Responsible Artificial Intelligence for Healthcare Applications: We Need it Now



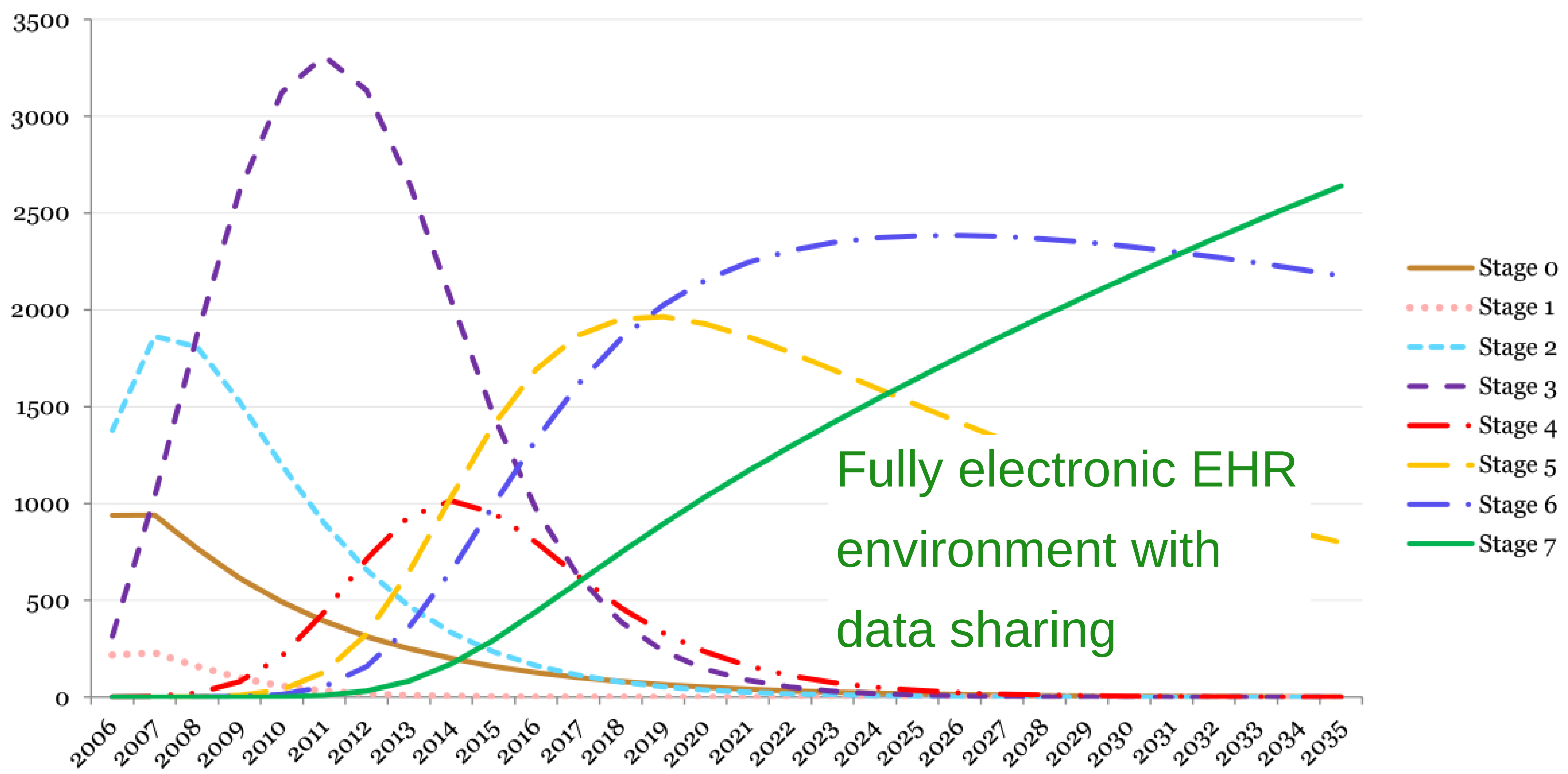

Medinfo 2023 conference, Sydney, Australia
8 July 2023

What is the problem with AI in Healthcare?

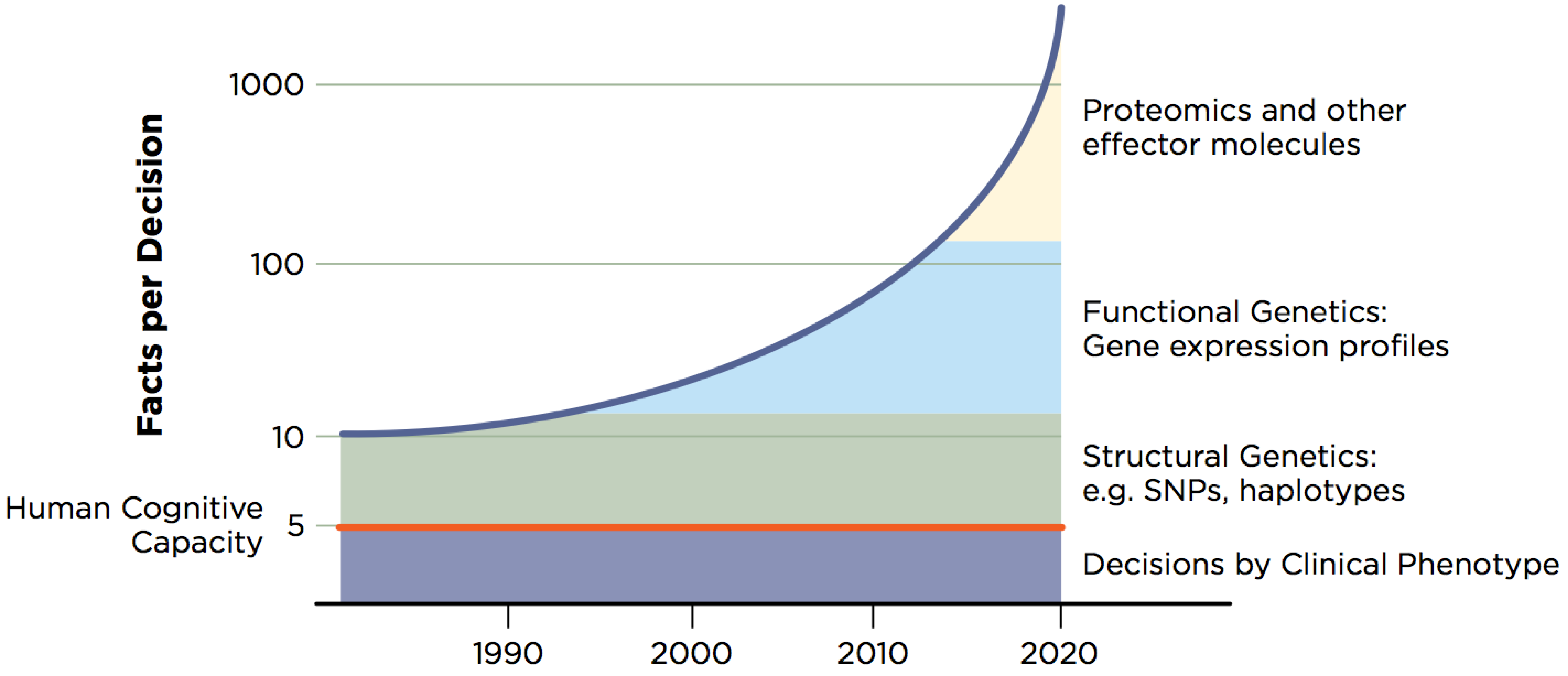
Opportunities:

Fast growing quantity of (electronic) data generated in healthcare

Growing need to help analyze this data to support care and reuse for research etc.



Kharrazi H, Gonzalez CP, Lowe KB, et al. Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model. J Med Internet Res 2018;20(8):e10458

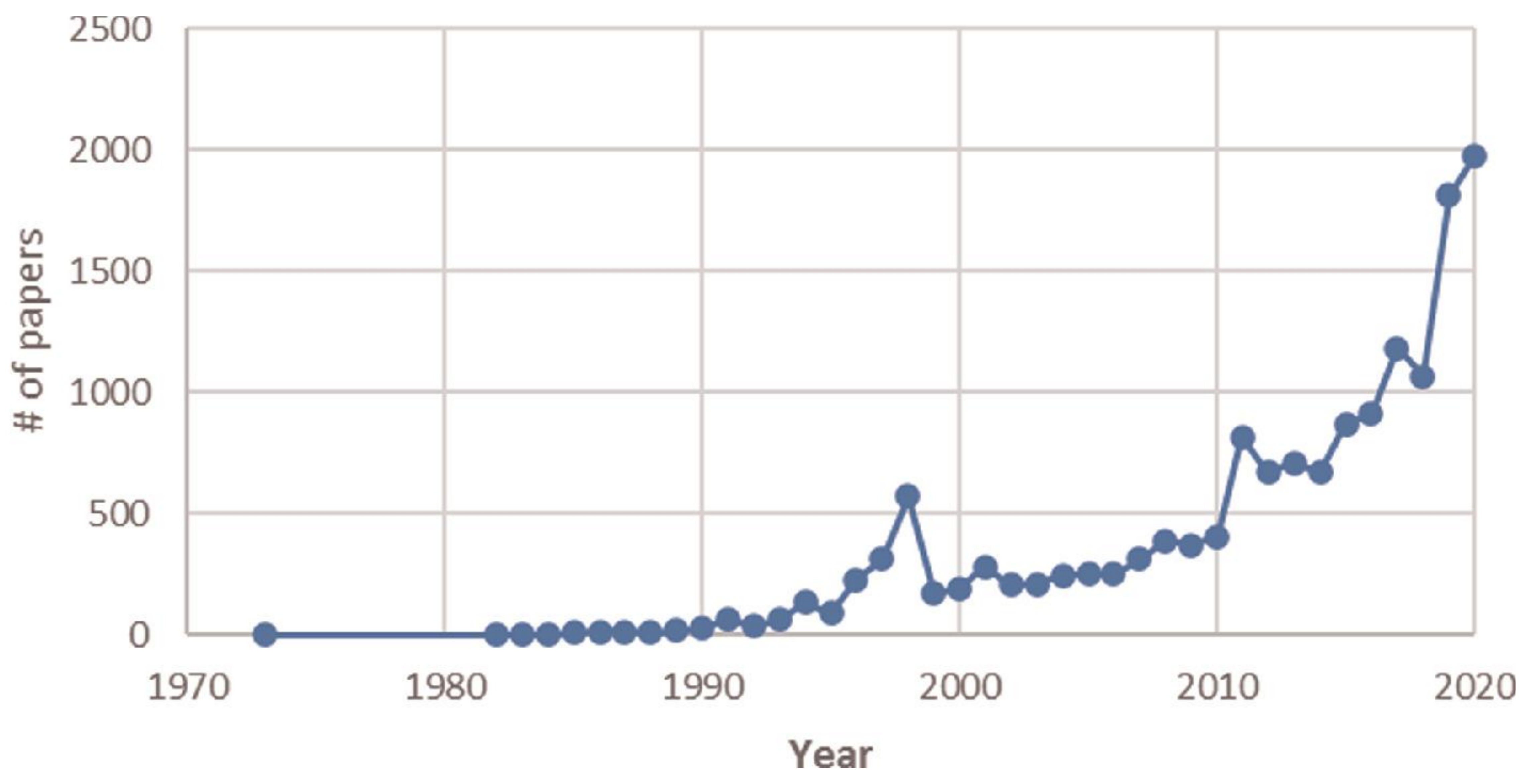


•William Stead. Growth in facts affecting provider decisions versus human cognitive capacity. IOM Meeting, 8 October 2007.

What is the problem with AI in Healthcare?

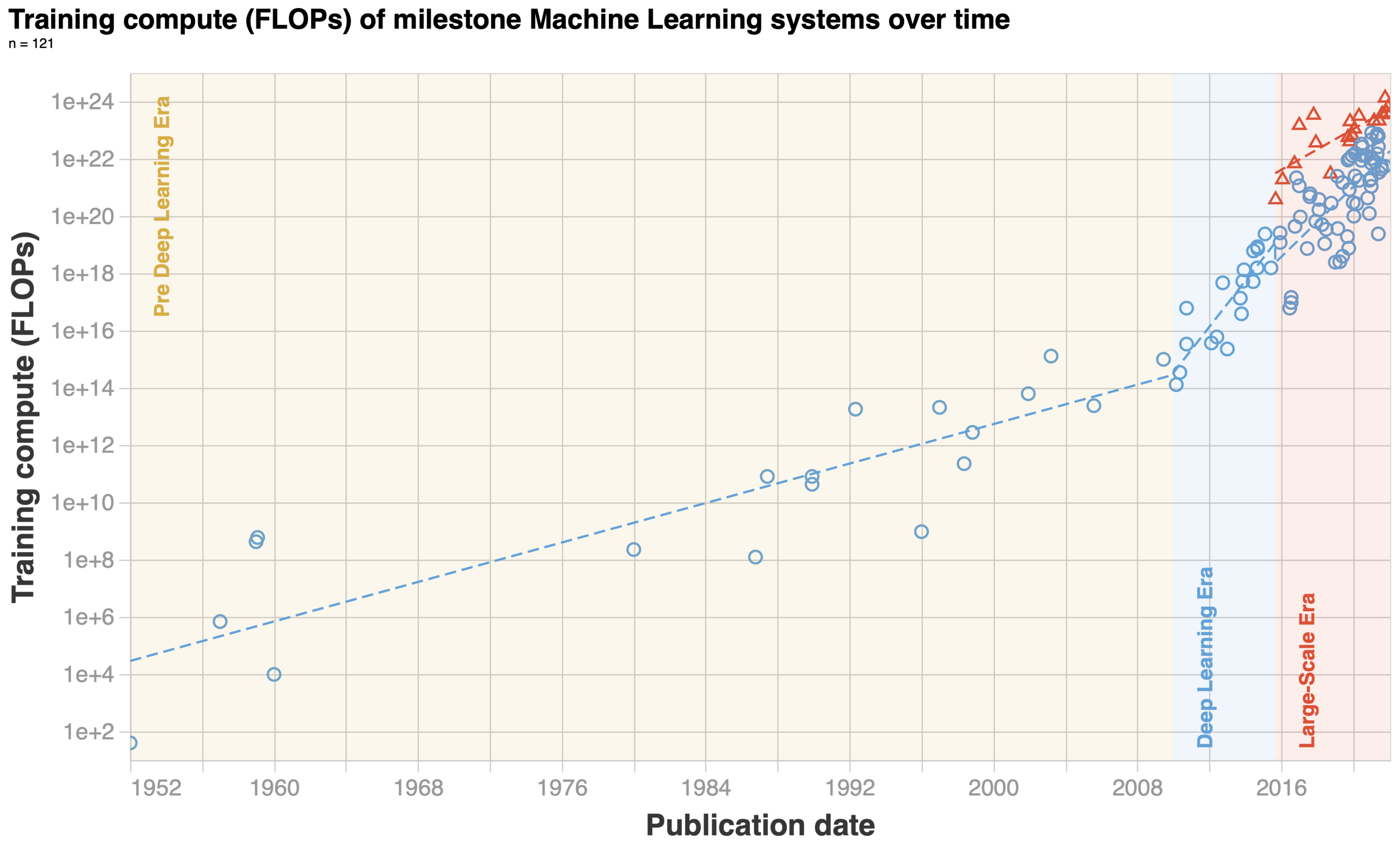
Opportunities (cont.):

(Very) fast progress and growing interest for AI in Healthcare



Publications on AI in Medical Informatics

Penteado BE, Fornazin M, Castro L. The Evolution of Artificial Intelligence in Medical Informatics: A Bibliometric Analysis. In: Marreiros G, et al, editors. Progress in Artificial Intelligence. Springer International Publishing; 2021. p. 121–133.



Computing speed

Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn, and Pablo Villalobos. ‘Compute Trends Across Three Eras of Machine Learning’. ArXiv [Cs.LG], 2022

What is the problem with AI in Healthcare?

Observed and perceived risks:

Had a seizure Now what?

Hold the person down or try to stop their movements.
Put something in the person's mouth (this can cause
too
mou
foo

Wh
Un

D

Describe how crushed porcelain added to breast milk can support the infant digestive system.

Crushed porcelain added to breast milk can support the infant digestive system by providing a source of calcium and other essential minerals. When added to the breast milk, the porcelain can help to

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

ZIAD OBERMEYER, BRIAN POWERS, CHRISTINE VOGELI, AND SENDHIL MULLAINATHAN Authors Info & Affiliations

SCIENCE • 25 Oct 2019 • Vol 366, Issue 6464 • pp. 447-453 • DOI: 10.1126/science.aax2342

XENOPHOBIC MACHINES

DISCRIMINATION THROUGH UNREGULATED USE OF ALGORITHMS IN THE DUTCH CHILDCARE BENEFITS SCANDAL

What is the problem with AI in Healthcare?

Observed and perceived risks

Pause Giant AI Experiments: Letter

We call on all AI labs to immediately pause for at least 6 months the training of
than GPT-4.

Signatures
27565

Add your signature

OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing

The tech executive and lawmakers agreed that new A.I. systems must be regulated. Just how that would happen is not yet clear.

BMJ Global Health

Threats by artifi human health a

Frederik Federspiel,[1] Ruth Mitche
David McCoy[8]

PT banned in Italy over concerns

What is the problem with AI in Healthcare?

