In A.I.
We Trust !

In A.I.
We Trust ?

A.I. provides great potential to aid in medical decision making

A.I. provides great potential to aid in medical decision making

A.I. in Medicine: it has failed before, why it would be different this time?

“It seems probable that in the not too distant future the physician and the computer will engage in frequent dialogue, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action.”

– William B. Schwartz

“It seems probable that in the not too distant future the physician and the computer will engage in frequent dialogue, the computer continuously taking note of history, physical findings, laboratory data, and the like, alerting the physician to the most probable diagnoses and suggesting the appropriate, safest course of action.”

– William B. Schwartz

Medicine and the computer. The promise and the problems of change. N Engl J Med. 1970;283:1257–64.

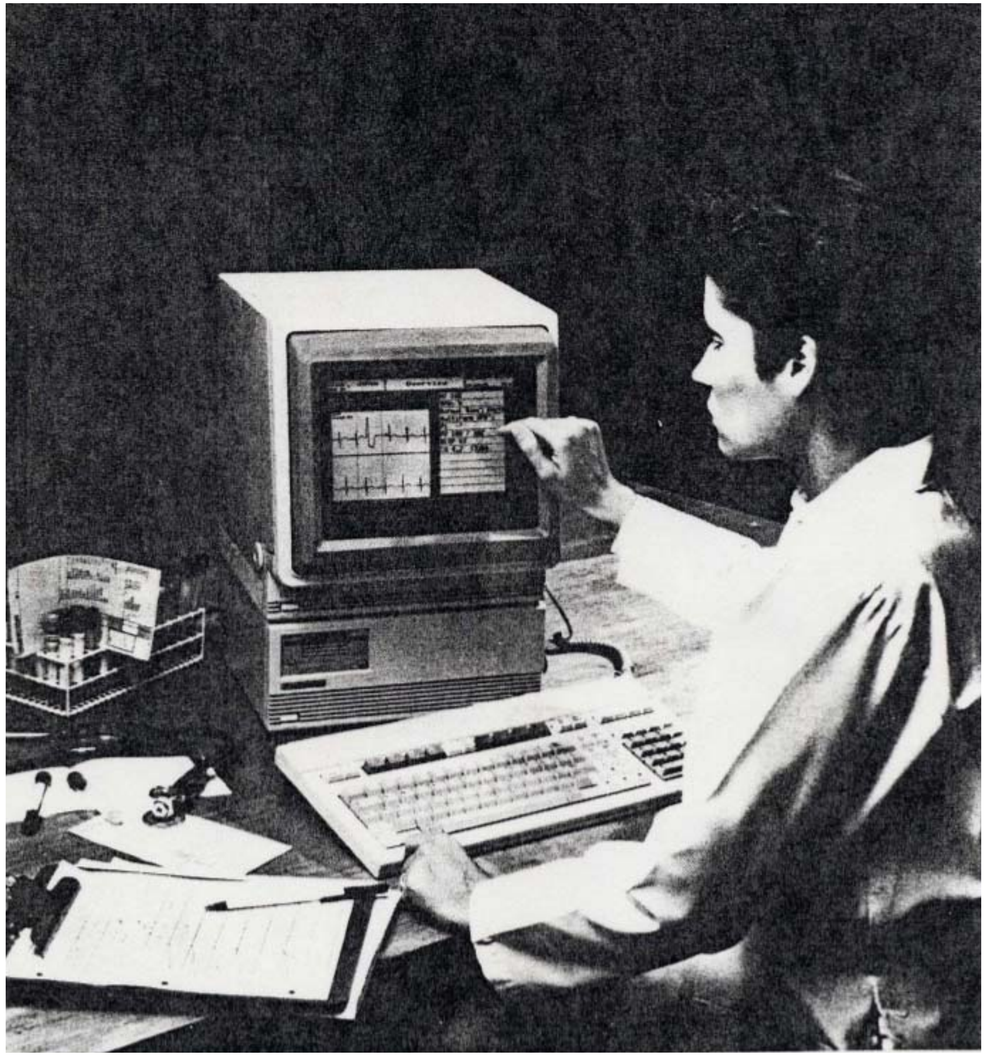
“Machines will be capable, within twenty years, of doing any work a man can do.”

Herbert Simon, 1965

“In from three to eight years we will have a machine with the general intelligence of an average human being.”

