Artificial Intelligence in Medicine: Context and Controversies

BE Chapman, PhD

2024-05-16

Outline

- *ethos, pathos, logos*

Outline

- *ethos*, *pathos*, *logos*
- a little about me, some context about healthcare, trends and controversies regarding AI in healthcare

Outline

- *ethos, pathos, logos*
- a little about me, some context about healthcare, trends and controversies regarding AI in healthcare
- But first, let's do some polls

Section 1

"*Ethos*: A little bit about me"

I'm honored to be invited to talk with you

Dennis Parker, Ph.D.: "Brian, your mistake is that you want to understand what it all means."

What are my credentials?

- PhD in medical informatics
- 30 years of professional experience of working on, collaborating with, and teaching about AI in healthcare
  - Primarily in medical imaging and natural language processing

What are my credentials?

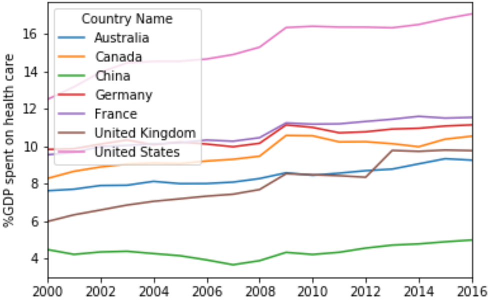
- Long-term patient
  - 4-time cancer survivor
  - Two childhood cancers (1976, 1983)
  - Two adult cancers
  - 12+ surgeries
  - 4 emergencies
  - Advanced peritonitis
  - Gangrenous strangulated bowel
  - Many resulting chronic issues!

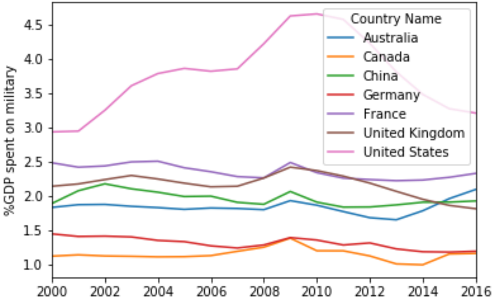
My “mobile medical record”

Section 2

Pathos: The Context of Healthcare

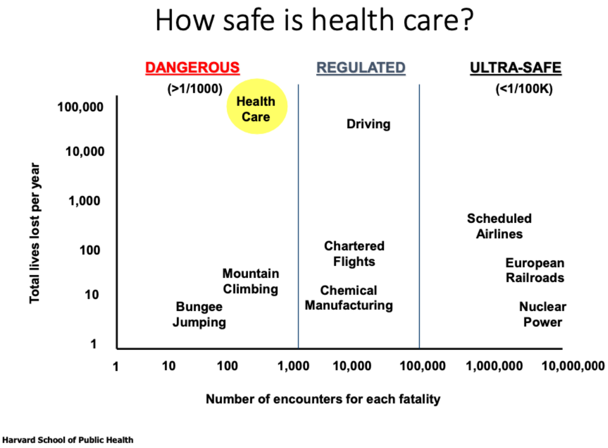
Healthcare is expensive [1]





[1] Data are taken from the World Bank

Healthcare is dangerous



Data collection can come at significant cost, discomfort, and risk [2]

These values are disputed (Meer et al. 2012) (see statistical fog below)

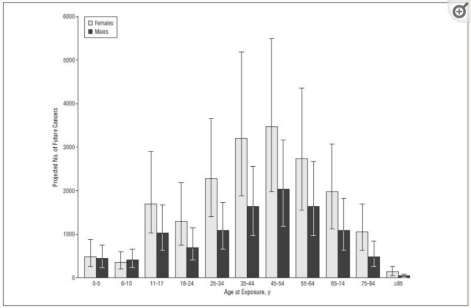


Figure 2

Projected number of future cancers (mean and 95% uncertainty limits) that could be related to computed tomographic scan use in the United States in 2007, according to age at exposure.

[2] lez et al. (2009)

Healthcare has delivery challenges (Braithwaite, Glasziou, and Westbrook 2020)

Braithwaite *et al. BMC Medicine* (2020) 18:102
https://doi.org/10.1186/s12916-020-01563-4

Received: 30 July 2019 Revised: 11 March 2020
Accepted: 17 March 2020 Published online: 04 May 2020

The three numbers you need to know about healthcare: the 60-30-10 Challenge

Jeffrey Braithwaite[1*], Paul Glasziou[2] and Johanna Westbrook[3]

Abstract

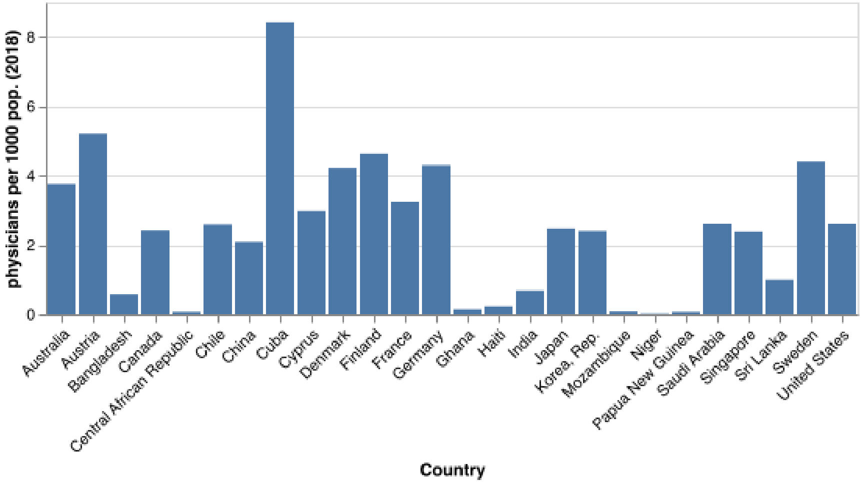
Background: Healthcare represents a paradox. While change is everywhere, performance has flatlined: 60% of care on average is in line with evidence- or consensus-based guidelines, 30% is some form of waste or of low value, and 10% is harm. The 60-30-10 Challenge has persisted for three decades.

Main body: Current top-down or chain-logic strategies to address this problem, based essentially on linear models of change and relying on policies, hierarchies, and standardisation, have proven insufficient. Instead, we need to marry ideas drawn from complexity science and continuous improvement with proposals for creating a deep learning health system. This dynamic learning model has the potential to assemble relevant information including patients' histories, and clinical, patient, laboratory, and cost data for improved decision-making in real time, or close to real time. If we get it right, the learning health system will contribute to care being more evidence-based and less wasteful and harmful. It will need a purpose-designed digital backbone and infrastructure, apply artificial intelligence to support diagnosis and treatment options, harness genomic and other new data types, and create informed discussions of options between patients, families, and clinicians. While there will be many variants of the model, learning health systems will need to spread, and be encouraged to do so, principally through diffusion of innovation models and local adaptations.

Conclusion: Deep learning systems can enable us to better exploit expanding health datasets including traditional and newer forms of big and smaller-scale data, e.g. genomics and cost information, and incorporate patient preferences into decision-making. As we envisage it, a deep learning system will support healthcare's desire to continually improve, and make gains on the 60-30-10 dimensions. All modern health systems are awash with data, but it is only recently that we have been able to bring this together, operationalised, and turned into useful information by which to make more intelligent, timely decisions than in the past.