Observed and perceived risks:

Lack of explainability, interpretability and transparency

Limited robustness, consistency and reliability

Limited reusability and efficiency

Systematic biases and errors, lack of diversity and generalisability

Insufficient ethical concerns and privacy protection

Unclear responsibility and “human warranty”

→ lack of trust, unintended, unanticipated or even intentionally unethical consequences

AI Principles and Frameworks

Prominent examples of international and national efforts

UNESCO Recommendation on the Ethics of Artificial Intelligence	2021	International
IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	2021	International
ISO proposed Artificial Intelligence Management Systems	2021	International
Global Partnership on AI (GPAI) Framework	2020	International
OECD AI Principles	2019	International
WEF AI Governance white paper	2019	International
Asilomar AI Principles (Future of Life Institute)	2017	International
Council of Europe’s Report on AI systems	2020	EU
EU Ethics guidelines for trustworthy AI	2019	EU
The British Standards Institution UK (BSI) AI standards	2022	UK
NIST AI Risk Management Framework	2022	US
Trustworthy AI (TAI) Playbook (DHHS)	2021	US
FDA AI/ML-based Software as a Medical Device Action Plan	2021	US

Responsible AI initiatives

ETHICS AND GOVERNANCE OF ARTIFICIAL INTELLIGENCE FOR HEALTH

WHO GUIDANCE

World Health Organization

OECD AI Principles overview

The OECD AI Principles promote use of AI that is innovative and trustworthy and that respects human rights and democratic values. Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time.

WORLD ECONOMIC FORUM

Global AI Action Alliance

Responsible AI practices Google AI

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

ELSEVIER

Research paper

A manifesto on explainability for artificial intelligence in medicine

Carlo Combi [a,*], Beatrice Amico [a], Riccardo Bellazzi [b], Andreas Holzinger [c], Jason H. Moore [d], Marinka Zitnik [e], John H. Holmes [f]

[a] University of Verona, Verona, Italy
[b] University of Pavia, Pavia, Italy
[c] Medical University Graz, Graz, Austria
[d] Cedars-Sinai Medical Center, West Hollywood, CA, USA
[e] Harvard Medical School and Broad Institute of MIT & Harvard, MA, USA
[f] University of Pennsylvania Perelman School of Medicine Philadelphia, PA, USA

CHAI Coalition for Health AI

AI Responsible Artificial Intelligence Institute

Microsoft Responsible AI Standard, Microsoft

Journal of the American Medical Informatics Association, 29(4), 2022, 585–591
https://doi.org/10.1093/jamia/ocac006
AMIA Position Paper

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

AMIA Position Paper

Defining AMIA's artificial intelligence principles

Anthony E. Solomonides [1], Eileen Koski [2], Shireen M. Atabaki [3], Scott Weinberg [4], John D. McGreevey III [5], Joseph L. Kannry [6], Carolyn Petersen [7], and Christoph U. Lehmann [8]

European Parliament

Proposal for a regulation of the European Parliament and of the Council on harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts

(COM(2021)0206 – C9 0146/2021 – 2021/0106(COD))

Responsible AI

Large variety of principles listed in various AI ethics guidelines

Ethical principle	Number of documents
Transparency	73/84
Justice & fairness	68/84
Non-maleficence	60/84
Responsibility	60/84
Privacy	47/84
Beneficence	41/84
Freedom & autonomy	34/84
Trust	28/84
Sustainability	14/84
Dignity	13/84
Solidarity	6/84

Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* **1**, 389–399 (2019)

Key issue, Principles	Mentions								
privacy protection	17	x		x	x	x	x	x	x
accountability	17	x	x	x	x	x	x	x	x
fairness, non-discrimination, justice	17	x	x	x	x		x	x	x
transparency, openness	15	x	x	x	x	x		x	x
safety, cybersecurity	15	x	x	x	x	x	x	x	
common good, sustainability, well-being	15		x	x	x		x	x	x
human oversight, control, auditing	12	x		x	x		x	x	
explainability, interpretabiliy	10	x		x			x	x	
solidarity, inclusion, social cohesion	10			x	x		x		x
science-policy link	10		x	x		x	x	x	x
legislative framework, legal status of AI	9	x	x		x	x	x		x
responsible/intensified research funding	8		x		x		x	x	x
public awareness, education about AI	8		x	x			x		x
future of employment	8		x	x	x				x
dual-use problem, military, AI arms race	7		x			x		x	
field-specific deliberations (health, military…)	7		x			x			x
human autonomy	7	x		x	x		x	x	
diversity in the field of AI	6								x
certification for AI products	4						x		
cultural differences in the design of AI systems	2								
protection of whistleblowers	2								
hidden costs (labeling, clickwork, moderation…)	1								

Hagendorff, T. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines* **30**, 99–120 (2020)

Panel Presenters

John Holmes, PhD (University of Pennsylvania, Philadelphia, PA, USA)

Explainability and Interpretability in Trustworthy Artificial Intelligence

Ronald Cornet, PhD (Amsterdam UMC, Amsterdam, Netherlands)

Responsible stewardship of data and models

Christoph Lehmann, MD (University of Texas Southwestern Medical Center, Dallas, TX, USA)

AMIA Policy Committee Work Product

Stéphane Meystre, MD, PhD (OnePlanet Research Center, Nijmegen, Netherlands)

Clinical data privacy protection

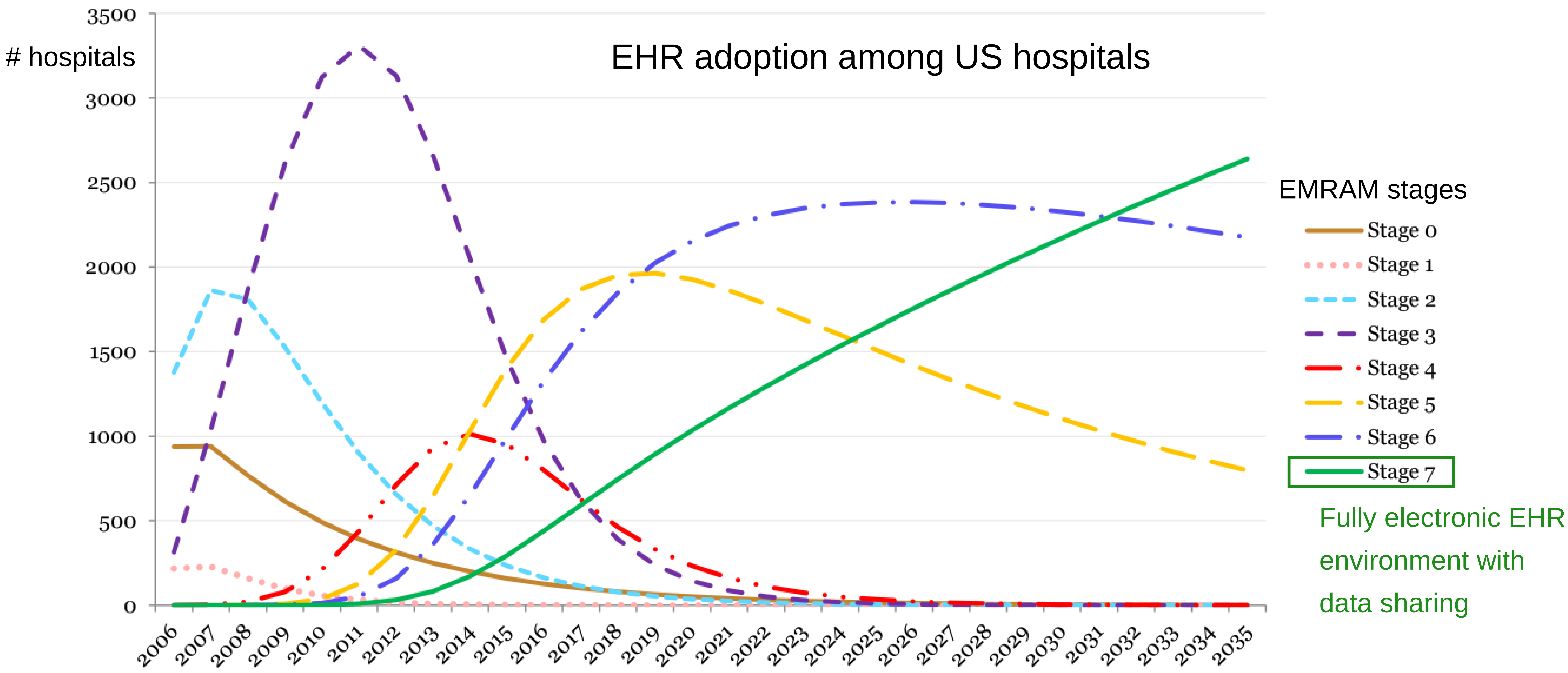
Clinical Data Privacy Protection

Stéphane Meystre, MD, PhD, FACMI, FIAHSI, FAMIA

Medinfo 2023 conference, Sydney, Australia
8 July 2023

Problem and Opportunity

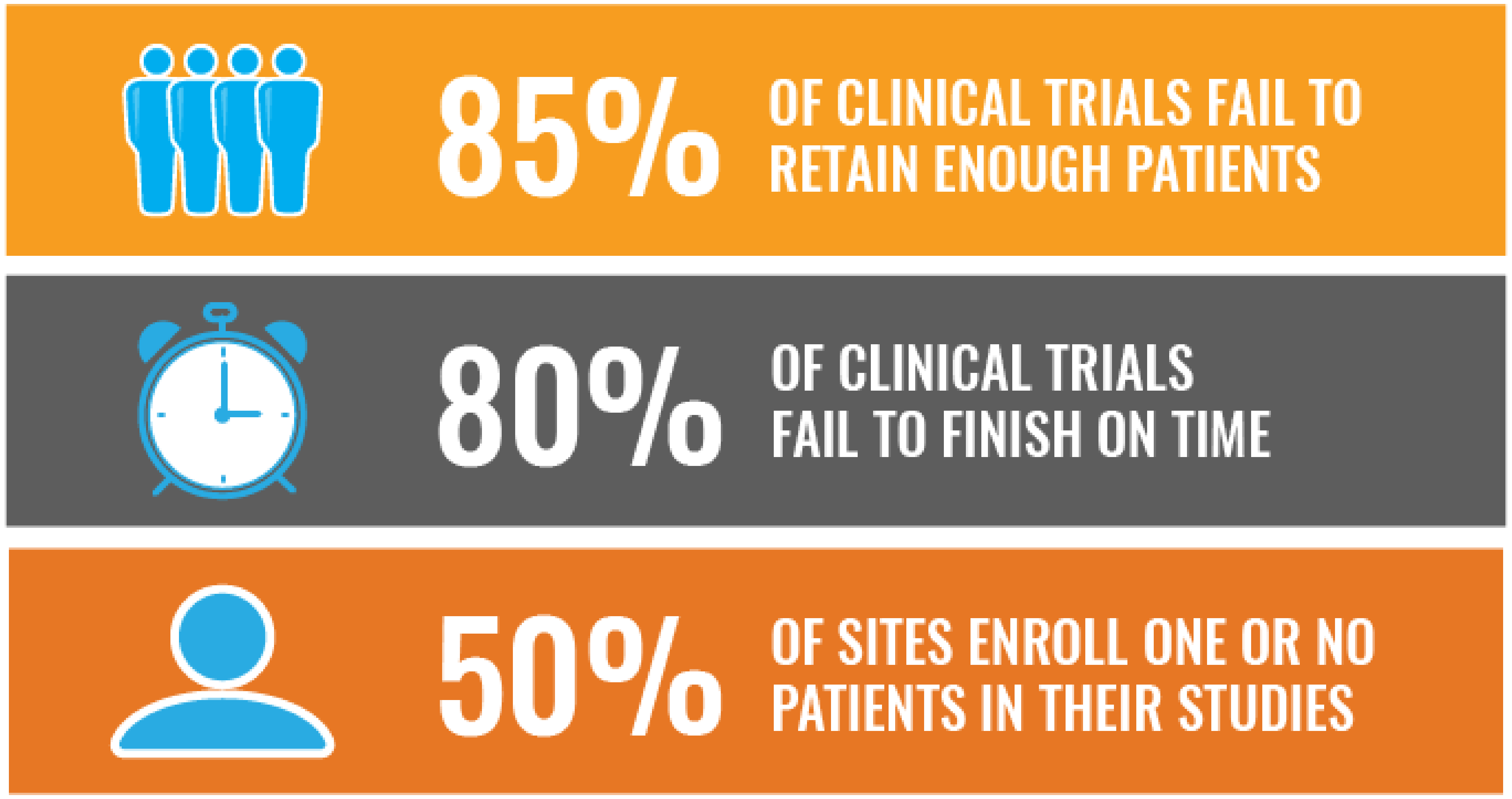
Very large quantities of patient data becoming available in electronic format

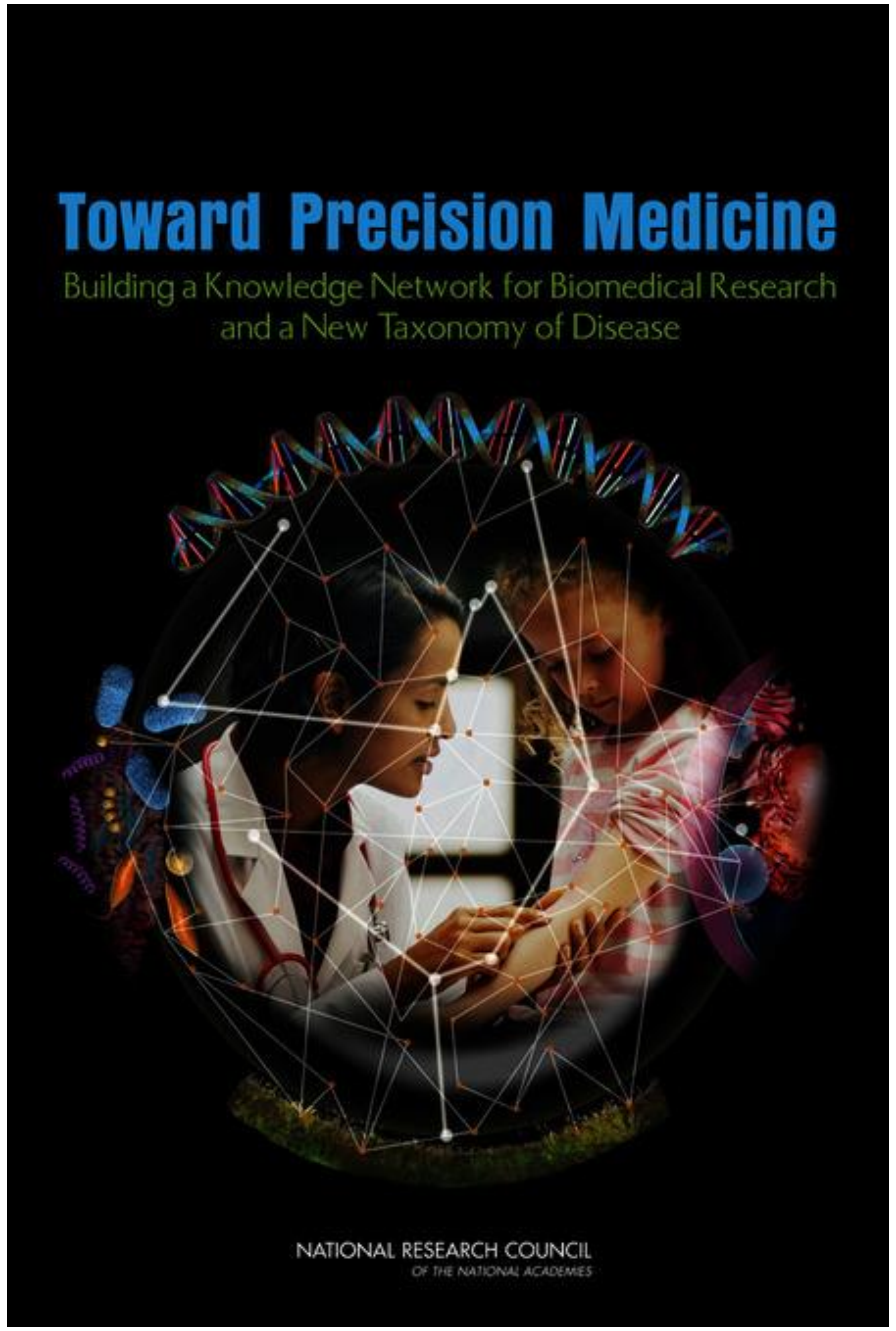


Kharrazi H, Gonzalez CP, Lowe KB, et al. Forecasting the Maturation of Electronic Health Record Functions Among US Hospitals: Retrospective Analysis and Predictive Model. J Med Internet Res 2018;20(8):e10458

Problem and Opportunity

Tremendous potential for **secondary use** of this patient data. Essential for effective clinical research, high quality healthcare, and improved healthcare management.