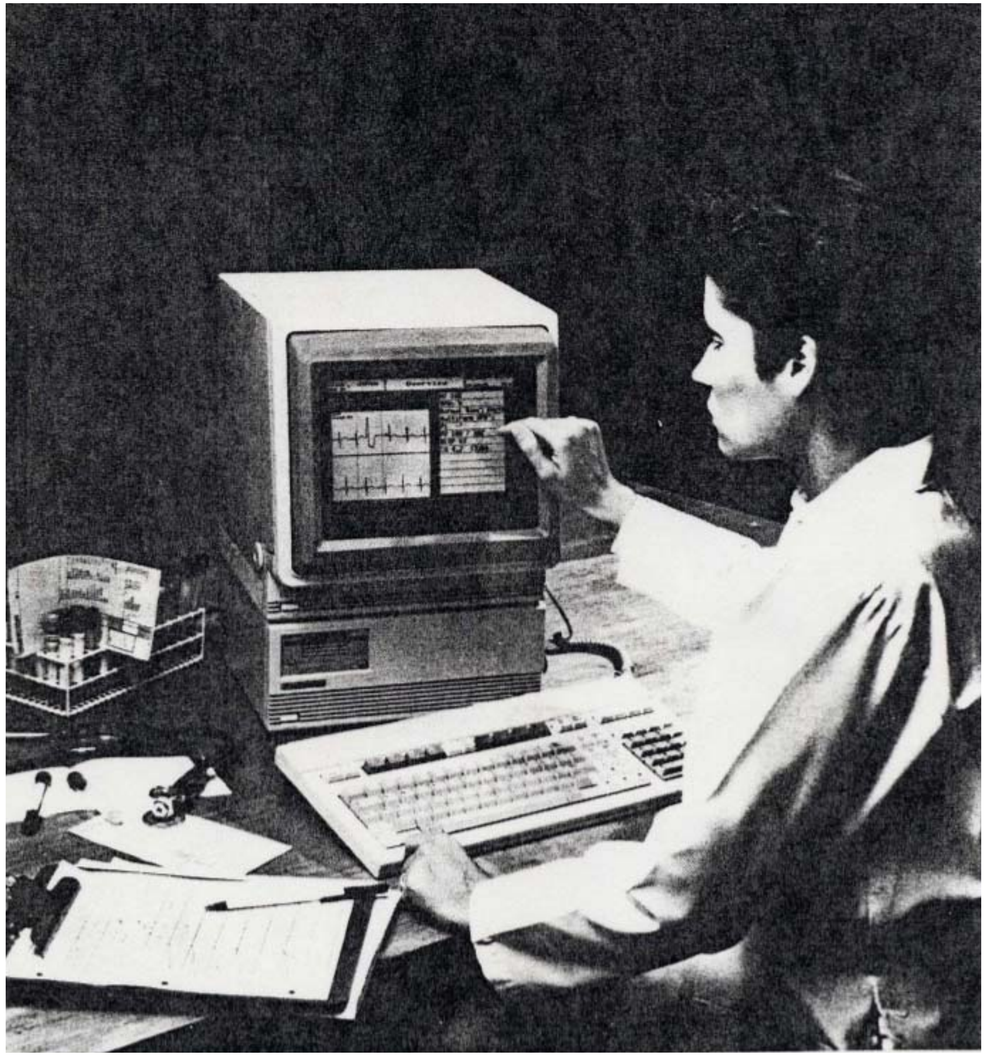
Marvin Minsky, 1970

The term “Artificial Intelligence” was first coined in 1956 at a conference at Dartmouth College.



The term “Artificial Intelligence” was first coined in 1956 at a conference at Dartmouth College.

