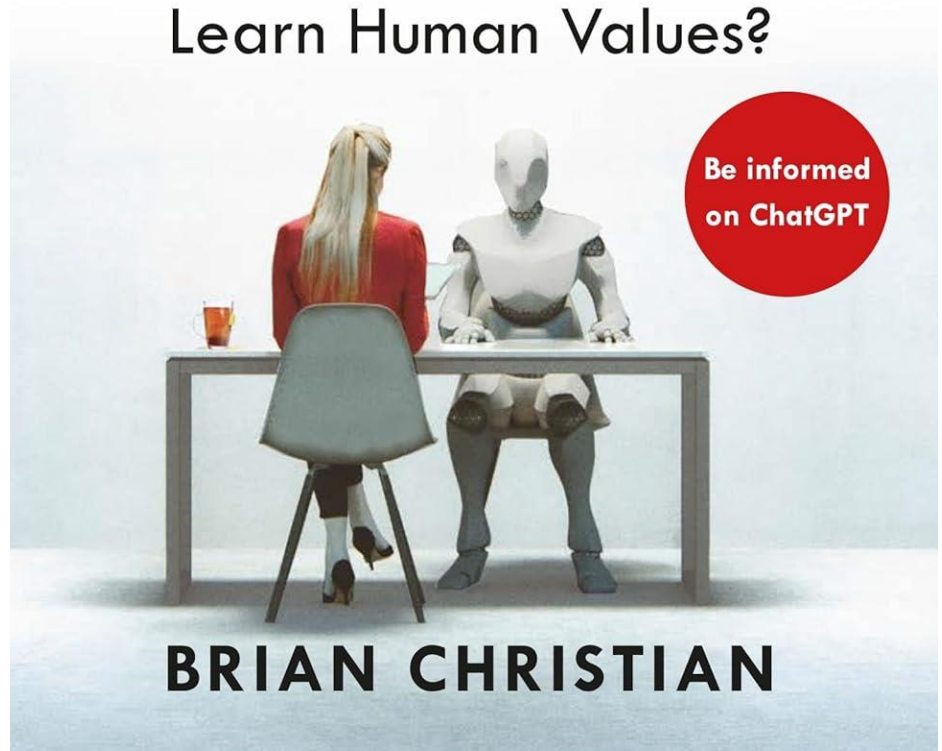
NOV 20, 2024

- Safety-centered human-AI system design thinking
- Equity-focused post-deployment continuous monitoring with system redesign as needed
- Revamping of legacy systems for education, innovation and regulation

‘Vital reading. This is the book on artificial intelligence that we need right now’ **Mike Krieger, co-founder of Instagram**

THE ALIGNMENT PROBLEM

How Can Artificial Intelligence Learn Human Values?



iScience. 2021 Jun 25; 24(6): 102656.
Published online 2021 Jun 10. doi: 10.1016/j.isci.2021.102656
PMCID: PMC8209268
PMID: 34169236

Village mentoring and hive learning: The MIT Critical Data experience

Christopher V. Cosgriff,[1] Marie Charpignon,[2] Dana Moukheiber,[3] Mary E. Lough,[4] Judy Gichoya,[5] David J. Stone,[6] and Leo Anthony Celi[7,8]