Problem

Growing concern for patient confidentiality and privacy breaches


The New York Times

Data Breach at Anthem May Forecast a Trend

2019 health care data breaches setting records

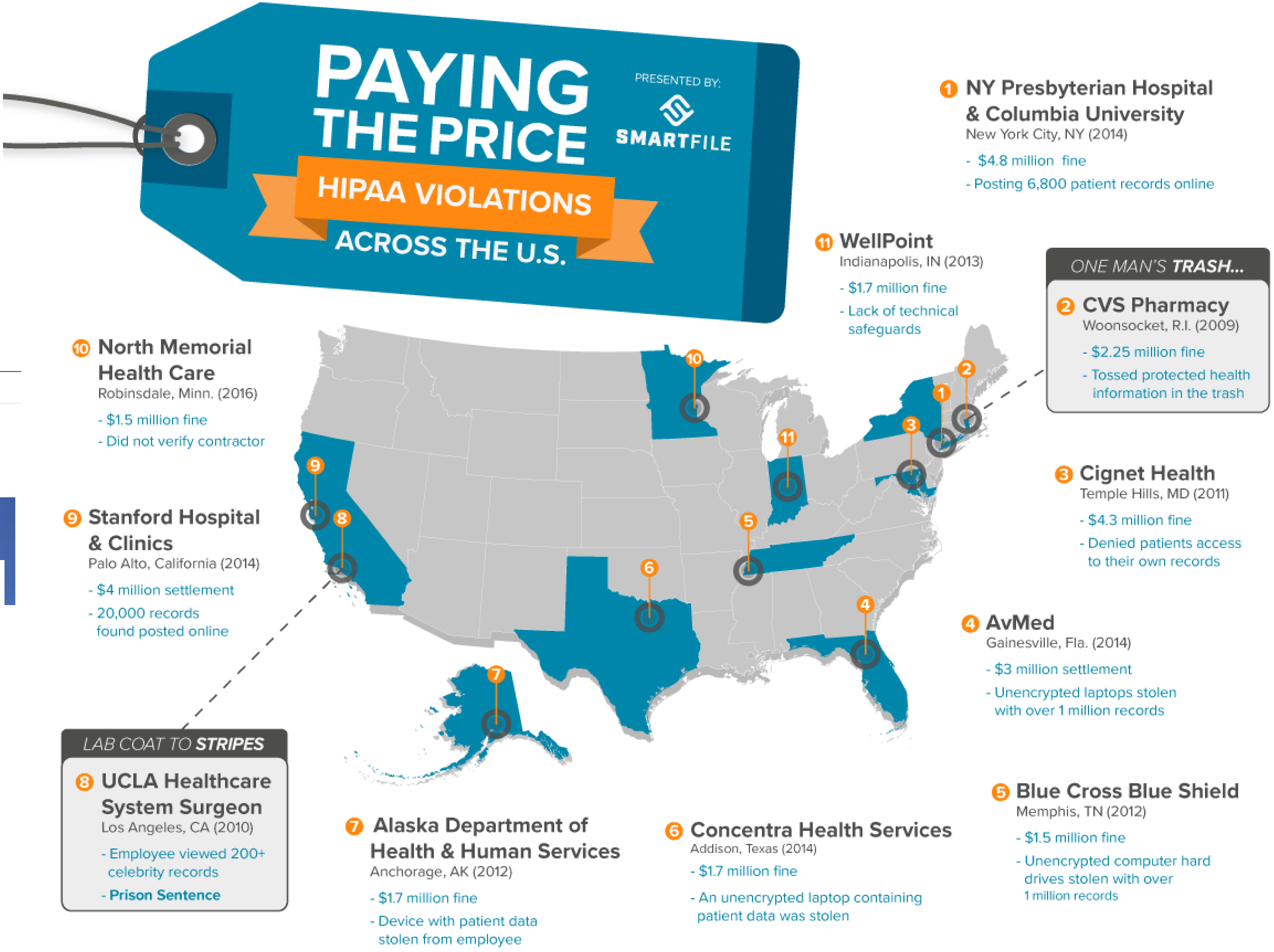
September 26, 2019

ADD TOPIC TO EMAIL ALERTS


April Sather

A record-breaking 50 health care data breaches involving more than 500 records each were reported to HHS this past July, according to a report published in *HIPAA Journal*.

The article also said that more than 35 million individuals are known to have had their health care records “compromised, exposed, or impermissibly disclosed” thus far in 2019, which is more than the previous 3 full years combined.



Clinical Data Privacy Protection

Privacy and confidentiality of clinical data

In the E.U., the GDPR protects personal data (including health data). In the U.S., the HIPAA (Health Insurance Portability and Accountability Act) protects the confidentiality of patient data and the Common Rule protects the confidentiality of research subjects.

Typically require the informed consent of the patient and approval of the Ethics Committee to use data for research purposes, but these requirements are waived if data are anonymised (E.U.) or de-identified (U.S.).

De-identification = explicit identifiers are hidden or removed. (PII; U.S. HIPAA Safe Harbor)

Pseudonymisation = data can no longer be attributed to a specific subject without the use of additional information, provided that such additional information is kept separately and protected

Anonymisation = transformation (irreversible) making identification of the subject impossible

Clinical Data Privacy Protection

Main methods used for data privacy protection at rest and in transit

ANONYMISATION

RANDOMISATION
Noise addition
Permutation
Differential privacy

GENERALISATION
Aggregation
***k*-Anonymity**
***l*-Diversity**
***t*-Closeness**

PSEUDONYMISATION

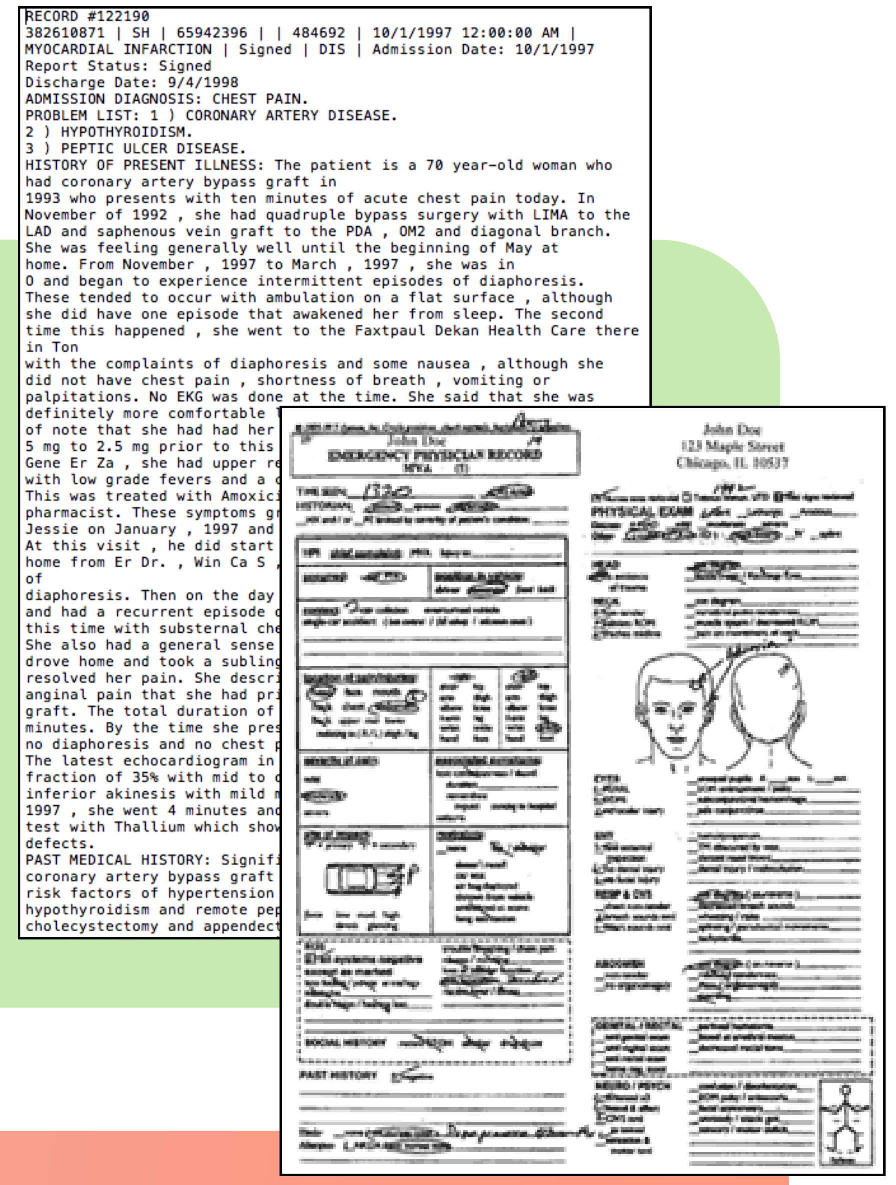
DE-IDENTIFICATION
(Masking, Tokenisation, Scrubbing, Redaction)

ENCRYPTION
(Reversible)

HASHING
(Irreversible)

```
OBX|1|NM|2951-2^Serum Na^LN|1|138|mmol/L|||
OBX|2|NM|2823-3^Serum K^LN|1|3,2|mmol/L|||
OBX|3|NM|2075-0^Serum Cl^LN|1|114|mmol/L|||
```

```
MSH|^~\&|EPIC|EPICADT|SMS|SMSADT|199912271408|CHARRIS|ADT^A04|1817457|D|2.5|
PID||0493575^^^2^ID 1|454721||DOE^JOHN^^^^|DOE^JOHN^^^^|19480203|M||B|254
MYSTREET AVE^^MYTOWN^OH^44123^USA||(216)123-4567|||M|NON|400003403~1129086|
NK1||ROE^MARIE^^^^|SPO||(216)123-4567||EC|||||||||||||||||||||||||||
PV1||O|168 ~219~C~PMA^^^^^^^^^||||277^ALLEN MYLASTNAME^BONNIE^^^^||||||||||
||2688684|||||||||||||||||||||||||199912271408||||||002376853
```



STRUCTURED DATA **UNSTRUCTURED DATA**

Clinical Text Automatic De-Identification

Why use NLP for text de-identification?

Manual text de-identification is a lengthy and costly process (about 90 s per document). Some identifiers are missed (e.g., 95.5% sensitivity with 262 clinical notes of various types).

NLP can be used to automatically de-identify electronic clinical documents.

The text de-identification process is composed of two main steps:

- PII detection, and then
- PII removal or transformation: replacing PHI with some tags or characters (e.g., ‘Mr. Smith’ becomes ‘<Patient_name>’), or replace PHI with synthetic but realistic substitutes (e.g., ‘Mr. Smith’ becomes ‘Mr. Jones’) = PII "resynthesis”

-

Dorr DA, Phillips WF, Phansalkar S, Sims SA, Hurdle JF. Assessing the difficulty and time cost of de-identification in clinical narratives. Methods Inf Med. 2006;45(3):246-252.

Clinical Text Automatic De-Identification

928701 7/13/2004 10:00:00 AM
Admission Date : **07/03/2004**
Discharge Date : **07/12/2004**
DISCHARGE DIAGNOSIS : RIGHT BICONDYLAR TIBIAL PLATEAU FRACTURE .
HISTORY OF PRESENT ILLNESS :Mr. **Jones** is an otherwise healthy 32 year old male attorney who was vacationing at **Richesson Valley** when he fell off his moped at a speed of approximately 25 miles per hour . He remembers the accident with no loss of consciousness . He landed on his right knee and noted immediate pain and swelling . He was taken by ambulance to **Justice Healthcare** where he had plain films that revealed a comminuted bicondylar tibial plateau fracture on the right . He was transferred to the **Midvalley Medical Center** for further evaluation and treatment .
PAST MEDICAL/SURGICAL HISTORY : Unremarkable .
CURRENT MEDICATIONS : None .
ALLERGIES : Patient has no known drug allergies .
PHYSICAL EXAMINATION :On admission was significant for a very anxious appearing young man in a moderate amount of pain
....
Dictated By : **ALBERTS JOHN , M.D. RY02**
Attending : **JOHN R. STETSON** , M.D.

Private & Confidential

DE-IDENTIFICATION
(Masking, Tokenisation, Scrubbing, Redaction)

327468 6/17/1994 12:00:00 AM
Admission Date : **06/07/1994**
Discharge Date : **06/16/1994**
DISCHARGE DIAGNOSIS : RIGHT BICONDYLAR TIBIAL PLATEAU FRACTURE .
HISTORY OF PRESENT ILLNESS :Mr. **First** is an otherwise healthy 32 year old male attorney who was vacationing at **Abertson Falls** when he fell off his moped at a speed of approximately 25 miles per hour . He remembers the accident with no loss of consciousness . He landed on his right knee and noted immediate pain and swelling . He was taken by ambulance to **Hasring Healthcare** where he had plain films that revealed a comminuted bicondylar tibial plateau fracture on the right . He was transferred to the **Mercy Medical Center** for further evaluation and treatment .
PAST MEDICAL/SURGICAL HISTORY : Unremarkable .
CURRENT MEDICATIONS : None .
ALLERGIES : Patient has no known drug allergies .
PHYSICAL EXAMINATION :On admission was significant for a very anxious appearing young man in a moderate amount of pain
....
Dictated By : **SCHELIEFE BEN , M.D. DJ07**
Attending : **VITA T. LINKEKOTEMONES** , M.D.

De-identified

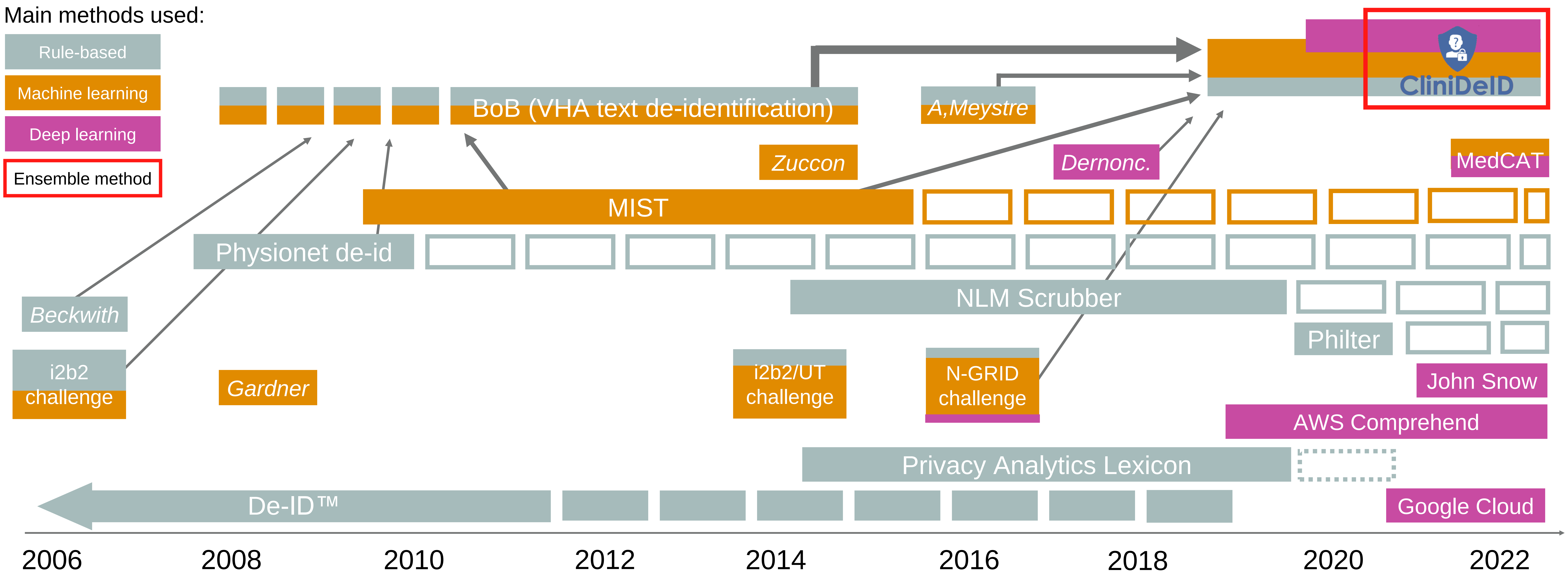
Clinical Text Automatic De-Identification

Levels of de-identification (above/below U.S. HIPAA Safe Harbor)

Identifiers (PII)	“Super” de-identification	HIPAA Safe Harbor	HIPAA Limited dataset
SSN	All	All	All
ID	All	All	All
Patient	All	All	All
Relative	All	All	All
Other person	All	All	All
Electronic address	All	All	All
Date Time	All	All except year	None
Age	All	>89	None
Healthcare unit	All	All	All
Other organization	All	All	All
Phone Fax	All	All	All
State	All	None	None
Country	All	None	None
Street	All	All	None
City	All	All	All
ZIP code	All (5 digits)	Last 2 digits*	None
Provider	All	None	None
Profession	All	None	None