“People were impressed by the memory, speed, and logical manipulations of the computer. Some felt that in [medical] decision making computers would rival the human brain.”[1]

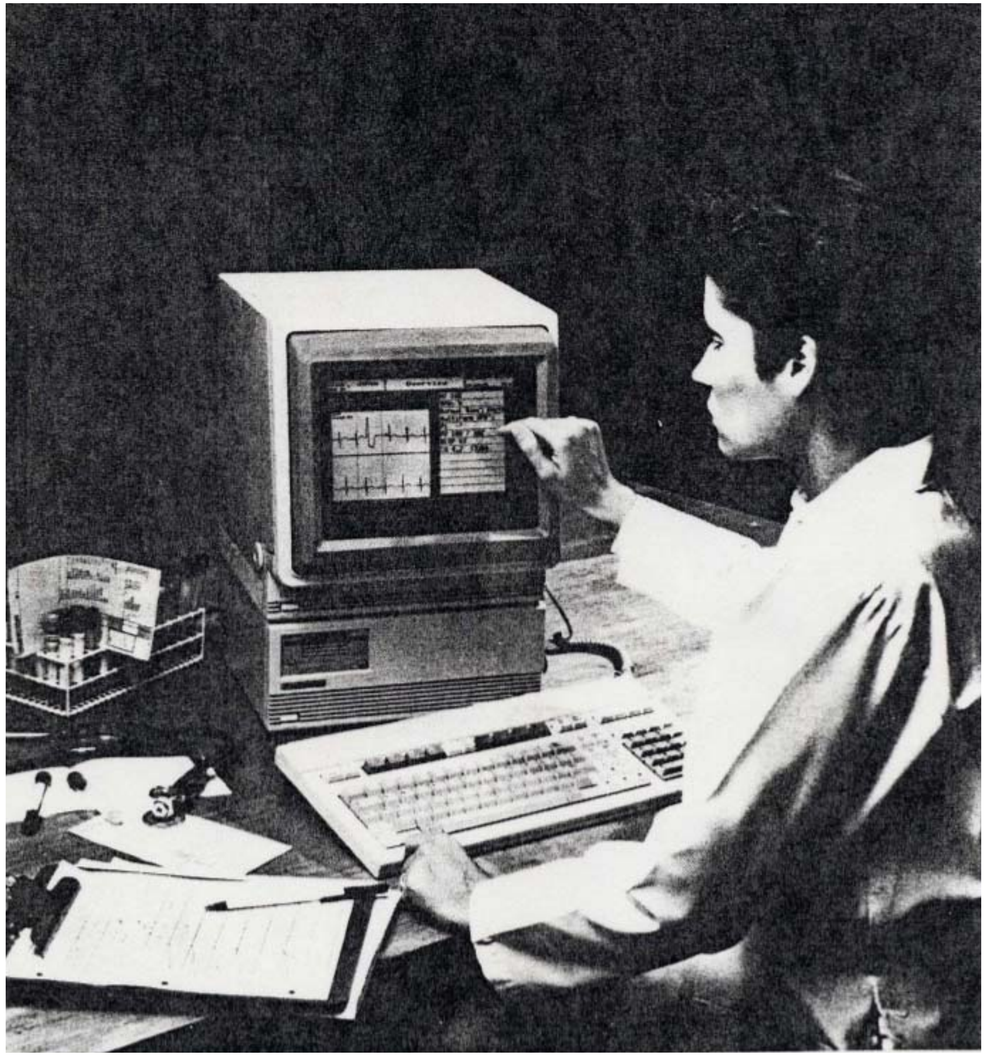


1. Engle RL Jr. Attempts to use computers as diagnostic aids in medical decision making: A thirty-year experience. Perspect Biol Med. 1992;35(2):207–19.

The term “Artificial Intelligence” was first coined in 1956 at a conference at Dartmouth College.

“People were impressed by the memory, speed, and logical manipulations of the computer. Some felt that in [medical] decision making computers would rival the human brain.”[1]

By end of 1960’s, there was a proliferation of medical diagnostic systems based on early breakthroughs in AI, e.g., MYCIN, HEME.