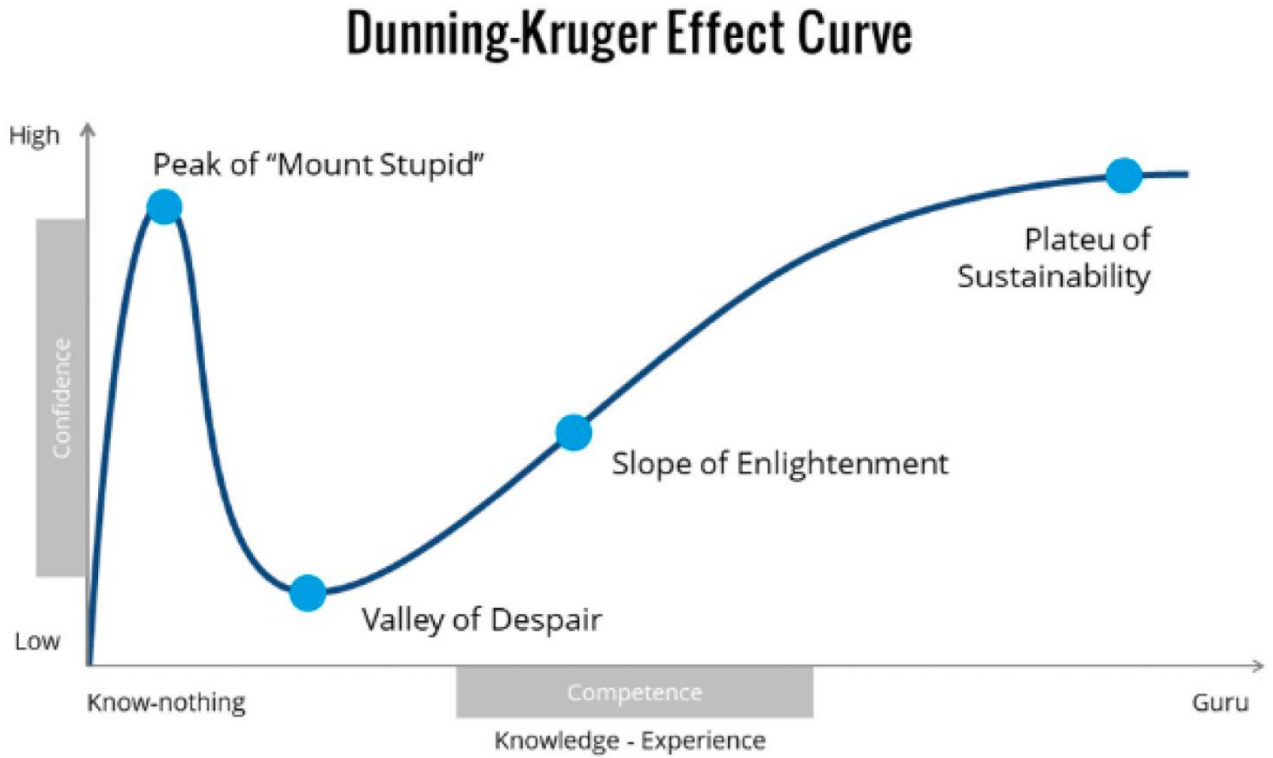
Author information Copyright and License information PMC Disclaimer

SECC-Best of SCCM Congress 2024 Taipei
暨 2024 急重症聯合學術年會

DISCOVERY
The Critical Care Research Network
UMASS BOSTON
MIT

ESICM
POLITECNICO
MIT Critical Data
Milan Critical Care Datathon
and ESICM's Big Data talk
Milan 1-3 February 2019

What do we teach about AI?