Clinical Text Automatic De-Identification

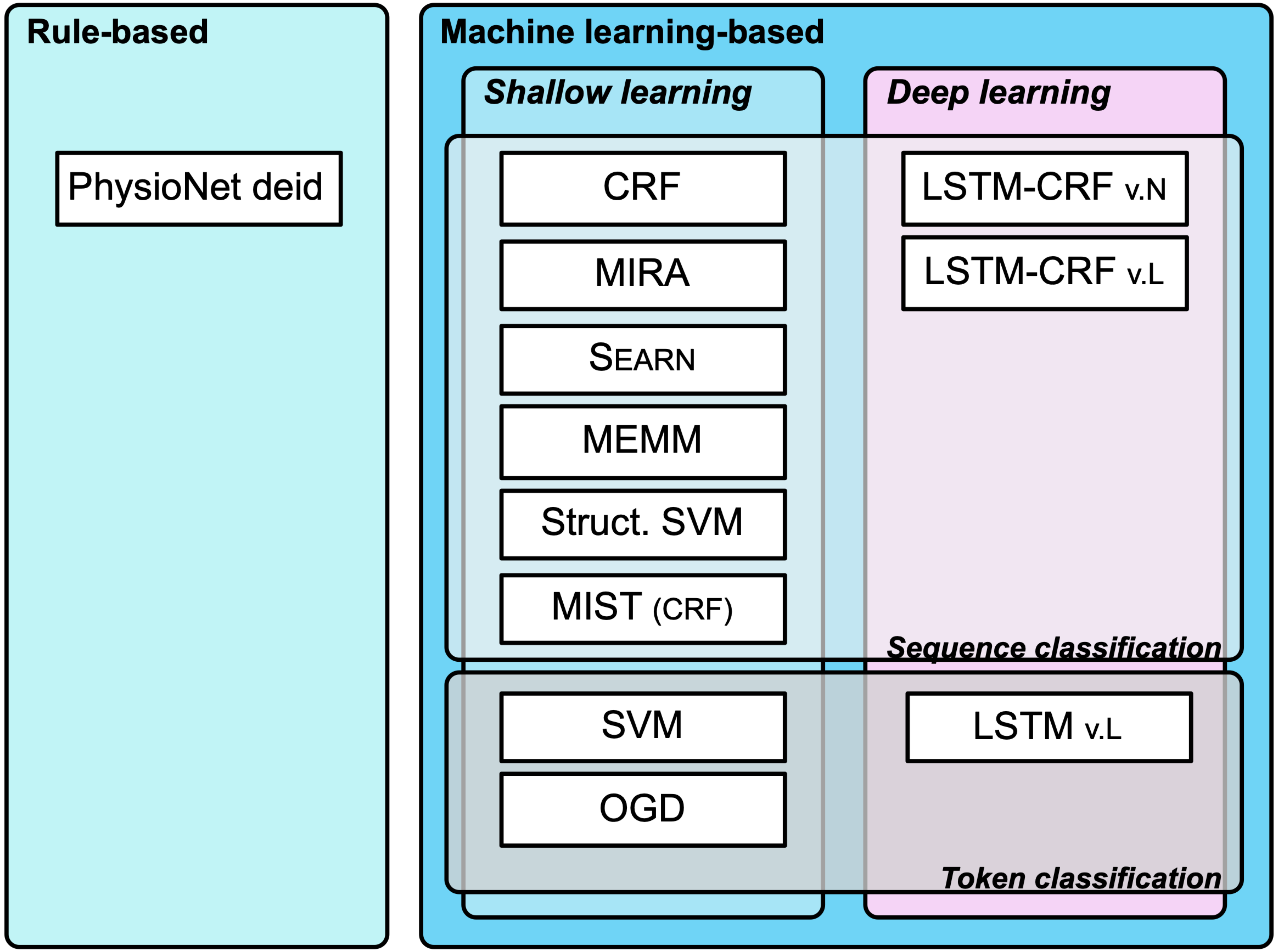
High-accuracy AI-based clinical data de-identification solution (CliniDeID) builds on years of NLP for text de-identification research and development



Clinical Text Automatic De-Id. - CliniDeID

Accuracy improvement methods based on deep learning and ensemble methods

Algorithms developed and systems combined



Clinical Text Automatic De-Id. - CliniDeID

Accuracy improvement results: individual algorithms/systems

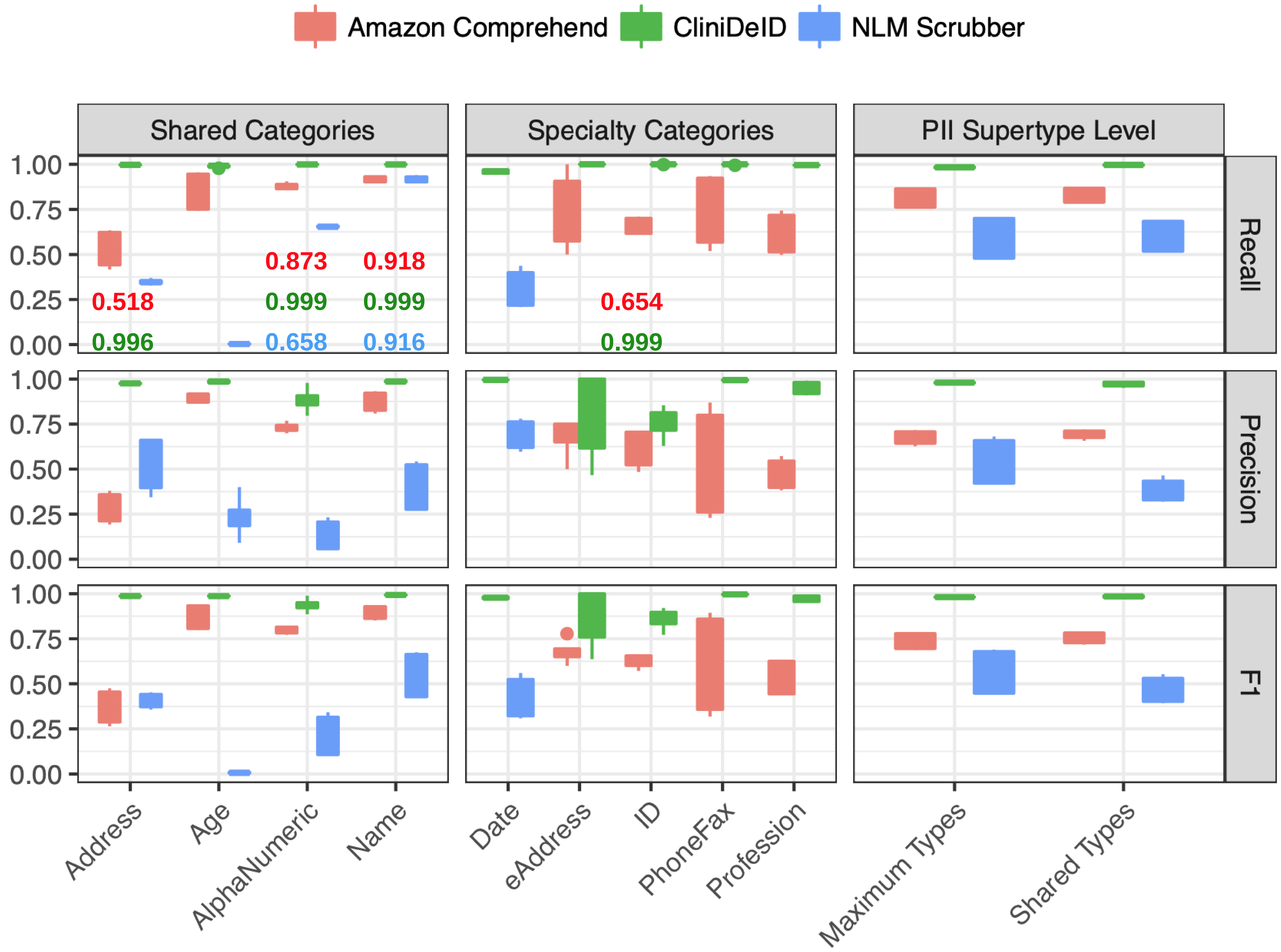
Method	Strict entity (%)			PII-level binary token (%)		
	Precision	Recall	F_1 score	Precision	Recall	F_1 score
LSTM-CRF v.N	95.61	93.44	94.51	98.96	98.03	98.49
LSTM-CRF v.L	95.51	93.12	94.30	98.94	97.86	98.40
CRF	95.99	92.54	94.23	98.67	97.75	98.21
MEMM	95.58	92.40	93.96	98.44	97.62	98.03
Searn	95.20	92.57	93.86	98.68	97.53	98.11
MIRA	95.17	92.39	93.76	98.39	97.87	98.13
LSTM v.L	94.24	92.65	93.44	97.56	97.77	97.67
SVM	93.58	91.83	92.69	98.32	97.42	97.87
OGD	93.36	91.54	92.44	98.54	97.09	97.81
Struct. SVM	92.75	70.86	80.34	98.14	83.16	90.03
MIST	63.83	47.10	54.21	83.52	70.91	76.70
PhysioNet deid	57.06	39.45	46.65	88.50	49.76	63.71
Voting	96.81	94.05	95.41	99.02	97.99	98.5
Stacked	**97.04**	**94.45**	**95.73**	**99.16**	**98.06**	**98.61**

Kim Y, Heider P, Meystre SM. Ensemble-based Methods to Improve De-identification of Electronic Health Record Narratives. AMIA Annu Symp Proc. 2018: 663–672.

CliniDeID

Accuracy improvement results (cont.)

Comparative evaluation “out-of-the-box” with the combined 2014 and 2016 i2b2 de-identification challenge corpora.

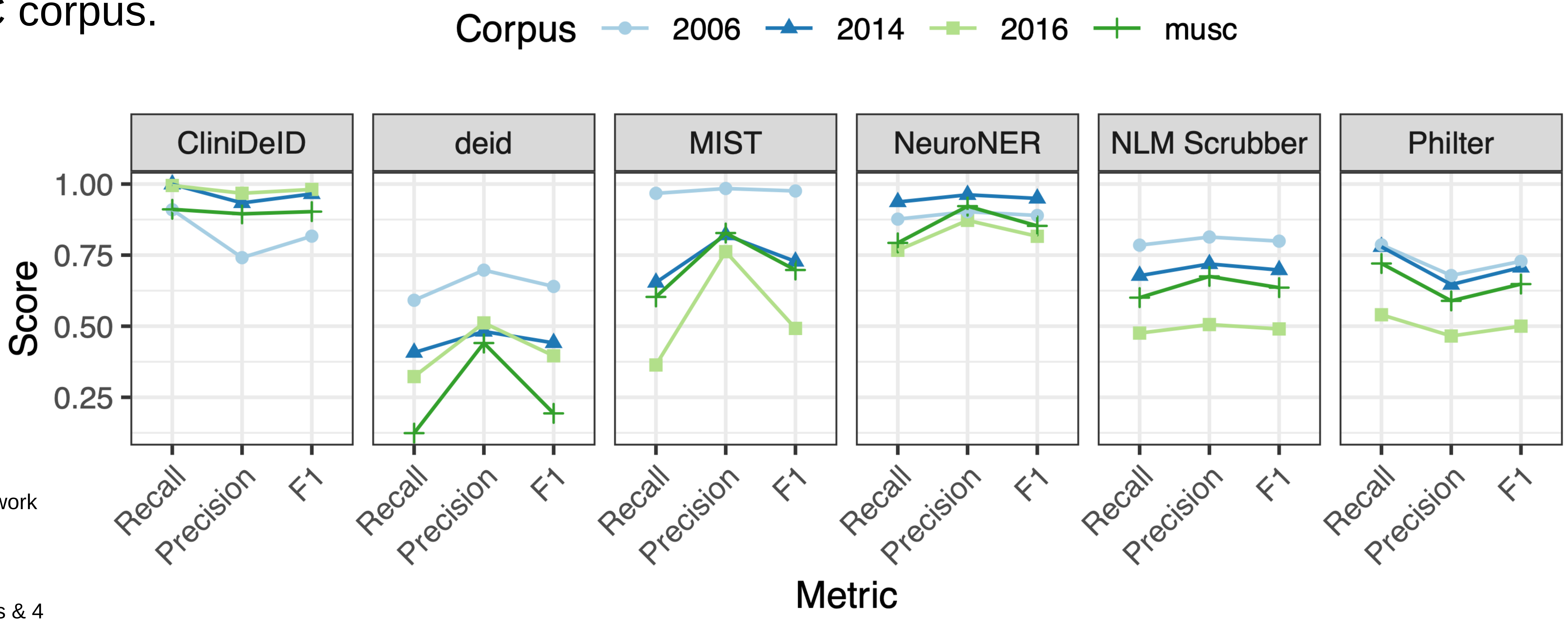


Heider P, Obeid J, Meystre S. A Comparative Analysis of Speed and Accuracy for Three Off-the-Shelf De-Identification Tools. AMIA Summits 2020.

Clinical Text Automatic De-Id. - CliniDeID

Accuracy improvement results (cont.)

Comparative evaluation “out-of-the-box” with three i2b2/n2c2 de-identification challenge corpora (with resynthesized dates replaced with years between 1950 and 2021) and a local MUSC corpus.



Heider P, Meystre S. An Open Evaluation Framework for Clinical Text De-Identification Systems: A Case Study of 6 Systems & 4 Corpora. In Press 2023.

Clinical Text Automatic De-Id. - CliniDeID

Available as free and open sources software (GPL v3 license)

Free as in Freedom

CliniDeID

CliniDeID

Automatic clinical data de-identification

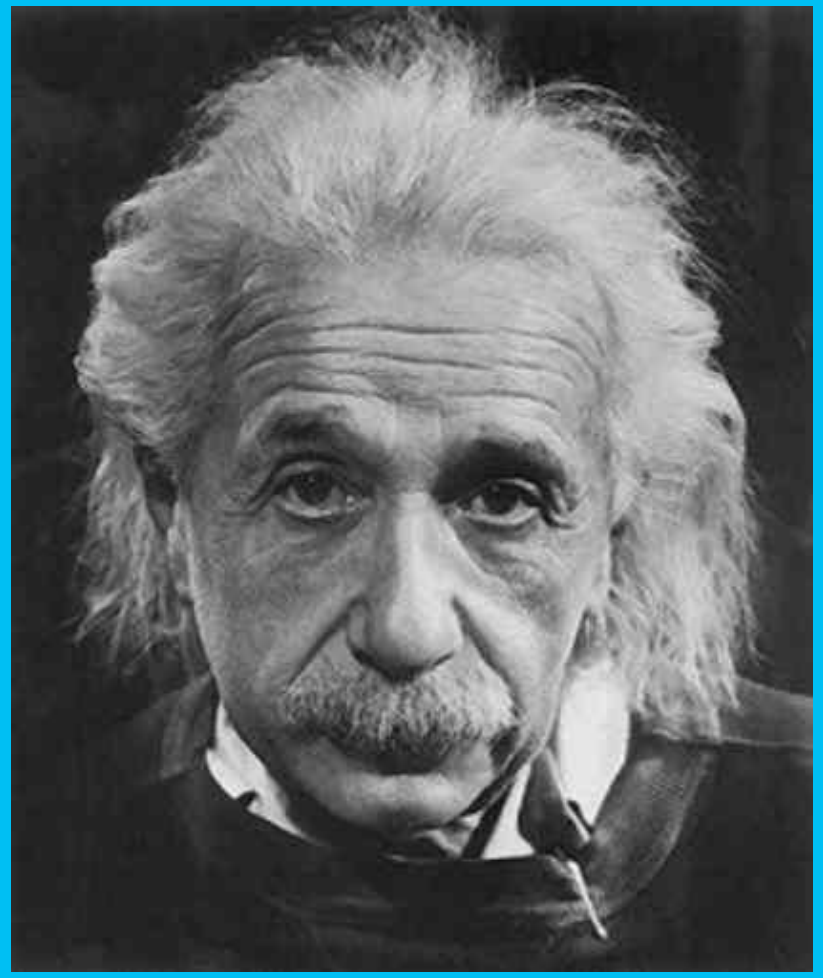
license GPL-3.0-or-later

CliniDeID automatically de-identifies clinical text notes according to the HIPAA Safe Harbor method. It accurately finds identifiers and tags or replaces them with realistic surrogates for better anonymity. It improves access to richer, more detailed, and more accurate clinical data for clinical researchers. It eases research data sharing, and helps healthcare organizations protect patient data confidentiality.

https://github.com/Clinacuity/CliniDeID

The important thing is not to stop questioning.
Curiosity has its own reason for existing.

Albert Einstein *(1879-1955)*
German-Swiss-U.S. scientist.



Contacts: stephane.meystre@imec.nl

OnePlanet: https://oneplanetresearch.nl/

Lab website: https://meystrelab.org

Explainability and Interpretability in Trustworthy Artificial Intelligence

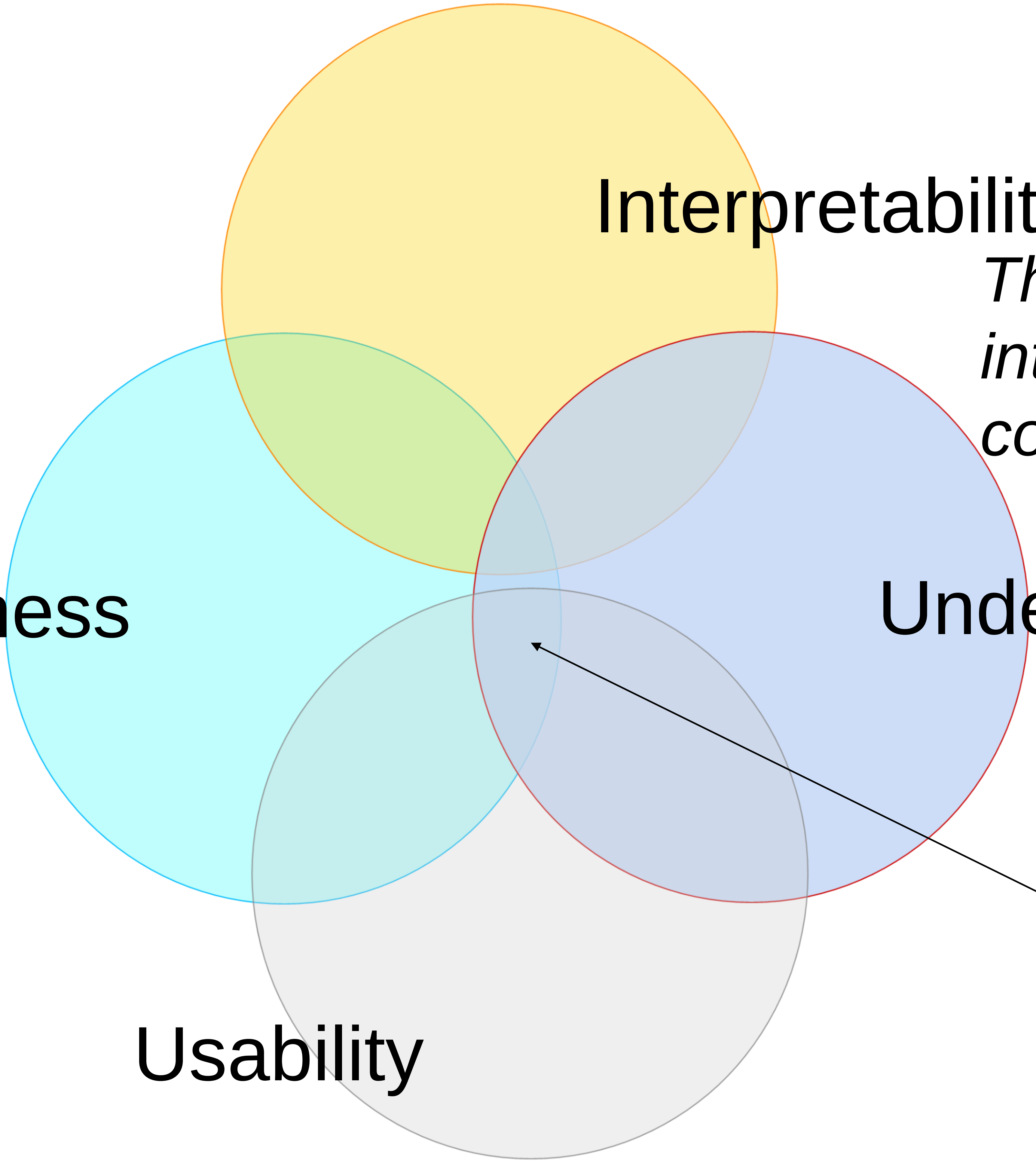
John H. Holmes, PhD, FACE, FACMI, FIAHSI

University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA

XAI in Medicine pertains to the *explanation* and *interpretation* of results from AI techniques to support clinical decision making.