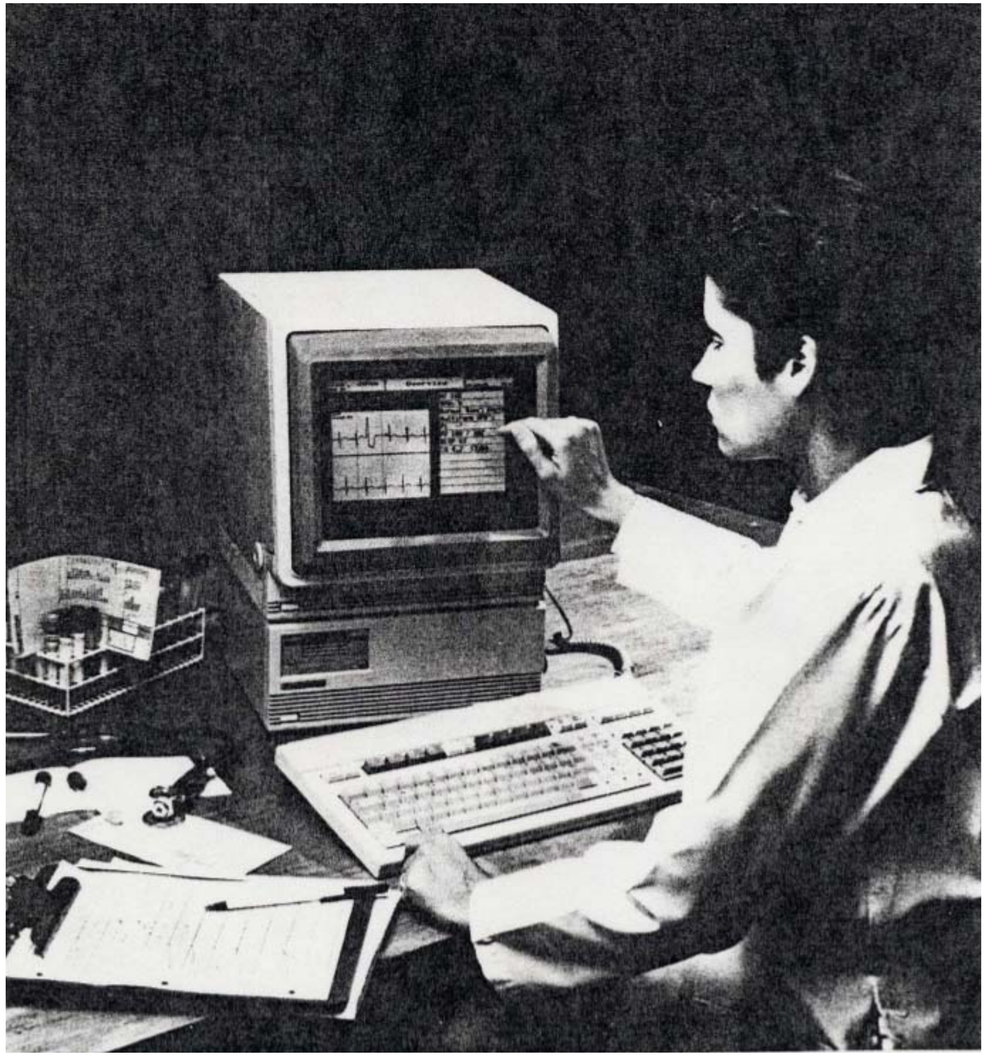
1. Engle RL Jr. Attempts to use computers as diagnostic aids in medical decision making: A thirty-year experience. Perspect Biol Med. 1992;35(2):207–19.

The term “Artificial Intelligence” was first coined in 1956 at a conference at Dartmouth College.

“People were impressed by the memory, speed, and logical manipulations of the computer. Some felt that in [medical] decision making computers would rival the human brain.”[1]

By end of 1960’s, there was a proliferation of medical diagnostic systems based on early breakthroughs in AI, e.g., MYCIN, HEME.

Most of these systems failed to be adopted.



1. Engle RL Jr. Attempts to use computers as diagnostic aids in medical decision making: A thirty-year experience. Perspect Biol Med. 1992;35(2):207–19.

ATTEMPTS TO USE COMPUTERS AS DIAGNOSTIC AIDS IN MEDICAL DECISION MAKING: A THIRTY-YEAR EXPERIENCE

*RALPH L. ENGLE, JR.**

Introduction

In the 1950s, when electronic computers were first coming into use, there was already great interest in having the computer perform functions that had previously been performed only by the human brain. People were impressed by the memory, speed, and logical manipulations of the computer. Some felt that in decision making computers would rival the human brain. Some were even concerned that the computer of the future would be making all of the important decisions in certain fields. Still others felt that computers would never be able to match the decision-making capabilities of the brain. Yet all agreed that computers would be useful in various ways to decision making. We still do not have final answers to this basic question of the computer's ability to make decisions, and most experts have not changed their minds about it.