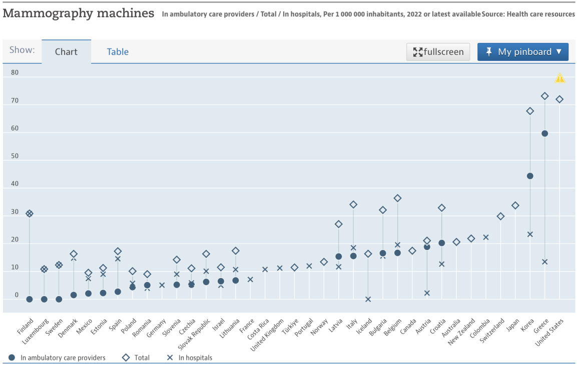
Keywords: Learning health system, Complexity, Complexity science, Change, Evidence-based care, Clinical networks, Quality of care, Patient safety, Policy, Healthcare systems

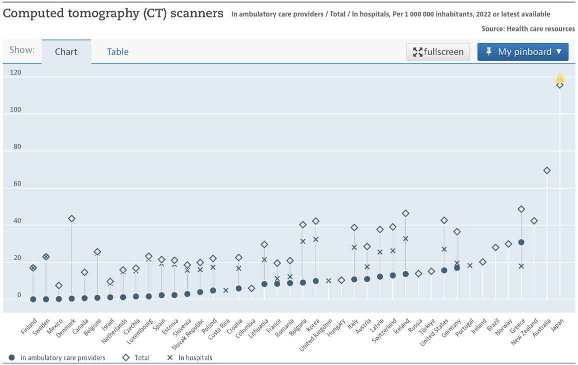
Healthcare has delivery challenges [3]



[3] Data are taken from the World Bank

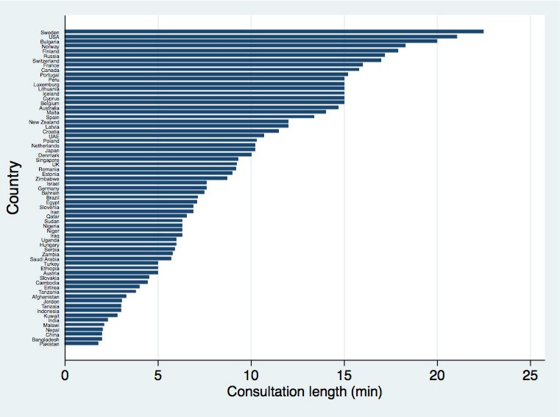
Society organizes healthcare resources differently [4]





[4] Data are taken from the OECD

Healthcare occurs in time-constrained settings (2017) (Irving et al. 2017)



Healthcare occurs within complex organizations

"time and motion studies have identified as many as 50 steps between the moment the doctor wrote the order and the moment the nurse finally administered the medication." (Wachter 2015, 129)

Medical knowledge is "impoverished"

Meet Ted Shortliffe and Bruce Buchanan

Medical knowledge is "impoverished"

- Ted Shortliffe: In medicine "you never tend to know absolutely anything"

Medical knowledge is “impoverished”

“Randomized controlled trials and other population-based studies are designed to discover whether or not health care interventions are effective. They are not designed to discover how health care interventions work (when they do work), or to come up with new ideas about mechanisms, new theories about disease processes, or new technologies for medical interventions. Health care interventions are judged effective when there is a correlation between the intervention and positive outcomes. . . ”

Figure 1: Miriam Solomon, Wikipedia

Medical knowledge is "impoverished"

"Often it is not too much of a leap to infer that the intervention causes the positive outcome." **But the resulting knowledge is rather impoverished: it is knowledge of what works, without knowledge of how it works (or why it does not work), or how to make it work better. It is knowledge of effects without knowledge of underlying mechanisms."** (Solomon 2015, 117)

Figure 2: Miriam Solomon, Wikipedia

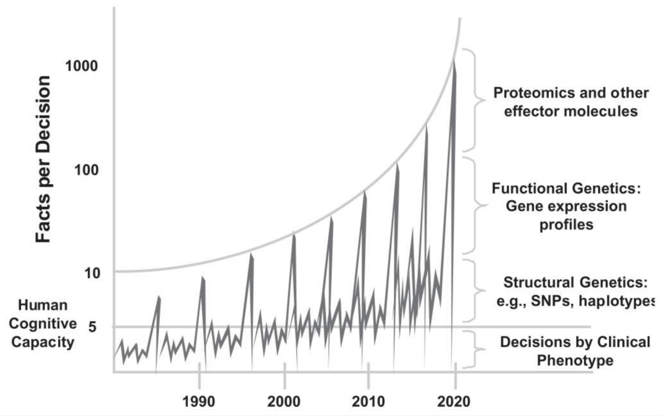
Medical Knowledge is "impoverished"

"Statistical fog" (Nicholas Rescher)

Figure 3: Wikipedia

Nonetheless (or because of this impoverished knowledge)...

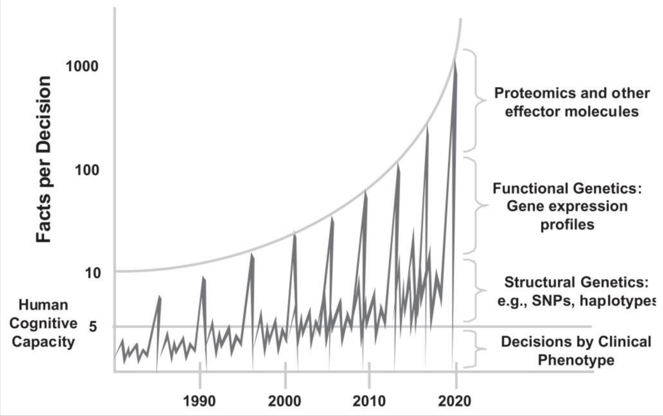
Overwhelming amount of information (Medicine et al. 2008)



Section 3

Context: Human Cognition

Limited working memory of human cognition



What are brains good for? Practicing medicine?

What, then, is the role of the biological brain....It is expert at recognizing patterns, at perception, and at controlling physical actions, but it is not so well designed...for complex planning and long, intricate, derivations of consequences. **It is, to put it bluntly, bad at logic and good at Frisbee.** (Clark 2003)

 [a]

[a] The New Yorker

Human brain exceptionalism

[W]hat is special about human brains, and what best explains the distinctive features of human intelligence, is precisely their ability to enter into deep and complex relationships with nonbiological constructs, props, and aids. This ability, however, does not depend on physical wire-and-implant mergers, so much as on our openness to information-processing mergers. (Clark 2003)

“Information-processing mergers” (Clark 2003)

- External objects that aid the mind’s reasoning
  - paper and pencil to do arithmetic
  - books for storing and recalling knowledge
  - Cell phones, computers
- Extended mind theory

"Man-Computer Symbiosis" (Licklider 1960)

4 IRE TRANSACTIONS ON HUMAN FACTORS IN ELECTRONICS March

Man-Computer Symbiosis*

J. C. R. LICKLIDER†

Summary—**Man-computer symbiosis is an expected development in cooperative interaction between men and electronic computers. It will involve very close coupling between the human and the electronic members of the partnership. The main aims are 1) to let computers facilitate formulative thinking as they now facilitate the solution of formulated problems, and 2) to enable men and computers to cooperate in making decisions and controlling complex situations without inflexible dependence on predetermined programs. In the anticipated symbiotic partnership, men will set the goals, formulate the hypotheses, determine the criteria, and perform the evaluations. Computing machines will do the routinizable work that must be done to prepare the way for insights and decisions in technical and scientific thinking. Preliminary analyses indicate that the symbiotic partnership will perform intellectual operations much**

will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.