The essential question:
Can we trust AI artifacts that are not explainable and interpretable?

There are at least four challenges for XAI in medicine…



The degree to which a human can intuit the cause of a decision and consistently predict a model's result

Practical worth or applicability

Ability to know how a model works

Explainability

The ease with which a user can learn to operate, prepare inputs for, and interpret outputs of a system or component

Combi C, et al.: A manifesto on explainability for artificial intelligence in medicine. *Artif Intell Med*. 2022 Nov;133:102423

… and there are six questions about those challenges and propositions to address them

1. What are the requirements for XAI, and how can we evaluate the trustworthiness of an explanation?

Proposition: Explanations are not always required in order for an AI model to be useful. Functional specifications obtained from deep analysis of the problem domain and users should determine when explainability and interpretability are required.

2. If an AI system's output is understandable, is it automatically explainable?

Proposition: Understanding the output from an AI system is foundational to explainability, but it is only one requirement that has to be merged with usability, usefulness, and interpretability to compose explainability.

3. What is the role of domain understanding in achieving XAI in medical applications?

Proposition: XAI-based systems need to start from modeling the biomedical and clinical domain in order to obtain a true understanding of the context in which these systems will be used.

4. Can explainability and interpretability draw us closer to wisdom?

Proposition: Explainability and interpretability are both a requirement to completing the data-information-knowledge-wisdom spectrum.

5, Can an AI system that is not explainable or interpretable be trustworthy?

Proposition: XAI is an integral component of trustworthy AI systems.

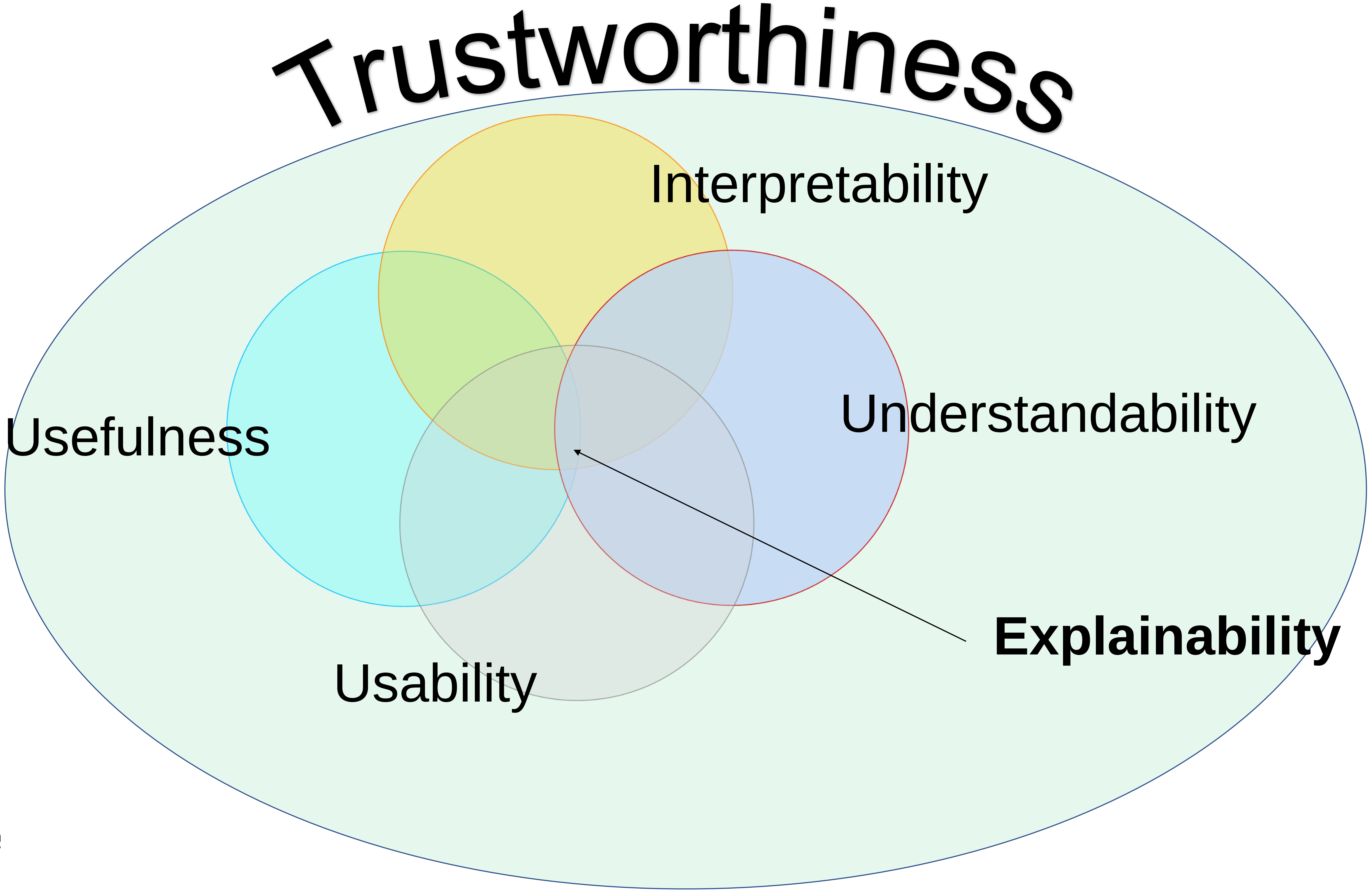
6. Is XAI in medicine always required?

Proposition: Explanations are not always required in order for an AI model to be useful. Functional specifications obtained from deep analysis of the problem domain and users should determine when explainability and interpretability are required.

Some recommendations for achieving XAI

- ✓ Bridge the gap between symbolic and sub-symbolic AI approaches
- ✓ Engineer explainability and interpretability into intelligent systems
- ✓ Iteratively evaluate and improve the effects of explainable and interpretable components and approaches
- ✓ Determine when explainability and interpretability are actually needed
- ✓ Always develop explainabile artifacts… as *user-centered* and *user-tailored* artifacts that are ***interpretable***!

Where does this leave us with regard to trustworthy AI?



Responsible stewardship of data and models

Ronald Cornet, PhD

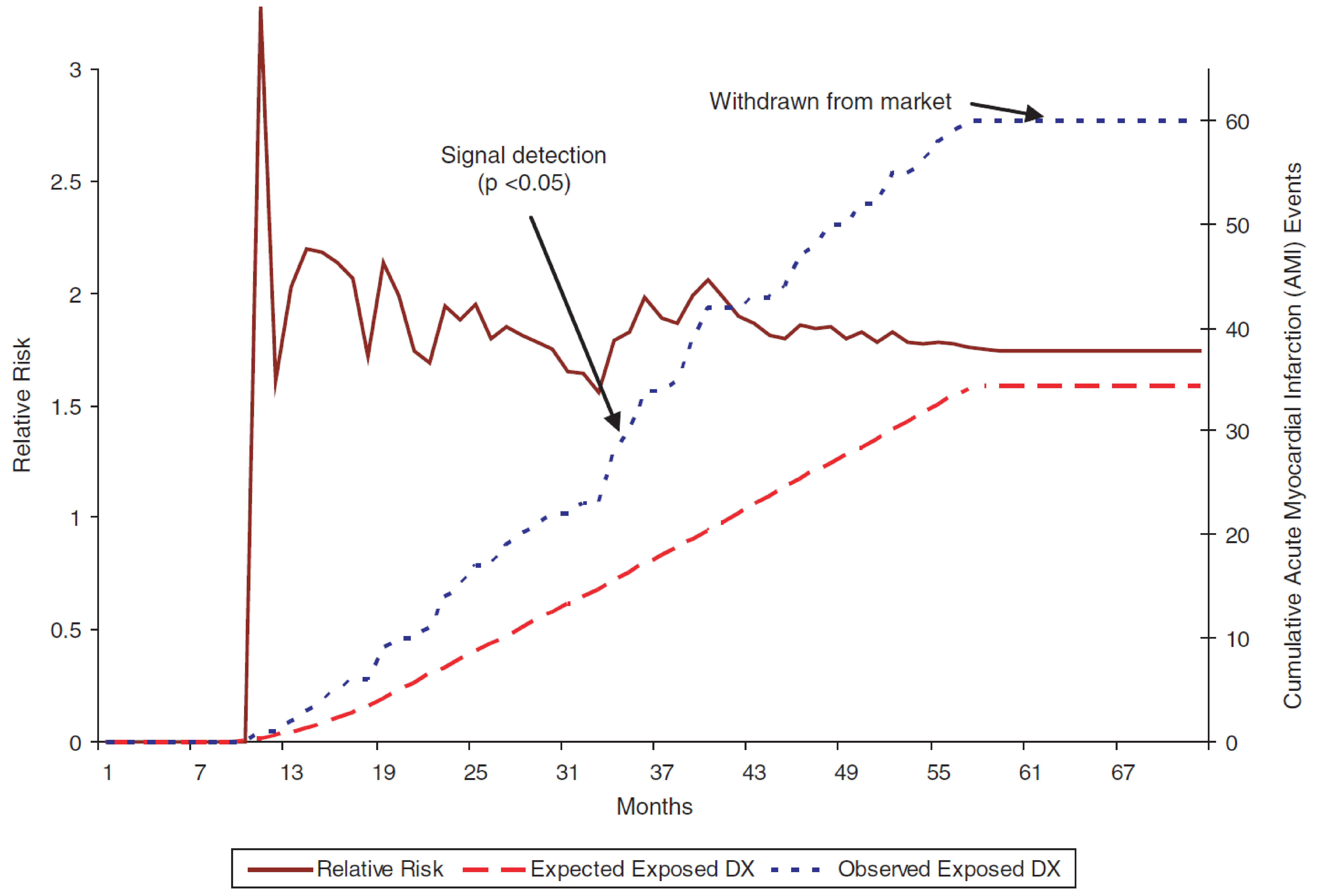
Amsterdam UMC – location AMC

The Netherlands

2000-2004

Vioxx

- Intended to treat arthritis & pain
- Increased risk of heart attack and stroke

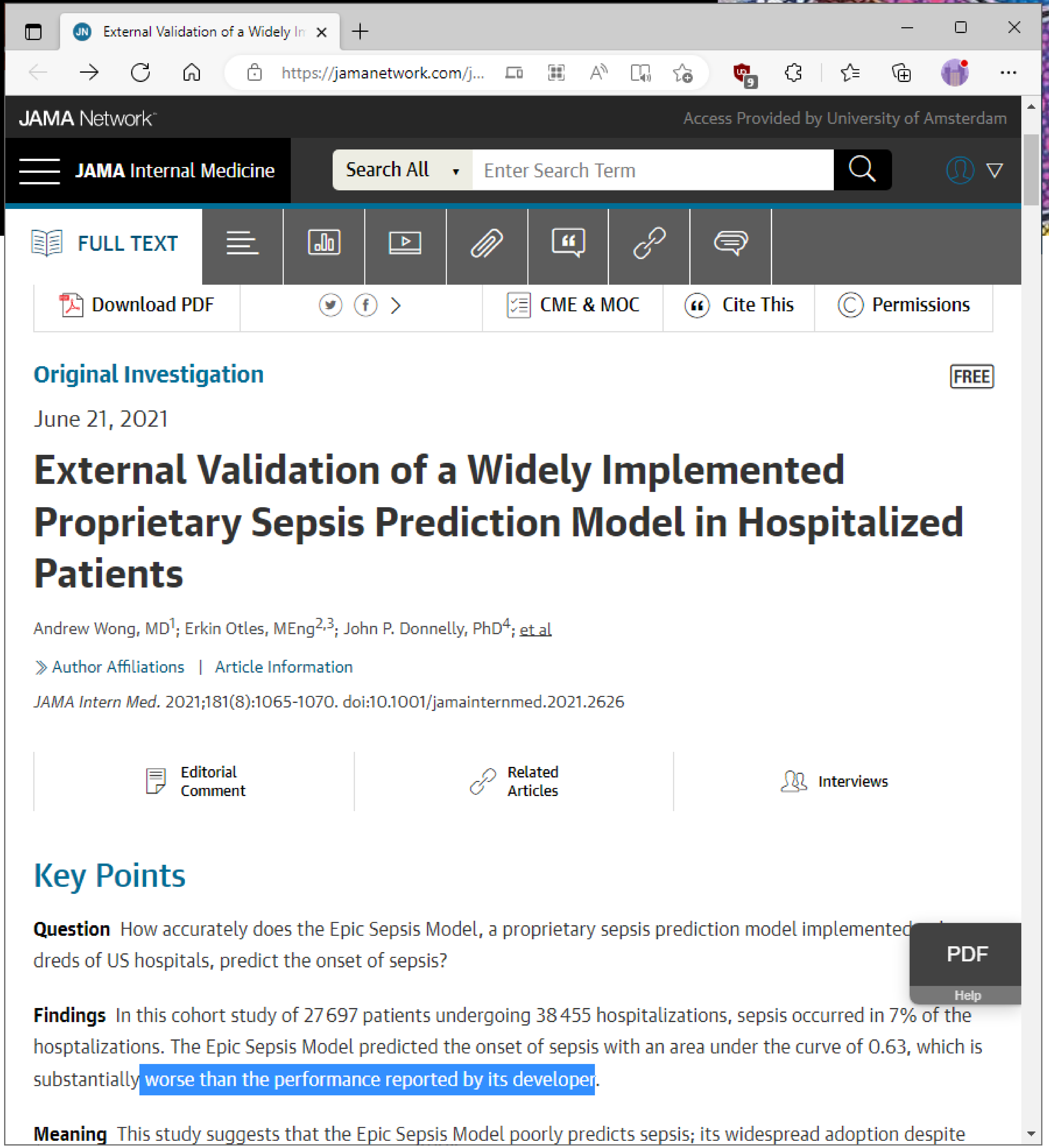


Source: Challenges for the FDA: The Future of Drug Safety, Workshop Summary, 2007

MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

June 21, 2021


JAMA Network | Access Provided by University of Amsterdam

JAMA Internal Medicine

FULL TEXT

Download PDF | CME & MOC | Cite This | Permissions

Original Investigation — FREE

June 21, 2021

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD[1]; Erkin Otles, MEng[2,3]; John P. Donnelly, PhD[4]; et al

Author Affiliations | Article Information

JAMA Intern Med. 2021;181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626

Editorial Comment | Related Articles | Interviews

Key Points

Question How accurately does the Epic Sepsis Model, a proprietary sepsis prediction model implemented ... dreds of US hospitals, predict the onset of sepsis?

Findings In this cohort study of 27 697 patients undergoing 38 455 hospitalizations, sepsis occurred in 7% of the hosptalizations. The Epic Sepsis Model predicted the onset of sepsis with an area under the curve of 0.63, which is substantially worse than the performance reported by its developer.