Computers have found an important role as aids to decision making in the field of radiology, for example, where many modern imaging techniques depend entirely on their capabilities. Also, computers as analytic tools have been used to aid in decision making in many fields of medicine, as in the interpretation of electrocardiograms. Even less direct

Will the Wave Finally Break? A Brief View of the Adoption of Electronic Medical Records in the United States

Eta S. Berner, EdD, Don E. Detmer, MD, MA, Donald Simborg, MD

Abstract For over thirty years, there have been predictions that the widespread clinical use of computers was imminent. Yet the "wave" has never broken. In this article, two broad time periods are examined: the 1960's to the 1980's and the 1980's to the present. Technology immaturity, health administrator focus on financial systems, application "unfriendliness," and physician resistance were all barriers to acceptance during the early time period. Although these factors persist, changes in clinicians' economics, more computer literacy in the general population, and, most importantly, changes in government policies and increased support for clinical computing suggest that the wave may break in the next decade.

■ **J Am Med Inform Assoc.** 2005;12:3–7. DOI 10.1197/jamia.M1664.

In a series of reports over the last 15 years, the Institute of Medicine (IOM) has highlighted that wider use of information technology in health care is essential for major improvements in the quality of care.[1-6] The publicity surrounding these reports has led to an increased interest in electronic medical records (EMR), including computer-based physician, or provider, order entry (CPOE) and clinical decision support systems (CDSS), all of which were the focus of some of the earliest work in medical informatics. The prospect of finally seeing these systems in widespread use is gratifying to those who have labored for more than 35 years to develop and promote them, but this latest flush of interest is not the first time such enthusiasm has flourished. In 1970, Schwartz,[7] proposed that it was probable that clinical computing would be commonplace "in the not too distant future." Yet almost 35 years later, when CPOE use is estimated to have at best approximately 10% market penetration,[8] we are again speculating over having reached a "tipping point." The title of this article comes from one of the authors (DS) who referred in a 2004 American College of Medical Informatics presentation to the prediction of the widespread use of the electronic medical record as "the wave that never breaks." In this article we discuss some of the factors that at different times seemed to presage the "age of clinical computing" and explore reasons why there may be more reason for optimism in today's health care climate than at earlier times.

The Beginning of Computers in Healthcare—1960s to 1980s

Computers were first used for administrative and fiscal functions in hospital settings in the early 1960s, following prior use in business and in research settings. At the same time, the early work in medical informatics focused on clinical computing with a clear goal—to improve clinical decisions and reduce medical errors—essentially through electronic access to procedure results, faster access to relevant medical information in the literature, and, from the beginning, decision support functions such as reminders and alerts.[9,10] It had been widely hypothesized that physicians' errors of omission and commission were at least as frequently related to their lack of information about the patient as they were to lack of

Technology immaturity

Physician resistance

Application “unfriendliness”

Lack of financial incentives for adoption

Technology immaturity

Physician resistance

Application “unfriendliness”

Lack of financial incentives for adoption

Technology immaturity

Physician resistance

Application “unfriendliness”

“Our experience confirms the great difficulty, and even the impossibility, of incorporating the complexity of human thought into a system that can be handled by a computer.”