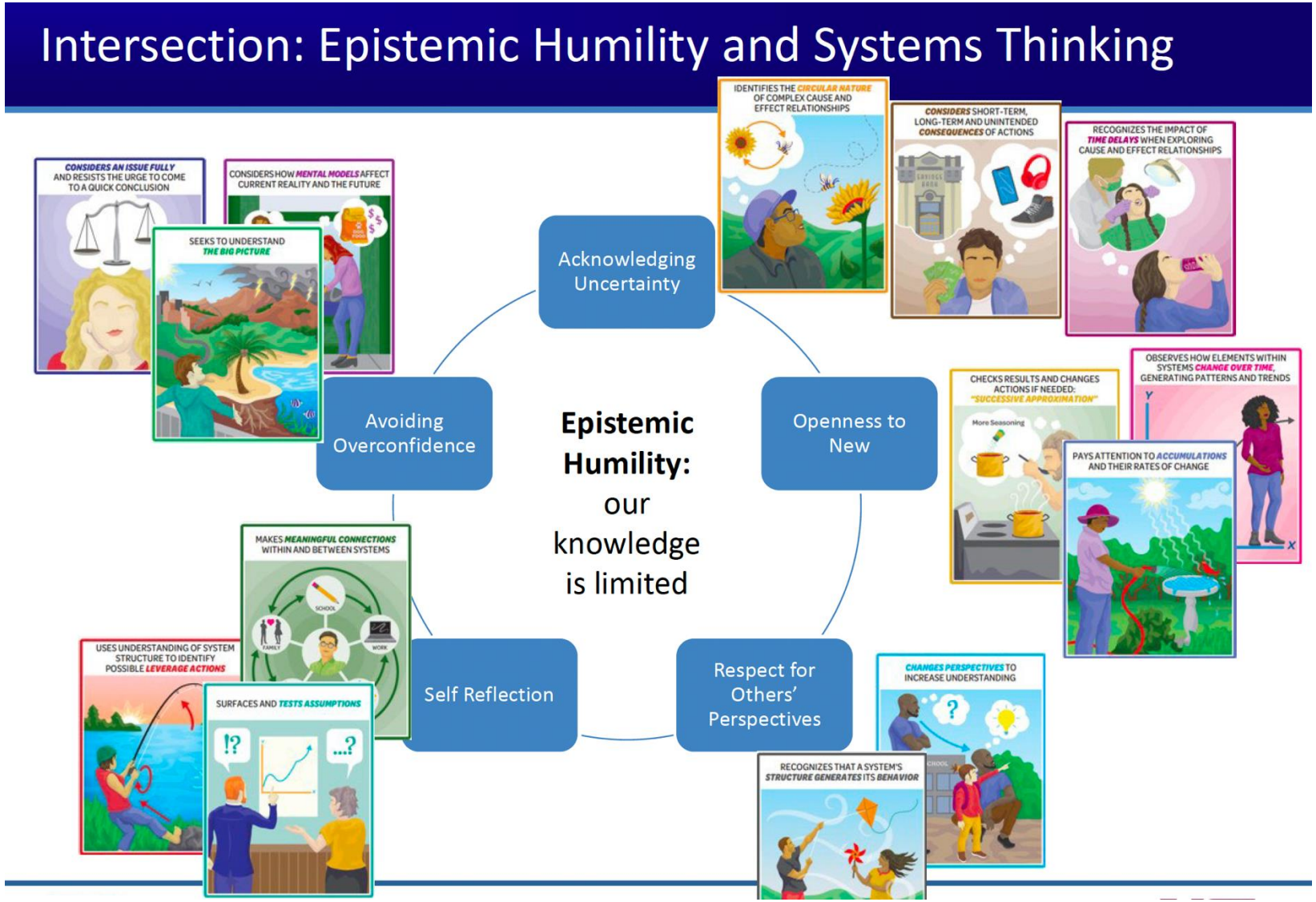


What do we teach about AI?

- Health system science / systems thinking / critical thinking
- Epistemic humility and pursuit of plurality
- “When everyone is thinking alike, no one is truly thinking.”
- Regular reflection on what motivates scientific thinking: publications & citations, validation & ego, professional identity
- Reflection only works when done amid those who think differently (and when there is psychological safety to challenge each other)

AI, Health Equity, and Ethics Symposium
January 16th, 2025 | 10:00AM - 3:30PM | E90-1201
This event calls for epistemic humility and diverse perspectives, alongside meaningful community involvement, to transform the way we think, learn, and collaborate to prevent AI from perpetuating existing disparities.
Keynote Speaker, Panel Discussions, and Workshops
Lunch Provided
Institute Community & Equity Office
PKG
PRISCILLA KING GRAY PUBLIC SERVICE CENTER
Critical Data
imes
INSTITUTE FOR MEDICAL ENGINEERING & SCIENCE

What do we teach about AI?



Human-AI Systems Embedded with Humility, Curiosity & Joy from Learning

Connection Science

Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-2010)

Jürgen Schmidhuber

ISSN: 0954-0091 (Print) 1360-0494 (Online) Journal homepage: www.tandfonline.com/journals/ccos20

Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts

Jürgen Schmidhuber

SSRN

Gender Diversity in AI Research

15 Pages • Posted: 1 Aug 2019

Konstantinos Stathoulopoulos

National Endowment for Science, Technology and the Arts (NESTA)

Juan C Mateos-Garcia

National Endowment for Science, Technology and the Arts (NESTA)

AINOW

2023 Landscape Our Work People Careers About Us

REPORT

Disability, Bias, and AI – Report

AI Now Institute Nov 20, 2019

NBER | NATIONAL BUREAU of ECONOMIC RESEARCH

Subscribe | Media | Open Calls | Summer Institute 2025

Login

Research Programs & Projects Conferences Affiliated Scholars NBER News Career Resources About

Search

Home › Research › Working Papers › Learning, Diversity and Adaptation in...

Learning, Diversity and Adaptation in Changing Environments: The Role of Weak Links

Daron Acemoglu, Asuman Ozdaglar & Sarath Pattathil

Health AI Systems Thinking for Community (HASTE)

Every system
is perfectly
designed to
achieve
exactly the
results it gets