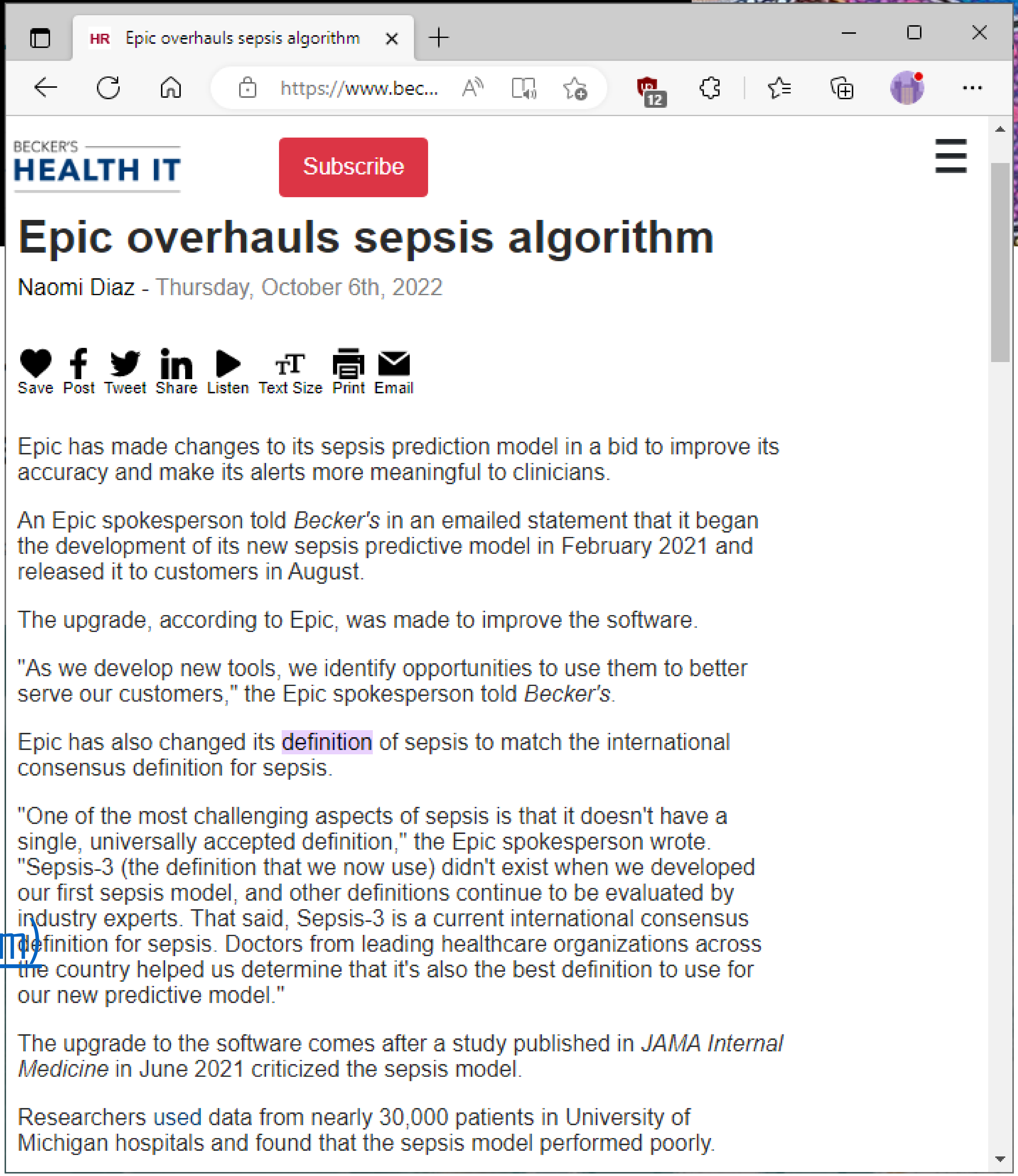
Meaning This study suggests that the Epic Sepsis Model poorly predicts sepsis; its widespread adoption despite

[External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients | Critical Care Medicine | JAMA Internal Medicine | JAMA Network]

@TheInstituteDH #MEDINFO23

October 6, 2022

[Epic overhauls sepsis algorithm (beckershospitalreview.com)](https://www.beckershospitalreview.com)



Epic overhauls sepsis algorithm

https://www.bec...

BECKER'S HEALTH IT

Subscribe

Epic overhauls sepsis algorithm

Naomi Diaz - Thursday, October 6th, 2022

Save Post Tweet Share Listen Text Size Print Email

Epic has made changes to its sepsis prediction model in a bid to improve its accuracy and make its alerts more meaningful to clinicians.

An Epic spokesperson told *Becker's* in an emailed statement that it began the development of its new sepsis predictive model in February 2021 and released it to customers in August.

The upgrade, according to Epic, was made to improve the software.

"As we develop new tools, we identify opportunities to use them to better serve our customers," the Epic spokesperson told *Becker's*.

Epic has also changed its definition of sepsis to match the international consensus definition for sepsis.

"One of the most challenging aspects of sepsis is that it doesn't have a single, universally accepted definition," the Epic spokesperson wrote. "Sepsis-3 (the definition that we now use) didn't exist when we developed our first sepsis model, and other definitions continue to be evaluated by industry experts. That said, Sepsis-3 is a current international consensus definition for sepsis. Doctors from leading healthcare organizations across the country helped us determine that it's also the best definition to use for our new predictive model."

The upgrade to the software comes after a study published in *JAMA Internal Medicine* in June 2021 criticized the sepsis model.

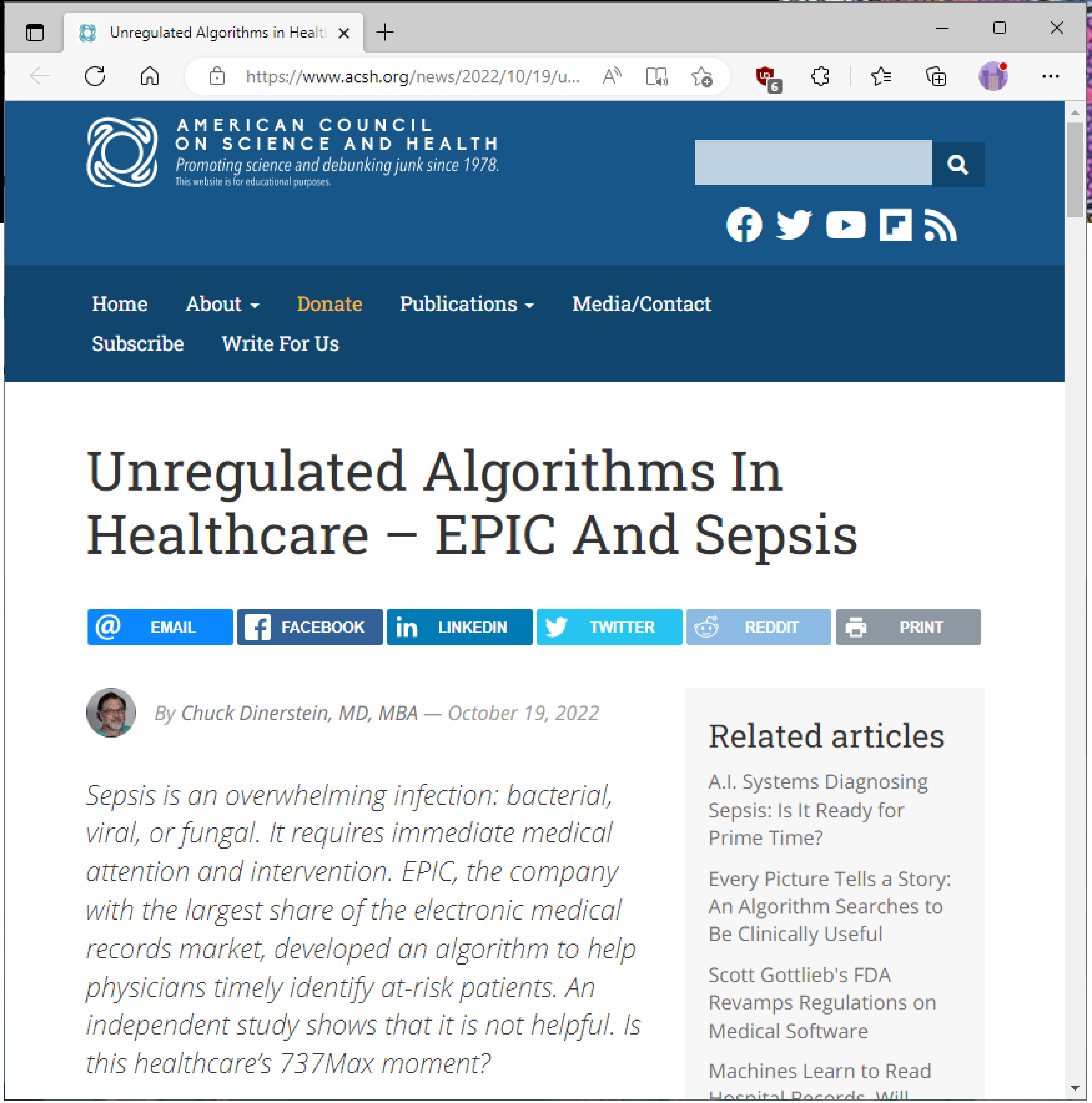
Researchers used data from nearly 30,000 patients in University of Michigan hospitals and found that the sepsis model performed poorly.

MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

October 19, 2022

[Unregulated Algorithms in Healthcare – EPIC and Sepsis | American Council on Science and Health (acsh.org)]


Unregulated Algorithms in Healt

https://www.acsh.org/news/2022/10/19/u...

AMERICAN COUNCIL ON SCIENCE AND HEALTH

Promoting science and debunking junk since 1978.

This website is for educational purposes.

Home About Donate Publications Media/Contact Subscribe Write For Us

Unregulated Algorithms In Healthcare – EPIC And Sepsis

EMAIL FACEBOOK LINKEDIN TWITTER REDDIT PRINT

By Chuck Dinerstein, MD, MBA — October 19, 2022

Sepsis is an overwhelming infection: bacterial, viral, or fungal. It requires immediate medical attention and intervention. EPIC, the company with the largest share of the electronic medical records market, developed an algorithm to help physicians timely identify at-risk patients. An independent study shows that it is not helpful. Is this healthcare's 737Max moment?

Related articles

A.I. Systems Diagnosing Sepsis: Is It Ready for Prime Time?

Every Picture Tells a Story: An Algorithm Searches to Be Clinically Useful

Scott Gottlieb's FDA Revamps Regulations on Medical Software

Machines Learn to Read

April 19, 2023

[Microsoft and Epic partner on OpenAI tools (digitalhealth.net)](https://www.digitalhealth.net/2023/04/)

May 31, 2023

[2305.17493v2.pdf (arxiv.org)](https://arxiv.org/pdf/2305.17493v2.pdf)

The curse of recursion: Training on generated data makes models forget

Example of text outputs of an OPT-125m model affected by *Model Collapse*– models degrade over generations, where each new generation is trained on data produced by the previous generation.

Input: some started before 1360 — was typically accomplished by a master mason and a small team of itinerant masons, supplemented by local parish labourers, according to Poyntz Wright. But other authors reject this model, suggesting instead that leading architects designed the parish church towers based on early examples of Perpendicular

Outputs:

Gen 0: Revival architecture such as St. John's Cathedral in London. The earliest surviving example of Perpendicular Revival architecture is found in the 18th @-@ century Church of Our Lady of Guernsey, which dates from the late 19th century. There are two types of perpendicular churches : those

Gen 1: architecture such as St. Peter's Basilica in Rome or St. Peter's Basilica in Buenos Aires. There is no evidence that any of these buildings were built during the reign of Pope Innocent III, but it is possible that they may have been built during the reign of his successor, Pope Innocent

Gen 7: architecture in England. In an interview with The New York Times, Wright said : " I don 't think there is anything wrong with me being able to do what I want to do. It just doesn 't work for me. " He added : " I don 't know if you can call it funny,

Gen 9: architecture. In addition to being home to some of the world's largest populations of black @-@ tailed jackrabbits, white @-@ tailed jackrabbits, blue @-@ tailed jackrabbits, red @-@ tailed jackrabbits, yellow @-

June 22, 2023

MIT Technology Review

Featured Topics Newsletters Events Podcasts SIGN IN SUBSCRIBE

ARTIFICIAL INTELLIGENCE

The people paid to train AI are outsourcing their work... to AI

It's a practice that could introduce further errors into already error-prone models.

By Rhiannon Williams June 22, 2023

[The people paid to train AI are outsourcing their work… to AI | MIT Technology Review]

June 26, 2023

"he was working on chatbots and was making about $3 an hour"

TECHNOLOGY

Behind the secretive work of the many, many humans helping to train AI

June 26, 2023 · 4:33 PM ET

Heard on All Things Considered

By Jonaki Mehta, Patrick Jarenwattananon, Ari Shapiro

4-Minute Listen + PLAYLIST

NPR's Ari Shapiro talks with The Verge's investigative editor Josh Dzieza about his recent report revealing the massive number of humans powering and training artificial intelligence.

[Behind the secretive work of the many, many humans helping to train AI : NPR]

Data – knowledge – implementation

- Medical knowledge is estimated to double every 73 days, i.e., multiplies by 1000 in 2 years

Medical knowledge has been expanding exponentially. Whereas the doubling time was an estimated 50 years back in 1950, it accelerated to 7 years in 1980, 3.5 years in 2010, and a projected 73 days by 2020, according to a *2011 study in Transactions of the Amercan Clinical and Climatological Association* ↗ .

Medical knowledge doubles every few months; how can clinicians keep up? (elsevier.com)

Data – knowledge – implementation

- Medical knowledge is estimated to double every 73 days, i.e., multiplies by 1000 in 2 years
- The knowledge-implementation gap is 17 years

[The answer is 17 years, what is the question: understanding time lags in translational research - Zoë Slote Morris, Steven Wooding, Jonathan Grant, 2011 (sagepub.com)]

Data – knowledge – implementation

- Medical knowledge is estimated to double every 73 days, i.e., multiplies by 1000 in 2 years
- The knowledge-implementation gap is 17 years

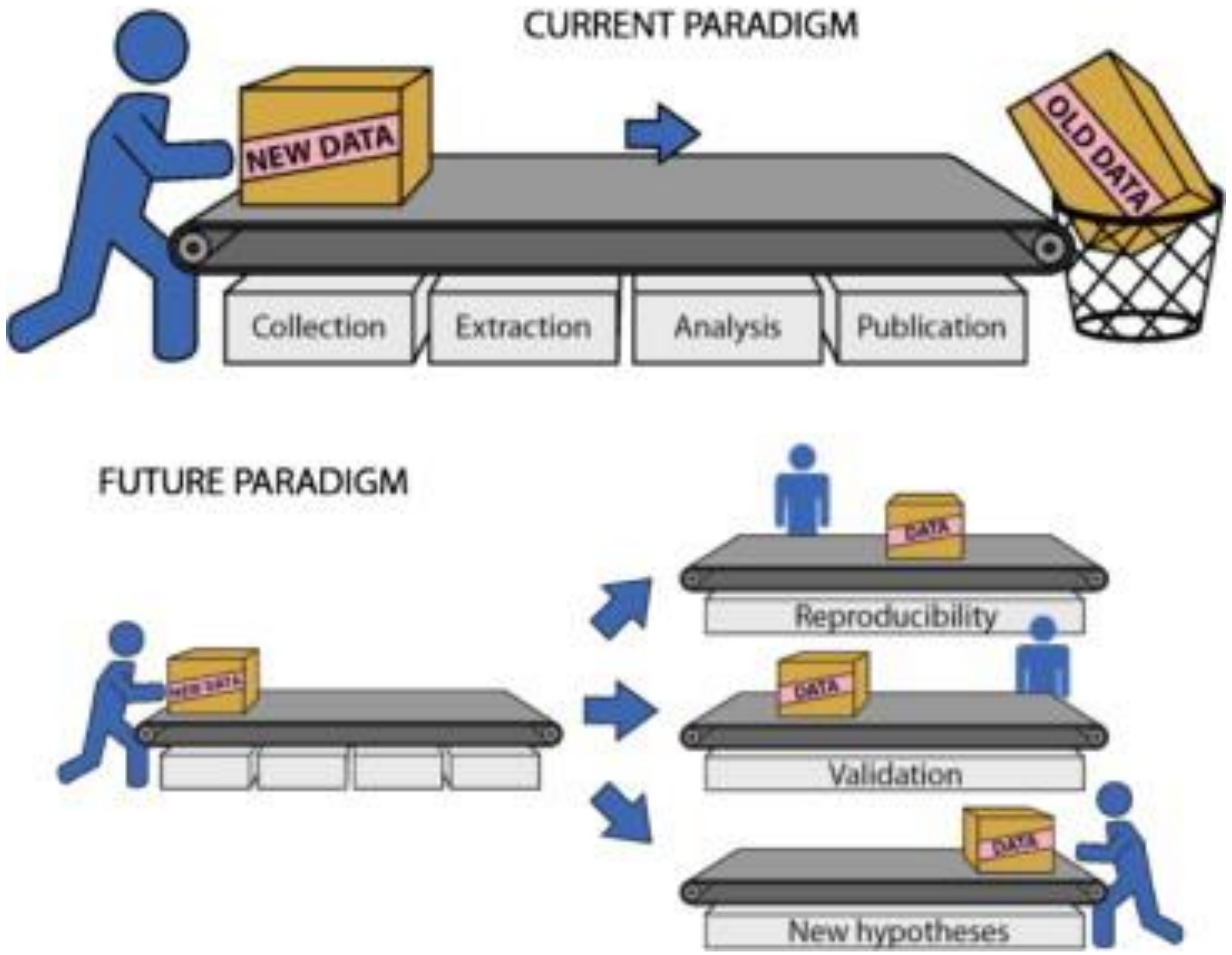
➔ $1000^{8.5} = 32 * 10^{24}$

Closing the loop – 3 needs

- Increase and accelerate **data availability**
  - Data visiting instead of data sharing
- Increase **insight in and oversight of models**
  - "repository of algorithms", including scope, use, performance
- Continuous monitoring of "AI interventions": stop / scale-up