B. Between "Mechanically Extended Man" and "Artificial Intelligence"

As a concept, man-computer symbiosis is different in an important way from what North[2] has called "mechanically extended man." In the man-machine systems of the past, the human operator supplied the initiative, the direction, the integration, and the criterion. The

"The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly and that the resulting partnership will think as no human brains as ever thought. . . ."

Section 4

Current AI Applications in Healthcare

Medicine makes extensive use of extended senses

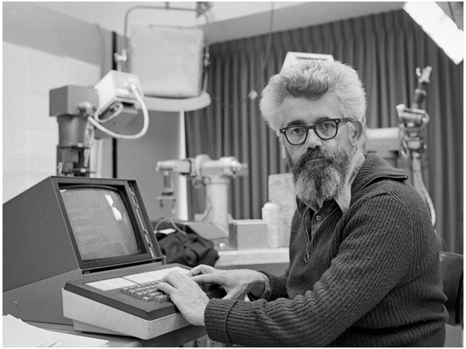
- Eye glasses and hearing aides
- Diagnostic imaging
- Stethescopes
- Microscopes
- Is the extended mind different?

Section 5

Logos: AI in healthcare

What is AI?

John McCarthy: "As soon as it works, no one calls it AI anymore." (Mitchell 2019)



[a]

[a] The Independent

What is AI? What I was taught! (by Al Pryor)

"Artificial intelligence is mostly artificial and not very intelligent."

What is AI?[5]

	Human-based	Ideal Rationality
Reasoning-Based	Systems that think like humans.	Systems that think rationally.
Behavior-Based	Systems that act like humans.	Systems that act rationally.

Which cell seems the most appropriate for healthcare?

[5] (Bringsjord and Govindarajulu 2022)

Three flavors of Medical AI (in order of historical significance)

- Probability (1960s, and 1990s)
  - Represent knowledge as as probabilities and use Bayes' theorem to make inferences
- Logic/Expert Systems (1970s-1980s)
  - Represent knowledge as computable rules and relationships
- Machine Learning (2000 present)
  - Learn implicit or explicit relationships between data (input) and classifications, predictions, etc. (output)

Early Probabilistic AI in Healthcare: 1961

"A Mathematical Approach to Medical Diagnosis: Application to Congenital Heart Disease" (WARNER et al. 1961)

"Old cardiologists just couldn't believe that a computer could do something better than a human." (Quoted in (McGrayne 2011))

Probabilistic AI

- Probabilities are great
    - Optimal way of reasoning under uncertainty
- Probabilities are challenging
    - You need **a lot** of them
    - How do you calculate the probabilities?

Early Expert System AI in Healthcare: MYCIN (1973)

"An artificial intelligence program to advise physicians regarding antimicrobial therapy" (Shortliffe et al. 1973)

- Exploration of heuristics
  - How to pragmatically represent uncertainty?
  - How do real humans, as opposed to "rational actors", make decisions?

Early Expert System AI in Healthcare: MYCIN

"Our goal is to build a program that can assist working scientists with reasoning problems. You wouldn't expect a tool to have all of the power of a working scientist, **but you would hope that you could design a smart system to provide some of the solutions to subproblems.**" (D. 1977)

Experts Argue Whether Computers Could Reason, and if They Should

By LEE DEMBART

Can machines think? Should they? The computer world is in the midst of a fundamental dispute over those questions because an eminent computer scientist wrote a book arguing that machines could never be made to reason like people and, what was more, should not be.

Twenty years ago, in the infancy of the computer revolution, before problems began cropping up, the public was told that computers would be smarter than brains. Computer chess champions and machine translation, for example, were just around the corner. So far, neither has been accomplished successfully, and neither is likely to be any time soon.

Nor have computers had much success in making decisions that require judgment. They can rattle off the Manhattan telephone directory unerringly time after time, which no human can do, but they cannot begin to distinguish one face from another, as babies can.

Computer scientists have always said, "Give us more time. The problem is more complex than we thought." Then one of them, Joseph Weizenbaum, a professor of computer science at the Massachusetts Institute of Technology, wrote a book saying that the project was fundamentally unsound and dangerous to pursue, partly, he said, because the computers' and humans' ways of thought would always be alien, and because knowledge might become limited to what a computer could undertand.

The elders of the artificial intelligence community reacted with outrage. Even those who agreed with his premises criticized the book as being too harsh in tone, too personal in its attacks. Computer journals have bristled over the last year with reviews, comments and replies provoked by Professor Weizenbaum's book, "Computer Power and Human Reason" (W. H. Freeman & Co., 1976). Now the controversy has spilled into the prestigious publication Science,

Continued on Page 34, Column 3

erals of his intention to back Mr. Cuomo all the way in an apparent effort to strengthen Mr. Cuomo's bid for the Liberal Party's nomination.

Titular Head Turns Elsewhere

If Mr. Cuomo wins the Liberal nomination but loses that of his own party, it would mean that the Governor would be supporting the Liberal candidate for Mayor rather than the Democratic one.

When Mr. Beame was asked about the Carey move, disclosed in The New York Times yesterday, he said it was difficult to believe "the titular head of the Democratic Party would support a candidate of another party rather than his own, especially after all his pretensions about the importance of an open primary."

"Apparently the Liberal Party is more important to him than the Democratic

Continued on Page 23, Column 1

Casinos Bring Atlantic City Woes

By MARTIN WALDRON
Special to The New York Times

ATLANTIC CITY, May 7—Fires, the condemnation of slum housing and the destruction of small hotels are driving out many elderly and poor as this seaside resort begins to experience sociological consequences in a prelude to the opening of gambling casinos.

Under a measure in the New Jersey Legislature, casinos could be opened only in large hotels. State and city officials hope to rebuild the deteriorating city, once the nation's premier convention center, by attracting $500 million worth of these new facilities.

Hotels and other businesses may be erected in an area now by slum housing, much of sought by speculators.

City officials deny tha policy to try to drive the poor from Atlantic City. A third of the city's 40,000 permanent residents are on welfare, with unemployment over 20 percent.

About 10,000 of these—mostly black and Hispanic residents—live in the area expected to be in demand soon for the construction of casinos, restaurants and parking facilities.

In all, some 700 to 1,000 people have

Continued on Page 36, Column 3

Figure 4: NYT 1977

Early Machine Learning AI in Healthcare (1985)

"A program for machine learning of counting criteria: empirical induction of logic-based classification rules" (Spackman 1985) (Kent Spackman)

How can AI be used in healthcare?