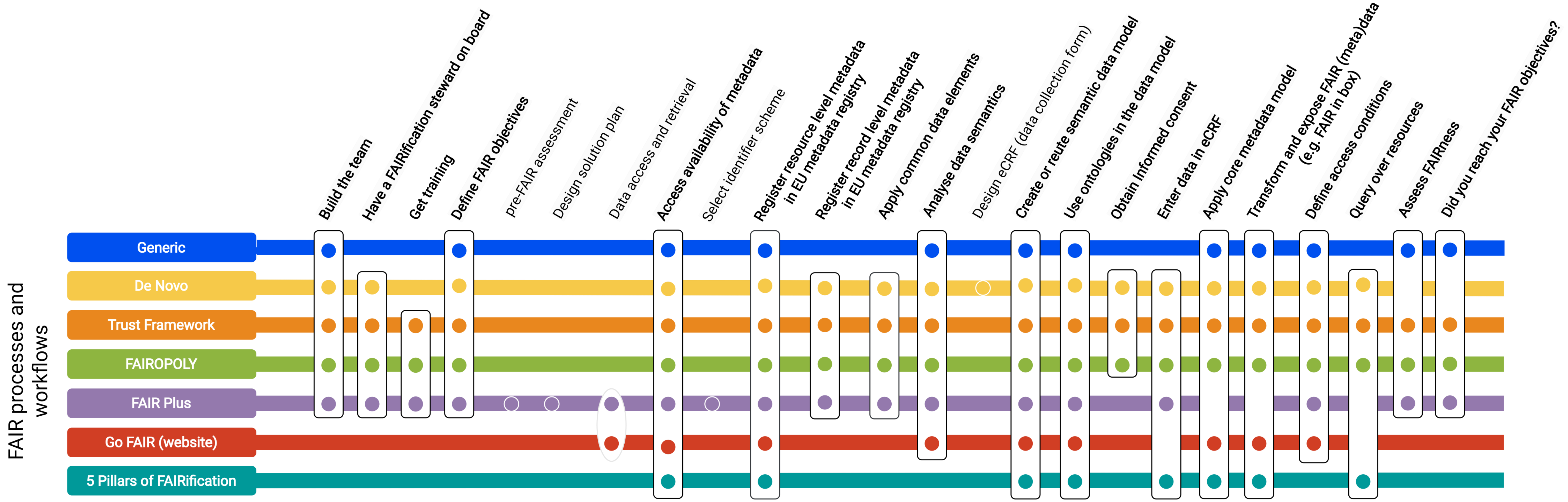
Responsible stewardship & use

- Data & Models
  - High quality
  - As open as possible
  - **FAIR:** Findable, Accessible, Interoperable, Reusable
  - Federated
- Questions:
  - who bears the burden of making FAIR and training models



https://doi.org/10.1016/j.radonc.2013.07.007

Making data FAIR in practice – the metroline



Mapped steps between FAIR processes and workflows

AMIA Policy Committee Work Product

- Based on work by the American Medical Informatics Association's Policy Committee 2020- 2021 and approved by the Board of Directors
- Solomonides AE, Koski E, Atabaki SM, Weinberg S, McGreevey JD, Kannry JL, Petersen C, Lehmann CU. Defining AMIA's artificial intelligence principles. J Am Med Inform Assoc. 2022 Mar 15;29(4):585-591. doi: 10.1093/jamia/ocac006. PMID: 35190824; PMCID: PMC8922174.
- https://academic.oup.com/jamia/article/29/4/585/6534106

Volume 29, Issue 4
April 2022

AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

Examples of Bias in ML

- Algorithm to predict complex health needs of patients to allocate resources
- Used health expenditure as a proxy for health status ("the more spent on healthcare, the worse a person's health must be")
- What do you think happened?

RESEARCH ARTICLE

ECONOMICS

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

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

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.

- Black patients with the same level of illness were less likely to be able to afford and access needed services
- The algorithm predicted lower future costs, incorrectly assessing better health and fewer needed services for this population

Historical Bias

- Use of historical data that may no longer reflect reality

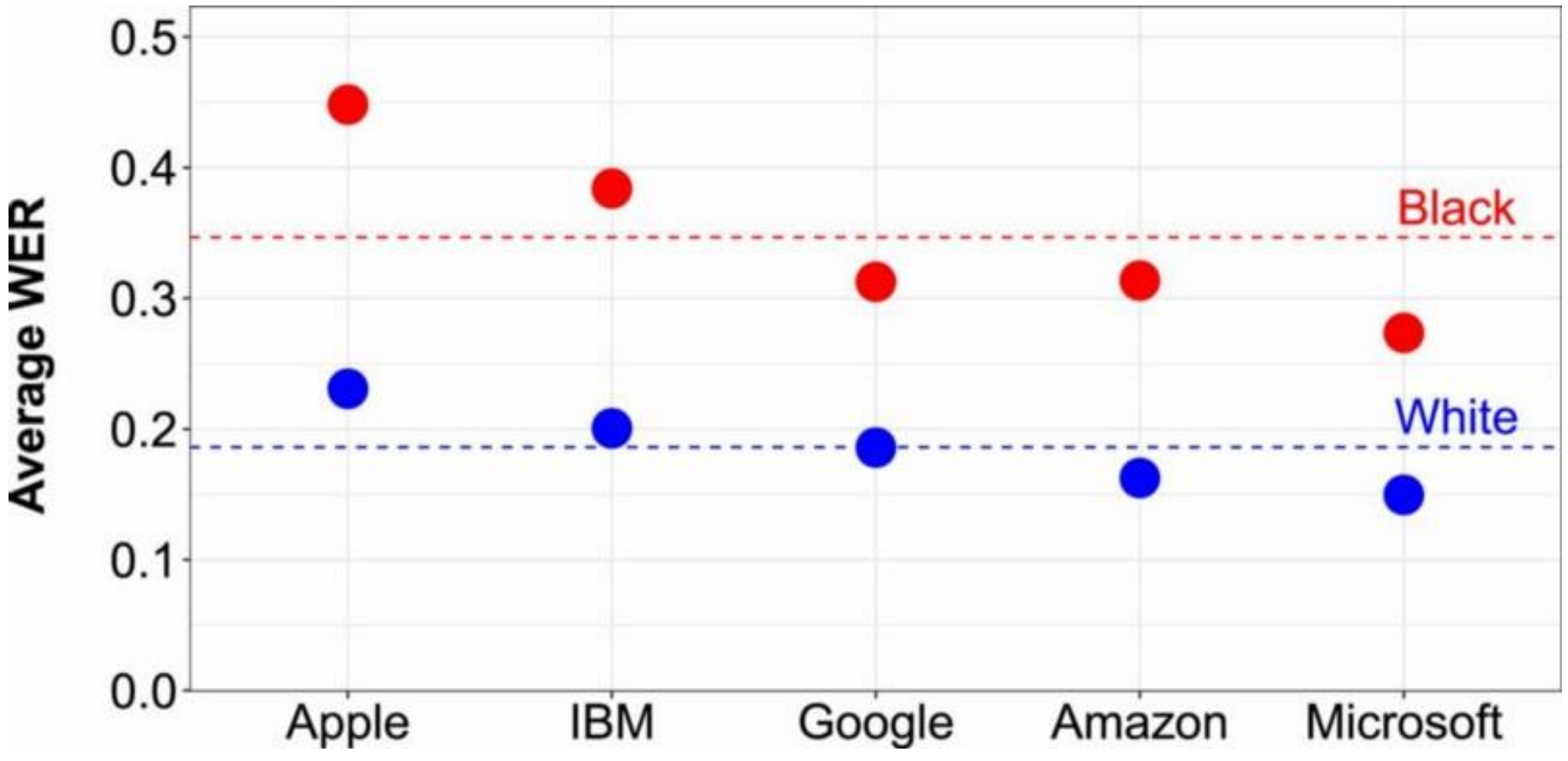
- 2014, Amazon built a system to screen job applicants from CVs
- Data from 2004 – 2014 where most employees were male

- Result: The system identified males as more suitable candidates
- Project was scrapped

https://www.seldon.io/6-types-of-ai-bias

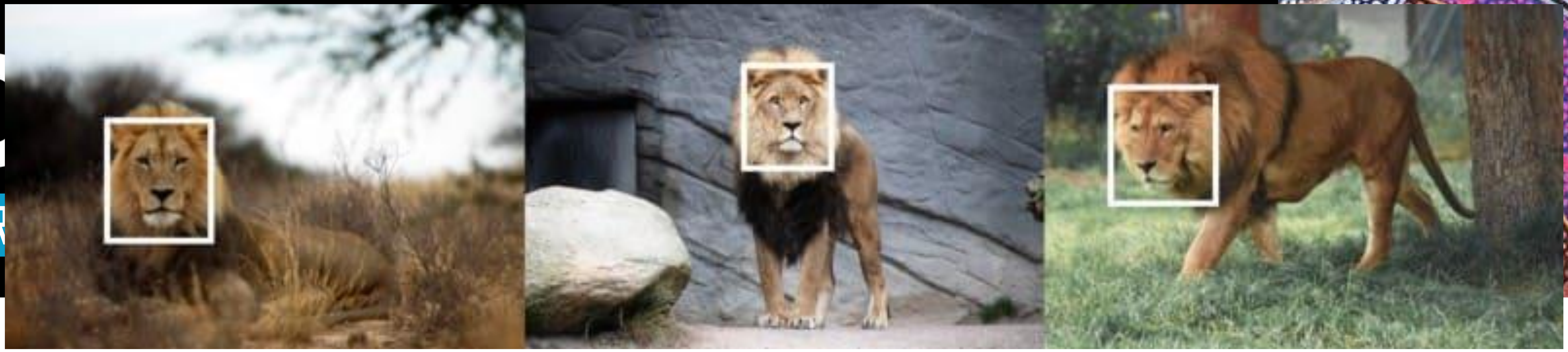
Sample Bias

- Training data do not accurately reflect the makeup of the real world
- Speech-to-Text System
- Trained on Audiobooks - narrated by well educated, middle aged, white men
- Underperforms with speakers from different socio-economic or ethnic backgrounds



Word Error Rate = WER

https://www.seldon.io/6-types-of-ai-bias



Label Bias

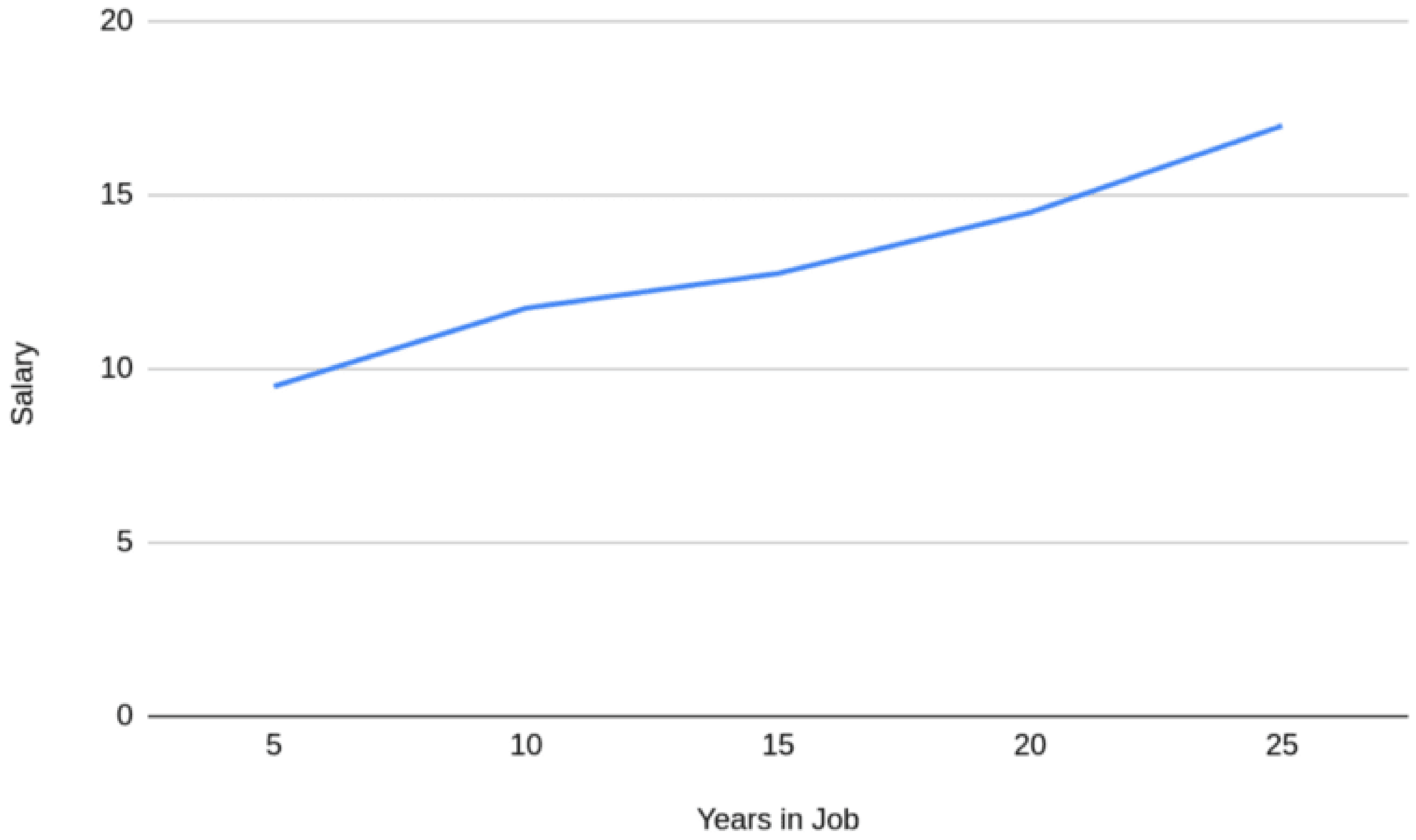
- ML models need labeled data - Labeling may vary
- Above only front facing lions are labeled;
- The system is unable to identify a lion from its side

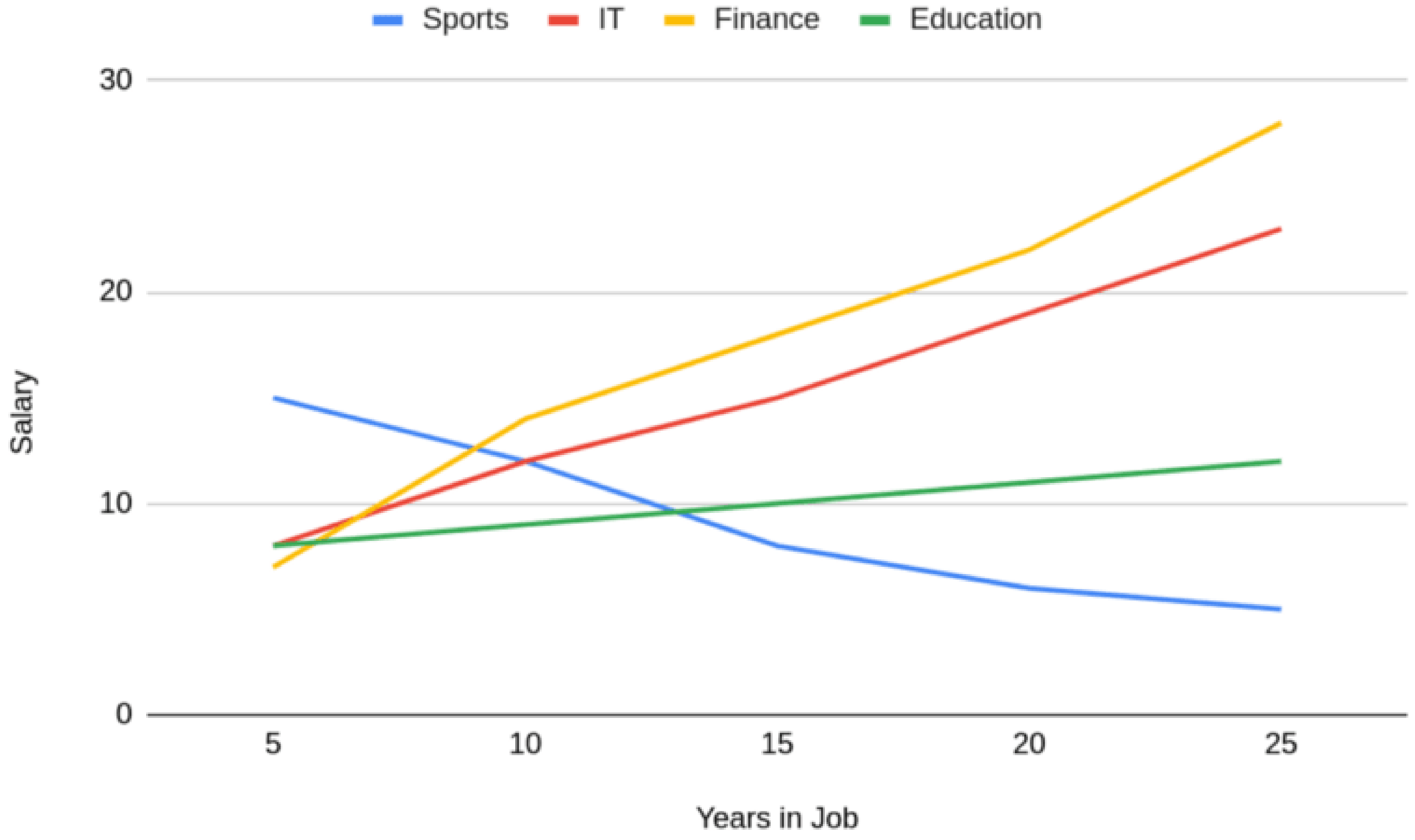
https://www.seldon.io/6-types-of-ai-bias

Aggregation Bias

- Aggregating data may introduce bias
- Graph shows salary and years on the job – linear correlations.