- Drug discovery
- Robotics
  - Surgery
  - Guided data acquisition (e.g. AI guided ultrasound acquisition)
  - Material delivery
- **decision support**

Drug discovery

- What are some of the ethical implications of AI being used for drug discovery?

Robotic/remote surgery

- What are some of the ethical implications of AI being used for robotic/remote surgery?
- What are some of the ethical implications of AI being used to guide data acquisition?

Robot Example: Pharmacy

"[A Swiss-made pharmacy robot] robot, installed in 2010 at a cost of $7 million, is programmed to pull medications off stocked shelves; to insert the pills into shrink-wrapped, bar-coded packages; to bind these packages together with little plastic rings; and then to send them. . . to locked cabinets on the patient floors. 'It gives us the first important step in eliminating the potential for human error,' said UCSF Medical Center CEO Mark Laret when the robot was launched." (Wachter 2015, 156)

- What kind of human errors do you think the CEO was concerned about?

Robot Example: Pharmacy

16-year-old Pablo Garcia had a rare genetic disease resulting in frequent infections and bowel problems. Each day Pablo took about 15 different medications, including an antibiotic to prevent recurrent skin infections. Pablo was checked into the hospital for a routine colonoscopy. As part of his admission, a doctor wrote prescriptions for each of the drugs Pablo would be taking during his hospitalization, including his daily dose of antibiotic. However, a medical error resulted in the doctor ordering a dose that was 39 times Pablo's usual dose.

Robot Example: Pharmacy

"The robot dutifully collected 38½ Septra tablets—with perfect accuracy—placed them on a half-dozen rings, and sent them to Pablo's floor, where they came to rest in a small bin waiting for the nurse to administer them at the appointed time. (Wachter 2015, 158)

- Imagine yourself as a pharmacy worker. What would you have done if you were given instructions to take 38.5 tablets and put each one of them in a separate, bar-coded plastic bag?
- What do you see as the ethical issues with this pharmacy robot?

Malicious Actors (Mirsky et al. 2019)

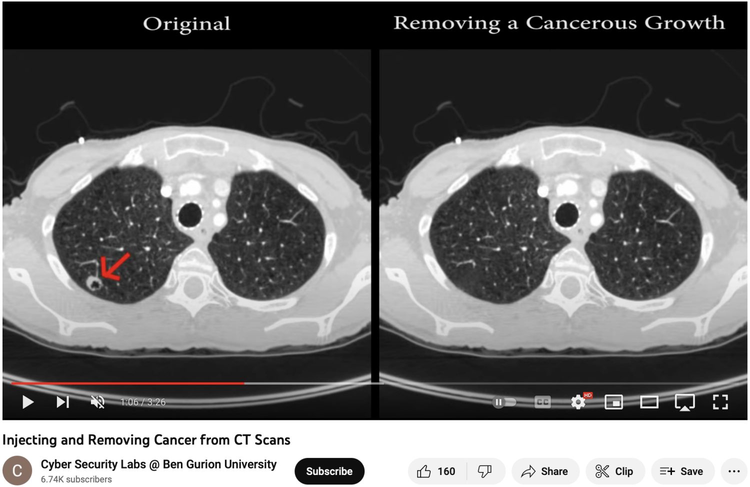
CT-GAN: Malicious Tampering of 3D Medical Imagery using Deep Learning

Yisroel Mirsky[1], Tom Mahler[1], Ilan Shelef[2], and Yuval Elovici[1]
[1]Department of Information Systems Engineering, Ben-Gurion University, Israel
[2]Soroka University Medical Center, Beer-Sheva, Israel
yisroel@post.bgu.ac.il, mahlert@post.bgu.ac.il, shelef@bgu.ac.il, and elovici@bgu.ac.il

Published in the 28th USENIX Security Symposium (USENIX Security 2019)

Demo video with pen-test: https://youtu.be/_mkRAArj-x0
Source code and datasets: https://github.com/ymirsky/CT-GAN

Malicious Actors



Imperfect Decision Making

- Healthcare is filled with imperfect decisions

Confusion matrix

	Actually Positive	Actually Negative
Called Positive	**True Positive**, power, **sensitivity**, recall, hit rate	**False Positive**, type I error, false alarm
Called Negative	**False Negative**, type II error	**True Negative**, **specificity**

Imperfect Models (Play along!)

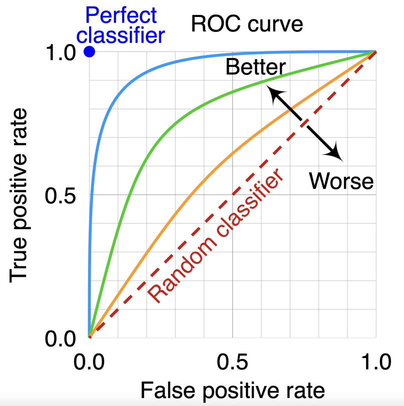
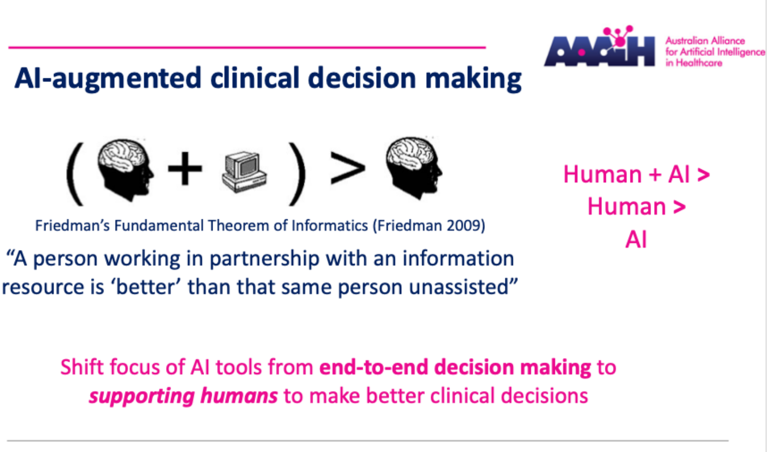


Figure 5: Wikipedia

Less Imperfect Models [6]



[6] Karin Verspoor

AI in Mammography (Satariano, Metz, and Akos 2023)

A.I. and Chatbots › | How Schools Can Survive A.I. | When Will the U.S. Regulate A.I.? | Smart Ways to Use Chatbots | Can A.I. Be Fooled?

Using A.I. to Detect Breast Cancer That Doctors Miss

Hungary has become a major testing ground for A.I. software to spot cancer, as doctors debate whether the technology will replace them in medical jobs.

"An A.I.-plus-doctor should replace doctor alone, but an A.I. should not replace the doctor," Mr. Kecskemethy said.
The National Cancer Institute has estimated that about 20 percent of breast cancers are missed during screening mammograms.
Dr. Constance Lehman, a professor of radiology at Harvard Medical School and a breast imaging specialist at Massachusetts General Hospital, urged doctors to keep an open mind.
"We are not irrelevant," she said, "but there are tasks that are better done with computers."

AI in Mammography

- What do you think about missing 20% of breast cancers?

AI in Mammography

- What do you think about missing 20% of breast cancers?
- Maybe that is actually high. “The result is mammography sensitivity of 60%, specificity of 80%, AUC of 0.73.” (Fitzjohn, Zhou, and Chase 2023)

AI in Mammography

- What do you think about missing 20% of breast cancers?
- Maybe that is actually high. “The result is mammography sensitivity of 60%, specificity of 80%, AUC of 0.73.” (Fitzjohn, Zhou, and Chase 2023)
- What would you do to improve this?

AI in Mammography (Jackson 2023)

Article • **Data & Analytics**

AI used to read breast cancer screenings a safe success

By **Amber Jackson**

August 03, 2023 • 4 mins

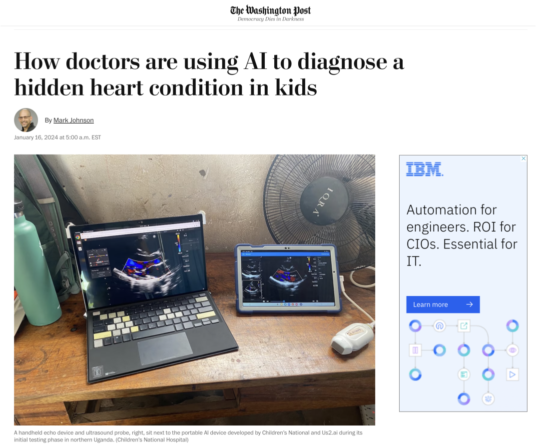
_Lead author of the study Dr Kristina Lång, from Lund University in Sweden, said: "The greatest potential of AI right now is that it could allow radiologists to be less burdened by the excessive amount of reading.

"While our AI-supported screening system requires at least one radiologist in charge of detection, it could potentially do away with the need for double reading of the majority of mammograms, easing the pressure on workloads and enabling radiologists to focus on more advanced diagnostics while shortening waiting times for patients."_

Collective intelligence and radiology

- What are the ethical issues at play here?
- Should every mammogram have two readers? three readers? and so on?
- Is radiologist+AI more or less ethical than a single radiologist or two radiologists?

How doctors are using AI to diagnose a hidden heart condition in kids


The Washington Post
Democracy Dies in Darkness

How doctors are using AI to diagnose a hidden heart condition in kids

By Mark Johnson
January 16, 2024 at 5:00 a.m. EST

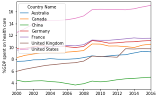
A handheld echo device and ultrasound probe, right, sit next to the portable AI device developed by Children's National and Us2.ai during its initial testing phase in northern Uganda. (Children's National Hospital)

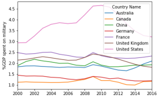
"Now, in an advance that shows the potential of artificial intelligence to aid medicine, researchers at Children's National have developed a new AI-powered tool for diagnosing rheumatic heart disease long before a patient needs surgery. Collaborating with staff at the Uganda Heart Institute, the team designed a system that will allow trained nurses to screen and diagnose children early on, when they can still be treated with penicillin for less than $1 a year. Early treatment could save thousands from having to undergo surgery." (See (Brown et al. 2024) for scientific paper.)

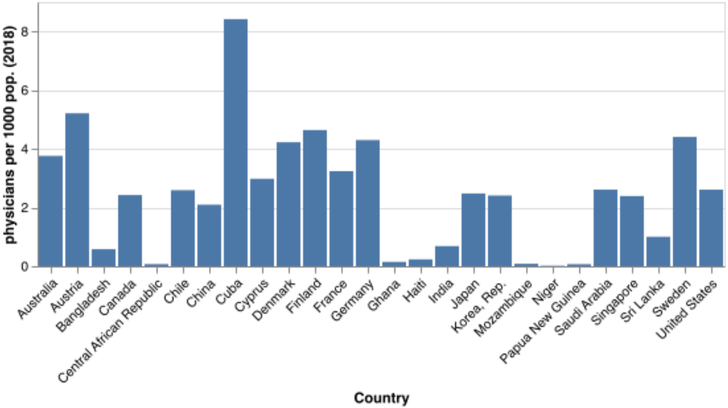
Question

If AI allows less skilled (less expensive) clinicians to accurately obtain clinical data in developing countries, should we let AI help less skilled (less expensive) clinicians to accurately obtain clinical data in developed countries?

Prioritizing AI Research

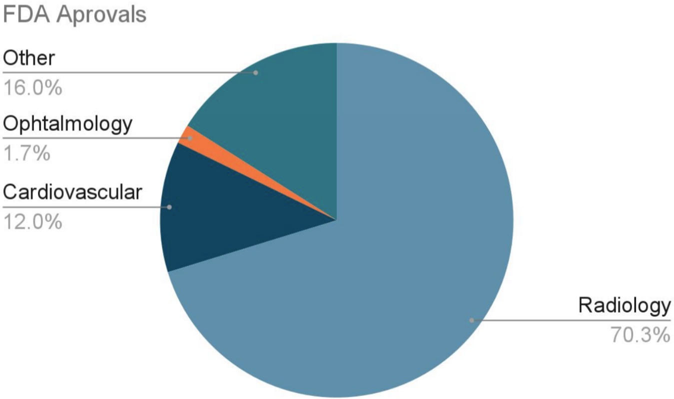






What ethical principles would you invoke regarding whether AI should prioritize reducing the cost of healthcare delivery in developed countries or providing “knowledge coupling” to developing countries?

The most FDA (USA) approved "AI" tools are in imaging (FDA 2022)



Why might imaging applications be so common?

- Are radiologist (relatively) stupid?

Why might imaging applications be so common?

- Are radiologist (relatively) stupid?
- Is radiology (relatively) hard?

Why might imaging applications be so common?

- Are radiologist (relatively) stupid?
- Is radiology (relatively) hard?
- Or is it about the availability of the data?

Example: Decision Support (circa 1970)

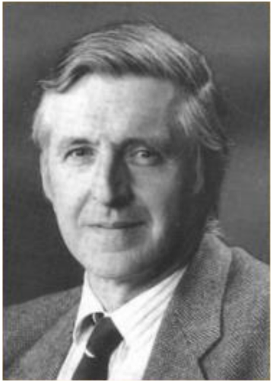


Figure 6: FT de Dombal

- AAPHelp: Bayesian diagnosis of abdominal pain

Example: Decision Support (circa 1970)

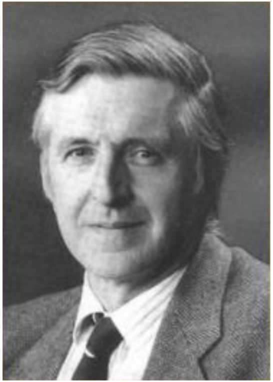


Figure 6: FT de Dombal

- AAPHelp: Bayesian diagnosis of abdominal pain
- "de Dombal's team collected data on thousands of patients who presented with acute abdominal pain. The researchers used data on clinical symptoms (e.g., pain severity, location, and quality) and physical signs (e.g., pulse and abdominal guarding) to derive probabilities for the computer system."