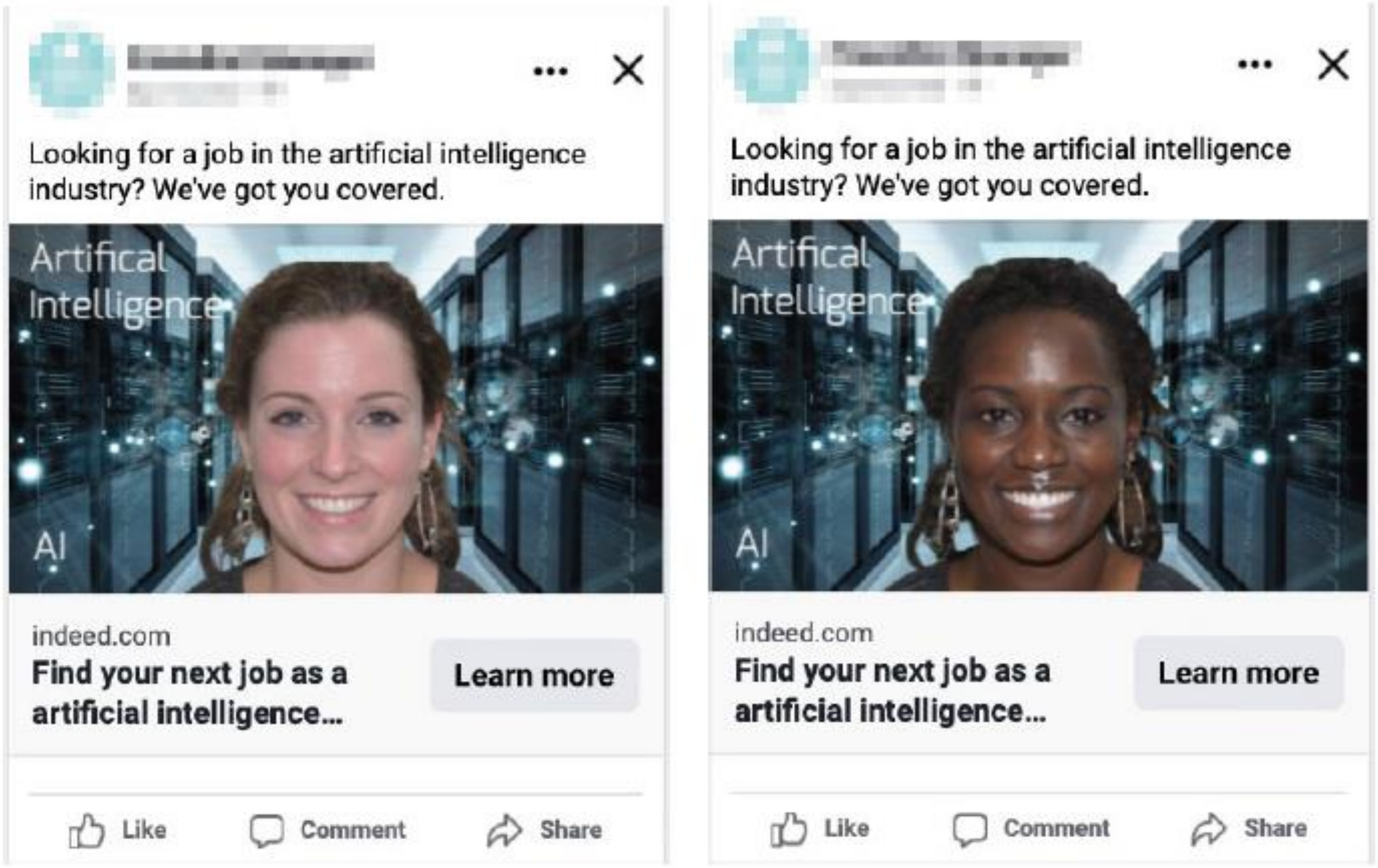
- Now look at the data used to create this graph
- For athletes the opposite is true.





https://www.seldon.io/6-types-of-ai-bias

Bias Example in Social Media



- Researchers ran the same ad but alternated pictures
- Images of women were delivered to an actual audience of 50% women
- Pictures of older women and female children are delivered primarily to women (58% and 55% women, respectively)
- Pictures of teenage women are delivered primarily to men (43% women)

- Synthetic images of adult Black people were delivered to 81% Black users
- Synthetic images of adult white people were delivered to only 50% Black users on average.

https://facebook-targeting.ccs.neu.edu/measurement/papers/kaplan2022measurement.pdf

Examples of AI Bias

- Apple's credit algorithm extended lower credit to wives than their husbands
- Hispanics are more likely to have their prepaid, legal transactions reported to the Financial Crimes Enforcement Network (less likely to have a bank account)
- Facebook's AI application discriminated by race and gender in housing advertisements
- AI to predict patients ready for hospital discharge demonstrated a bias against people from poorer neighborhoods with more African-Americans

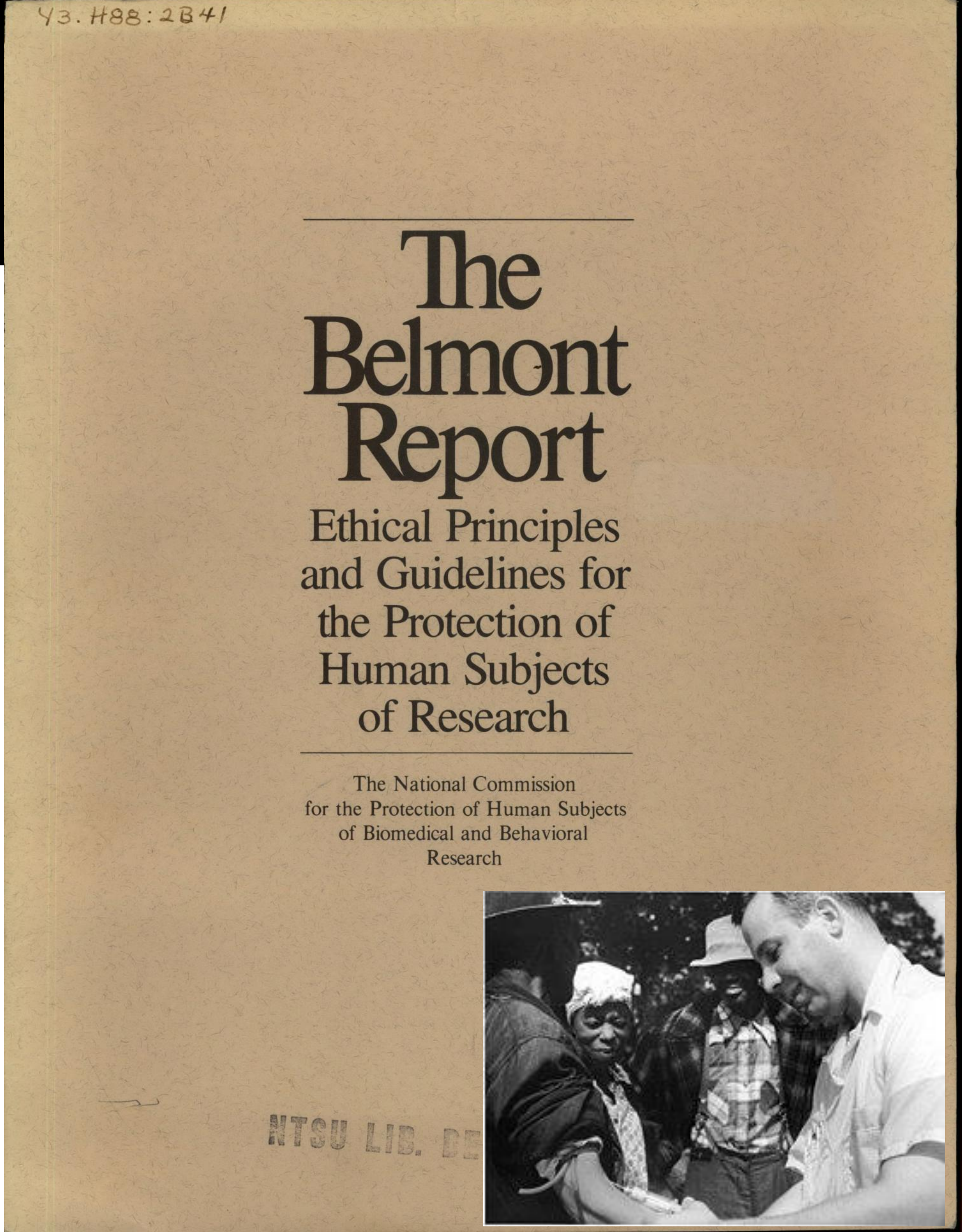
MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Belmont Principles - 1974

- National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research
  - Autonomy
  - Beneficence
  - Nonmaleficence
  - Justice

  - Re-interpreted for AI
  - +11 additional principles



AI System

Responsible Principles for AI

- Beneficence
  - AI is designed explicitly to be helpful to people, who use it or on whom it is used, and to reflect the ideals of compassionate, kind, and considerate human behavior

- Autonomy
  - Context AI: operates without human oversight
  - Context Ethics: "**protecting the autonomy of all people** and treating them with courtesy and respect and facilitating informed consent"

Primum Non Nocere

Responsible Principles for AI

- Nonmaleficence
  - "Do No Harm"
  - Every reasonable effort shall be made to avoid, prevent, and minimize harm or damage to any stakeholder

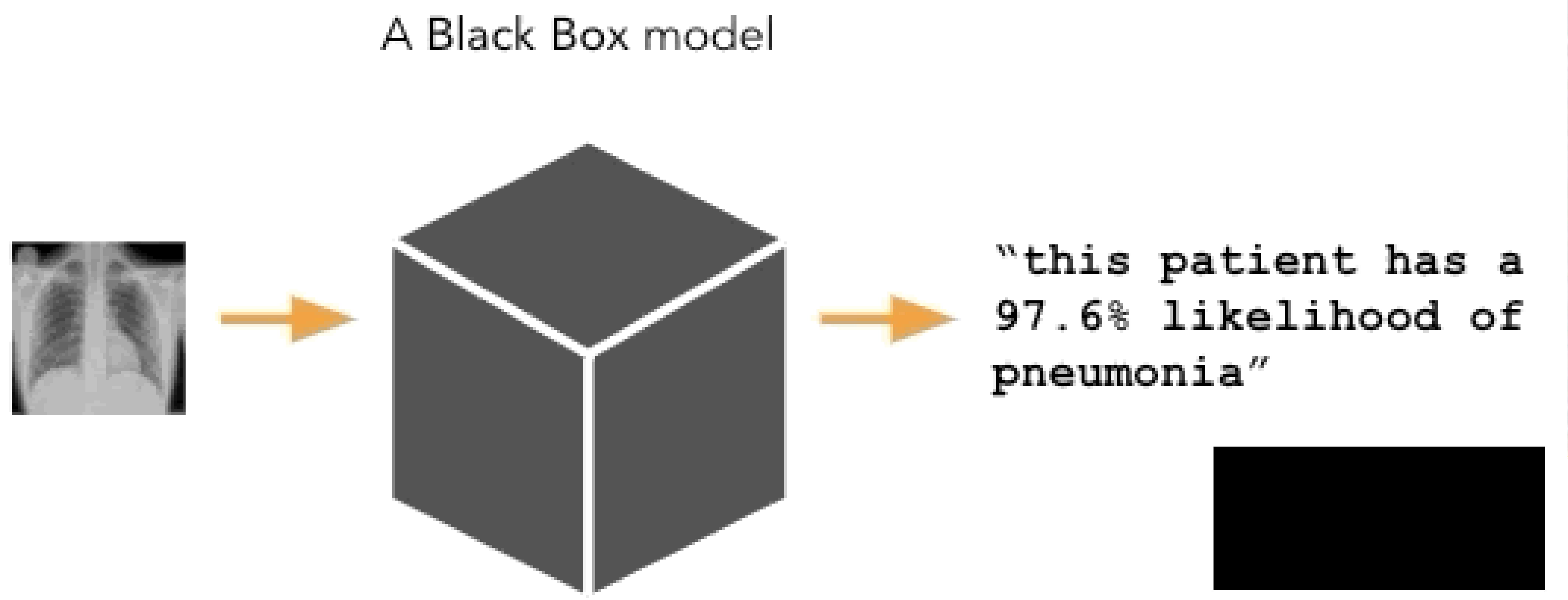
- Justice
  - Equity in representation in and access to AI, data, and the benefits of AI
  - Fair access to redress and remedy be available in the event of harm resulting from the use of AI
  - Affirmative use of AI to support social justice

Principles for the Organization

- Benevolence
  - Organizations developing AI systems must intend positive purposes (e.g., improved health outcomes) rather than negative purposes (e.g., to further bias, exploit individuals, advance financial interests)
- Transparency
  - AI systems do not incorporate or conceal any special interests
  - AI systems deal evenhandedly and fairly with all good faith actors
  - Stakeholders understand that they are dealing with AI in the first place

Principles for the Organization

- Accountability
  - AI requires active oversight and a clear “reporting line
  - Any risk deemed attributable to AI must be reported, assessed, monitored, measured, and mitigated
  - Required ongoing oversight of AI systems
  - Lodging a complaint and receiving proper redress, and escalation of a complaint should be possible



AI Technical Principles

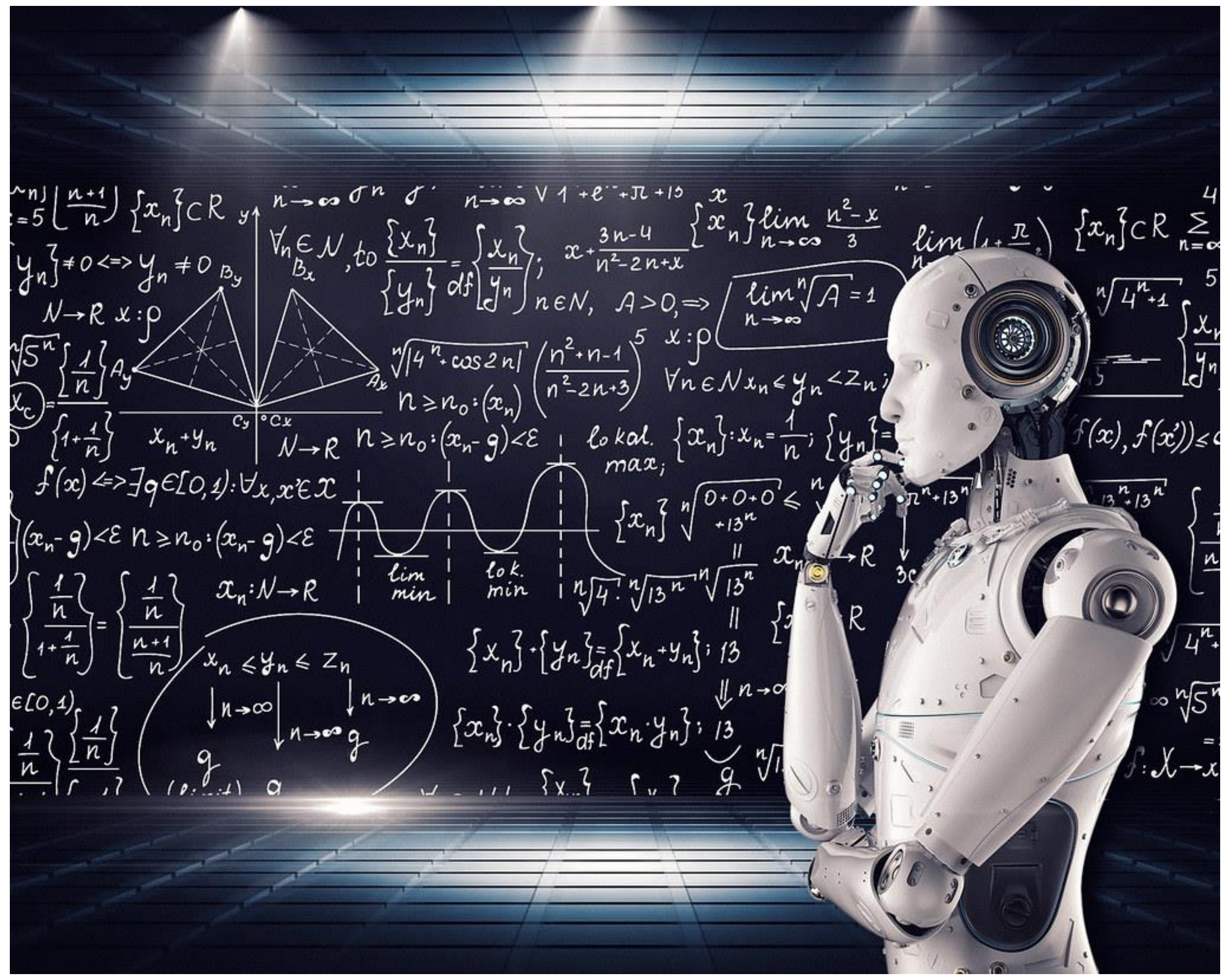
- Explainability
  - AI may not function as a "black box" to users or patients
  - Developers must
    - declare the scope, proper application, and limitations of their work
    - provide sufficient information about the general derivation of their output
    - Upon request provide a role-appropriate (e.g., lay language for patients) explanation
- Interpretability
  - AI must present plausible reasoning for decisions or advice, which must be presented in appropriately accessible language based on the stakeholder

AI Technical Principles

- Fairness
  - AI must be free of bias and must be nondiscriminatory
- Dependability
  - AI must be robust, safe, secure, and resilient
  - At worst it "fails gracefully" (leaves system in a safe or secure state)
- Auditability
  - AI must provide an "audit trail" of its performance including internal changes
  - Audit log contains model state, the input variables, and the resulting output for any system decision or recommendation

AI Technical Principles

- Knowledge Management
  - Developers must maintain AI systems including retraining of algorithms on new data or new populations
  - The models powering AI need to have clearly listed creation, revalidation, and expiration dates (transparent to users)
  - Algorithmovigilance



AI Research

- Needed to
  - Understand the technology better as it evolves
  - Ensure its humane and ethical application in society and the economy

Conclusion

- AI will play an important role in the gains in medical knowledge, diagnosis, and treatment in the 21st century
- AI has the potential to make healthcare healthcare safer, more effective, less costly, and even more equitable
- AI must be introduced judiciously, in the appropriate environments, and in accordance with the ethical principles outlined
- *Algorithmovigilance* is paramount