Example: Decision Support (circa 1970)

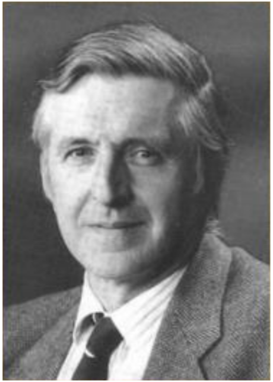


Figure 6: FT de Dombal

- AAPHelp: Bayesian diagnosis of abdominal pain
- "de Dombal's team collected data on thousands of patients who presented with acute abdominal pain. The researchers used data on clinical symptoms (e.g., pain severity, location, and quality) and physical signs (e.g., pulse and abdominal guarding) to derive probabilities for the computer system."
- Evaluated from 1971-1972. Outperformed senior clinicians with a correct diagnosis 91.8% of the time

Example: Decision Support (circa 1970)

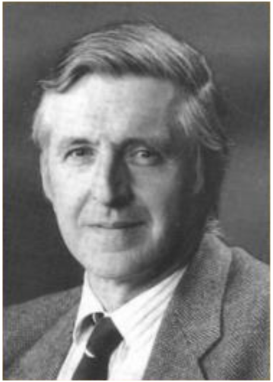


Figure 6: FT de Dombal

- AAPHelp: Bayesian diagnosis of abdominal pain
- "de Dombal's team collected data on thousands of patients who presented with acute abdominal pain. The researchers used data on clinical symptoms (e.g., pain severity, location, and quality) and physical signs (e.g., pulse and abdominal guarding) to derive probabilities for the computer system."
- Evaluated from 1971-1972. Outperformed senior clinicians with a correct diagnosis 91.8% of the time
- Amazing! This should be exported!

Example: Decision Support (circa 1970)

- "But when his group teamed up with researchers at Bispebjerg Hospital in Copenhagen in 1976 to test the system in a fresh clinical environment, its overall accuracy plummeted to 65%."

Example: Decision Support (circa 2023)

- "Sepsis is a life-threatening condition that occurs when the body's response to an infection damages its own tissues and organs. It can start from any type of infection, but rapidly worsens if not treated early, leading to severe complications like organ failure." (GPT4)
- Sepsis is a major research topic in medical AI
- A major international medical software vendor developed a deep learning-based sepsis detector in the USA. It worked remarkably well but when ported to Australia performed remarkably poorly.

What is going on?

- Does Danish biology differ significantly from English biology?
- Are Americans and Australians radically different?

What is going on?

- “databases don’t travel” (de Dombal)

What is going on?

- "databases don't travel" (de Dombal)
- "clinical data are formed, not found." (Lea and Jones 2024)

Data Scarcity (Johnson et al. 2022)

The Ghost in the Machine has an American accent: value conflict in GPT-3.

Rebecca L Johnson
The University of Sydney, Australia.
Rebecca.johnson@sydney.edu.au

Giada Pistilli
Sorbonne Université, France.
Giada.pistilli@paris-sorbonne.fr

Natalia Menédez-González
European University Institute, Spain.
Natalia.menendez@eui.eu

Leslye Denisse Dias Duran
Ruhr Universität Bochum, Germany.
Leslye.diasduran@ruhr-uni-bochum.de

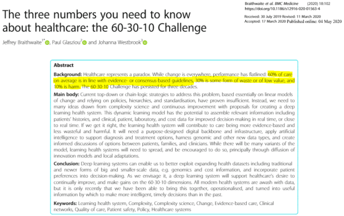
Enrico Panai
University of Sassari, Italy.
enricopanai@gmail.com

Julija Kalpokiene
Vytautas Magnus University, Lithuania.
Julija.kalpokiene@vdu.lt

Donald Jay Bertulfo
Delft University of Technology, Netherlands.
d.j.bertulfo@tudelft.nl

What are the implications for the development and use of AI in the USA (population ~333M) vs. Australia (population ~26M)?

Implications of data-driven AI (Braithwaite, Glasziou, and Westbrook 2020)



Braithwaite *et al. BMC Medicine* (2020) 18:102
https://doi.org/10.1186/s12916-020-01563-4
Received: 30 July 2019 Revised: 11 March 2020
Accepted: 17 March 2020 Published online: 04 May 2020

The three numbers you need to know about healthcare: the 60-30-10 Challenge

Jeffrey Braithwaite[1*], Paul Glasziou[2] and Johanna Westbrook[3]

Abstract

Background: Healthcare represents a paradox. While change is everywhere, performance has flatlined: 60% of care on average is in line with evidence- or consensus-based guidelines, 30% is some form of waste or of low value, and 10% is harm. The 60-30-10 Challenge has persisted for three decades.

Main body: Current top-down or chain-logic strategies to address this problem, based essentially on linear models of change and relying on policies, hierarchies, and standardisation, have proven insufficient. Instead, we need to marry ideas drawn from complexity science and continuous improvement with proposals for creating a deep learning health system. This dynamic learning model has the potential to assemble relevant information including patients' histories, and clinical, patient, laboratory, and cost data for improved decision-making in real time, or close to real time. If we get it right, the learning health system will contribute to care being more evidence-based and less wasteful and harmful. It will need a purpose-designed digital backbone and infrastructure, apply artificial intelligence to support diagnosis and treatment options, harness genomic and other new data types, and create informed discussions of options between patients, families, and clinicians. While there will be many variants of the model, learning health systems will need to spread, and be encouraged to do so, principally through diffusion of innovation models and local adaptations.

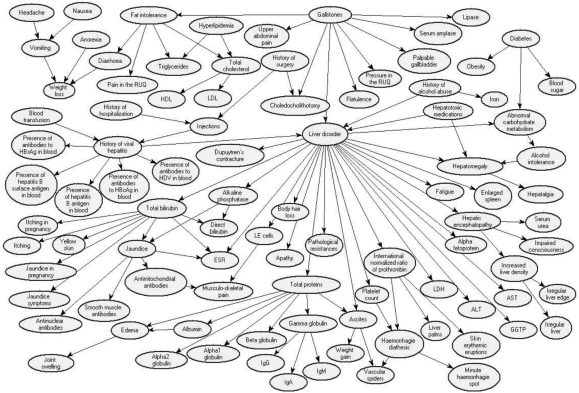
Conclusion: Deep learning systems can enable us to better exploit expanding health datasets including traditional and newer forms of big and smaller-scale data, e.g. genomics and cost information, and incorporate patient preferences into decision-making. As we envisage it, a deep learning system will support healthcare's desire to continually improve, and make gains on the 60-30-10 dimensions. All modern health systems are awash with data, but it is only recently that we have been able to bring this together, operationalised, and turned into useful information by which to make more intelligent, timely decisions than in the past.

Keywords: Learning health system, Complexity, Complexity science, Change, Evidence-based care, Clinical networks, Quality of care, Patient safety, Policy, Healthcare systems

These suboptimal processes will create suboptimal data. Should we be building AI tools based on these data?

Can some forms of AI be more portable?

Should we pay a lot of money to Amazon to heat up GPUs or pay experts to capture their knowledge?



Liver Model: Nodes and Edges

- Nodes without any parents (incoming edges) only require prior probabilities (baseline prevalence in the world you are modeling)
  - E.g. Prevalence of gallstones (0.15)
- Nodes with incoming edges require the conditional probabilities ($P(x|\text{parents})$)
  - E.g. $P(\text{liver disease}|\text{gallstones, Hx alchohol abuse, Hx viral hepatitis, Hx hepatotoxic medications})$
- Direction of edges are not statements about causality. They reflect how we chose to factor the joint probability distribution, which, however, might reflect our understanding of causality

How was the model created?

We elicited the structure of dependencies among the variables from our domain experts: Dr. Hanna Wasyluk (third author) from the Medical Center of Postgraduate Education, and two American experts, a pathologist, Dr. Daniel Schwartz, and an epidemiologist, Dr. John N. Dowling, both at the University of Pittsburgh. ***We estimate that elicitation of the structure took about 40 hours with the experts,*** *30 hours with Dr. Wasyluk and 10 hours with Drs. Schwartz and Dowling. This includes model refinement sessions, where previously elicited structure was reevaluated in a group setting. We started with an initial model comprising 40 variables of the highest diagnostic value (according to the expert) and gradually extended it by adding variables one at a time. ((Onisko et al. 1999))*

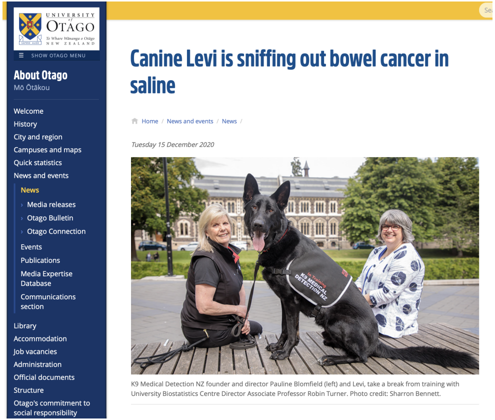
What about the parameters (probabilities)?

- These were calculated using data from their EMR
- Bayesian networks (structure and algorithms) greatly reduces the number of probabilities we need

Deep Learning vs Bayesian networks

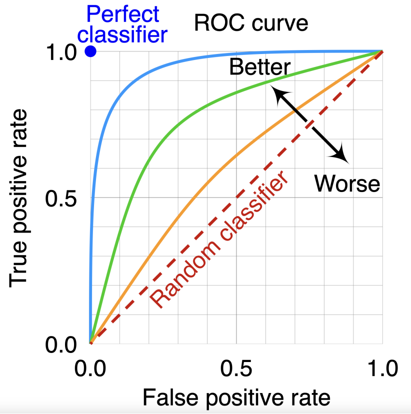
- Deep neural networks generally perform better than alternative models (e.g. Bayesian networks). However, they tend to be brittle, black boxes, that may not travel well. How much gain in diagnostic performance (AUC) would you trade off for explainability, portability, etc.?

Black Box AI 1



Do you have any concerns about Levi functioning in the healthcare system? Would you feel differently if it was an algorithm?

Mammography and explanation I



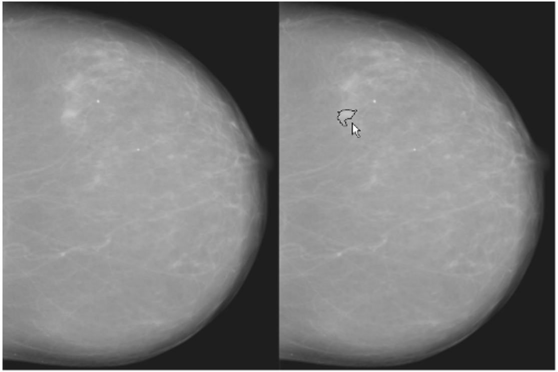
Scenario: Dr. Ziggy Stardust has developed an AI tool for mammography. It generates a simple "positive" (the patient has cancer) or "negative" (the patient does not have cancer) with a high AUC (0.9) but does not provide any explanation of why or where it is positive.

- Does this tool have any clinical value?
- How would you tune the sensitivity and specificity of the test to maximize its clinical value?

Is it all in the context?

- The utility of Dr. Stardust's tool seems to lie in the context of what comes next.
- Consider a sepsis tool that accurately (but not perfectly) predicts a patient is about to develop sepsis and die, but doesn't tell you anything else. Could this have value? Harms?

Mammography and explanation II



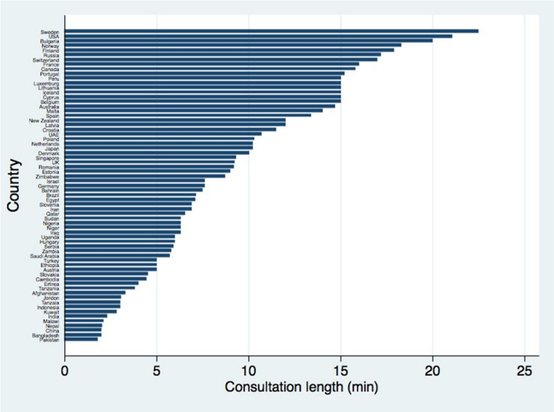
Scenario: Dr. Taylor Swift has developed an AI tool for mammography. It identifies suspicious masses in the mammogram and outlines them. The outlines are a form of explanation. The tool performs well with a very good AUC (e.g. >0.95). However, the outlines of the massses are very different from the outlines human radiologists create for the same masses and human performance is worse when provided the outline vs. just a cross-hair at the location of the mass.

Should the explanation be provided?

How to merge/partner with information resource is an open question (Cabitza et al. 2023)

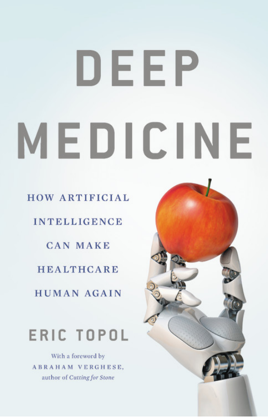
- AI support was found useful but XAI was associated with a null or detrimental effect.
- AI-first protocols had higher accuracy than human-first ones and humans or AI alone.

Scenario: Using LLMs to generate discharge summaries



- A time-consuming task in healthcare is writing a discharge summary?
- What are the benefits and liabilities of having
    - An experienced physician write the discharge summary?
    - A medical student write the discharge summary?
    - A large language model write the discharge summary?

Can AI help healthcare be more humane, more human centric?



- Eric Topol, M.D.: Hopes AI will give the gift of time
- Brian Chapman, Ph.D.: Hopes AI will give patients the gift of knowledge
    - Allowing/Demanding patients be treated as persons

Extended mind or cognitive outsourcing (Danaher 2018)

As Selinger and Frischmann (2016) have recently noted, usage of AI assistance is effectively a new form of outsourcing. ***Humans have long outsourced the performance of cognitive tasks to others.*** *I don't do my tax returns; my accountant does. I don't book my travel arrangements; my assistant does. Such humanistic outsourcing has its own ethical issues. . . .Michael Sandel (2012) has argued, there are some tasks that seem to ethically demand my personal involvement. For instance, outsourcing the writing of a best man's speech seems like a mark of disrespect and apathy, not a praiseworthy efficiency-maximising way to fulfil one's duties.*

Cognitive Outsourcing (Danaher 2018)

*One of those knock-on effects, according to Carr, **is that increased reliance on AI assistance will atrophy and degenerate our mental faculties.** So, far from freeing up mental resources, increased reliance on AI assistance will deplete mental resources. **We will no longer have the ability to think the important thoughts.** This will in turn reduce the quality of our personal lives because the ability to engage in deep thinking is both intrinsically and instrumentally valuable: it results in a better immediate conscious experience and engagement with life, and it helps one to solve personal problems.*

Question:

Should we think about what we are doing?

Limited Cognitive Capacity: Alfred North Whitehead

It is a profoundly erroneous truism, repeated by all copy-books and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. ***Civilization advances by extending the number of important operations which we can perform without thinking about them.*** *Operations of thought are like cavalry charges in a battle — they are strictly limited in number, they require fresh horses, and must only be made at decisive moments.* (Whitehead 1958)

Thank you

References I

Braithwaite, J., P. Glasziou, and J. Westbrook. 2020. "The three numbers you need to know about healthcare: the 60-30-10 Challenge." *BMC Med* 18 (1): 102.

Bringsjord, Selmer, and Naveen Sundar Govindarajulu. 2022. "Artificial Intelligence." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Fall 2022. https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/; Metaphysics Research Lab, Stanford University.

Brown, K., P. Roshanitabrizi, J. Rwebembera, E. Okello, A. Beaton, M. G. Linguraru, and C. A. Sable. 2024. "Using Artificial Intelligence for Rheumatic Heart Disease Detection by Echocardiography: Focus on Mitral Regurgitation." *J Am Heart Assoc* 13 (2): e031257.

References II

Cabitza, Federico, Andrea Campagner, Luca Ronzio, Matteo Cameli, Giulia Elena Mandoli, Maria Concetta Pastore, Luca Maria Sconfienza, Duarte Folgado, Marília Barandas, and Hugo Gamboa. 2023. "Rams, Hounds and White Boxes: Investigating Human–AI Collaboration Protocols in Medical Diagnosis." *Artificial Intelligence in Medicine* 138: 102506. https://doi.org/https://doi.org/10.1016/j.artmed.2023.102506.

Clark, A. 2003. *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence. Oxford University Press. https://books.google.com.au/books?id=8JXaK3sREXQC.

D., LEE. 1977. "Experts Argue Whether Computers Could Reason, and If They Should: Experts Dispute Whether Computers Could Reason." *New York Times (1923-)*. https://www.proquest.com/historical-newspapers/experts-argue-whether-computers-could-reason-if/docview/123325923/se-2.

References III

Danaher, John. 2018. "Toward an Ethics of Ai Assistants: An Initial Framework." *Philosophy and Technology* 31 (4): 629–53. https://doi.org/10.1007/s13347-018-0317-3.

FDA. 2022. "Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices." U.S. Food & Drug Administration. October 5, 2022. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices.

Fitzjohn, Jessica, Cong Zhou, and J. Geoffrey Chase. 2023. "Critical Assessment of Mammography Accuracy." *IFAC-PapersOnLine* 56 (2): 5620–25. https://doi.org/https://doi.org/10.1016/j.ifacol.2023.10.472.

Irving, G., A. L. Neves, H. Dambha-Miller, A. Oishi, H. Tagashira, A. Verho, and J. Holden. 2017. "International variations in primary care physician consultation time: a systematic review of 67 countries." *BMJ Open* 7 (10): e017902.

References IV

Jackson, Amber. 2023. "AI Used to Read Breast Cancer Screenings a Safe Success." *AI Magazine*. https://aimagazine.com/articles/ai-used-to-read-breast-cancer-screenings-a-safe-success.

Johnson, Rebecca L, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. "The Ghost in the Machine Has an American Accent: Value Conflict in GPT-3." https://arxiv.org/abs/2203.07785.

Lea, A. S., and D. S. Jones. 2024. "Mind the Gap - Machine Learning, Dataset Shift, and History in the Age of Clinical Algorithms." *N Engl J Med* 390 (4): 293–95.

lez, A., M. Mahesh, K. P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land. 2009. "Projected cancer risks from computed tomographic scans performed in the United States in 2007." *Arch Intern Med* 169 (22): 2071–77.

Licklider, J. C. R. 1960. "Man-Computer Symbiosis." *IRE Transactions on Human Factors in Electronics* HFE-1 (1): 4–11. https://doi.org/10.1109/THFE2.1960.4503259.

References V

McGrayne, S. B. 2011. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted down Russian Submarines, & Emerged Triumphant from Two Centuries of c.* Matematicas (e-Libro). Yale University Press. https://books.google.com.au/books?id=_Kx5xVGuLRIC.

Medicine, I., L. A. M. Olsen, E. G. Nabel, J. M. McGinnis, and M. B. McClellan. 2008. *Evidence-Based Medicine and the Changing Nature of Health Care: 2007 IOM Annual Meeting Summary.* Learning Healthcare System Series. National Academies Press. https://books.google.com.au/books?id=xA5kAgAAQBAJ.

Meer, A. B., P. A. Basu, L. C. Baker, and S. W. Atlas. 2012. "Exposure to ionizing radiation and estimate of secondary cancers in the era of high-speed CT scanning: projections from the Medicare population." *J Am Coll Radiol* 9 (4): 245–50.

References VI

Mirsky, Yisroel, Tom Mahler, Ilan Shelef, and Yuval Elovici. 2019. "CT-GAN: Malicious Tampering of 3D Medical Imagery Using Deep Learning." In *Proceedings of the 28th USENIX Conference on Security Symposium*, 461–78. SEC'19. USA: USENIX Association.

Mitchell, M. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus; Giroux. https://books.google.com.au/books?id=65iEDwAAQBAJ.

Onisko, Agnieszka, Marek J Druzdzel, Hanna Wasyluk, et al. 1999. "A Bayesian Network Model for Diagnosis of Liver Disorders." In *Proceedings of the Eleventh Conference on Biocybernetics and Biomedical Engineering*, 2:842–46. Citeseer.

Satariano, Adam, Cade Metz, and Stiller F. Akos. 2023. "Using A.I. To Detect Breast Cancer That Doctors Miss." New York Times (Online). 2023. https://www.proquest.com/blogs-podcasts-websites/using-i-detect-breast-cancer-that-doctors-miss/docview/2782623735/se-2.

References VII

Shortliffe, E. H., S. G. Axline, B. G. Buchanan, T. C. Merigan, and S. N. Cohen. 1973. "An artificial intelligence program to advise physicians regarding antimicrobial therapy." *Comput Biomed Res* 6 (6): 544–60.

Solomon, M. 2015. *Making Medical Knowledge*. Oxford University Press. https://books.google.com/books?id=___dsBwAAQBAJ.

Spackman, K. A. 1985. "A program for machine learning of counting criteria: empirical induction of logic-based classification rules." *Comput Methods Programs Biomed* 21 (3): 221–26.

Wachter, R. 2015. *The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age*. McGraw Hill LLC. https://books.google.com/books?id=qO-VBgAAQBAJ.

WARNER, H. R., A. F. TORONTO, L. G. VEASEY, and R. STEPHENSON. 1961. "A mathematical approach to medical diagnosis. Application to congenital heart disease." *JAMA* 177 (July): 177–83.

References VIII

Whitehead, A. N. 1958. *An Introduction to Mathematics*. A Galaxy Book, GB18. Oxford University Press. https://books.google.com.au/books?id=JmKM_0EZm5AC.