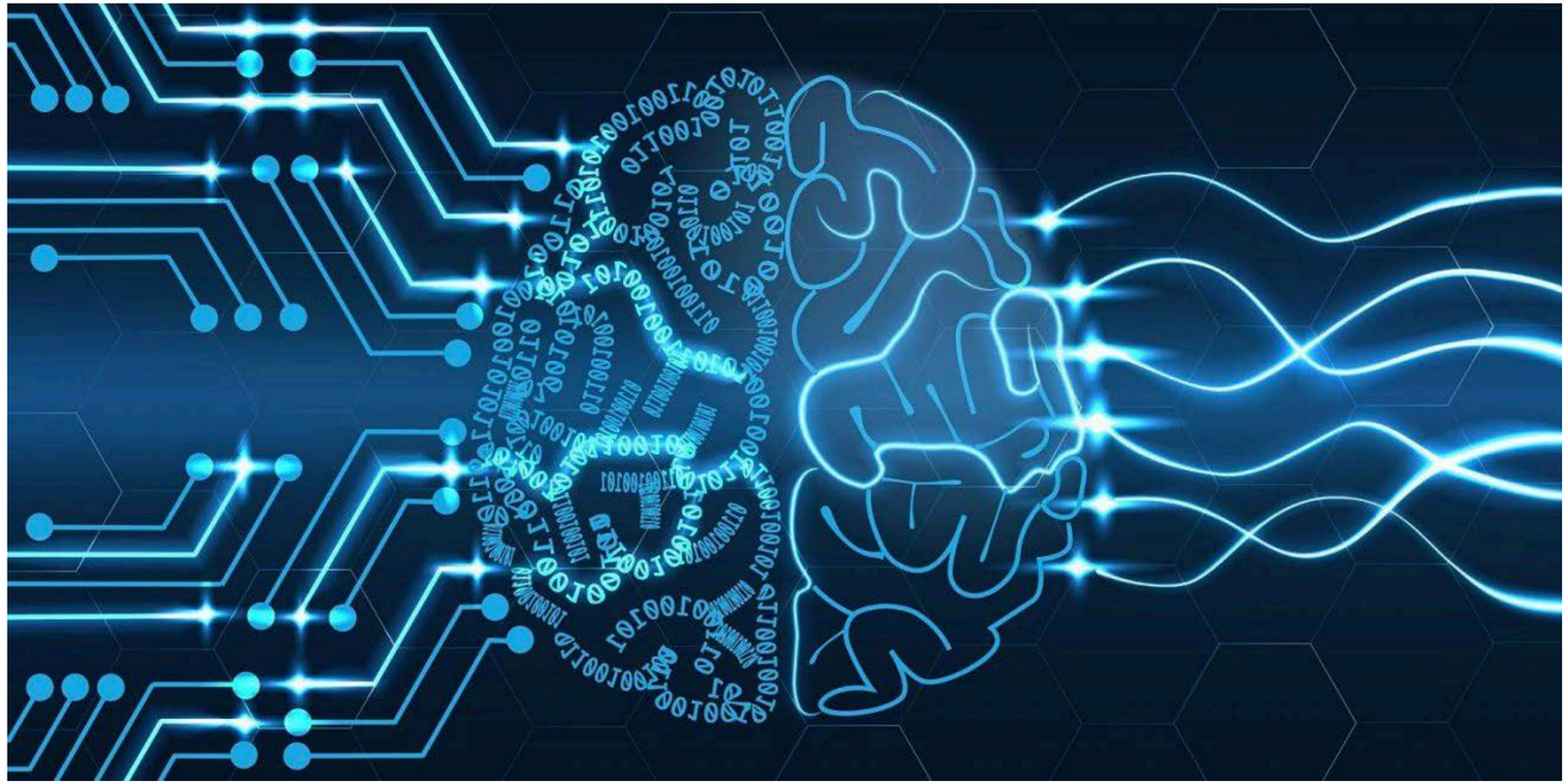
“Our experience confirms the great difficulty, and even the impossibility, of incorporating the complexity of human thought into a system that can be handled by a computer.”

“We concluded that we should stop trying to make a computer act like a diagnostician and concentrate instead on ways of making computer-generated relevant information available to physicians as they make decisions.”

More sophisticated neural network models

Significantly improved computing power

Better health information systems for collecting and managing clinical data



✔ Technology immaturity

Physician resistance

Application “unfriendliness”

✔ Technology immaturity

Physician resistance

Application “unfriendliness”

“Physicians continued to rely on their own autonomy and authority and may not have wanted to use decision support systems even if they were available. Often, it appeared that … those that potentially diminished the doctor’s independence were resisted.”

“Physicians continued to rely on their own autonomy and authority and may not have wanted to use decision support systems even if they were available. Often, it appeared that … those that potentially diminished the doctor’s independence were resisted.”

“In 1982, Komaroff characterized the prevailing attitude as, ‘I am being regimented if you give algorithms to me, but I am being systematic if I develop algorithms for myself.’”The intrinsic rigidity of such models reinforced the perception of regimentation.

✔ Technology immaturity

✘ Physician resistance

Application “unfriendliness”

✔ Technology immaturity

✘ Physician resistance

Application “unfriendliness”

Even when systems were clearly demonstrated to have improved quality and cost, if they impeded clinicians' workflow they were not widely adopted.

Even when systems were clearly demonstrated to have improved quality and cost, if they impeded clinicians' workflow they were not widely adopted.

The early decision support systems functioned more like the "Greek Oracle" rather than permitting a more flexible, interactive approach. The intrinsic rigidity of such models reinforced the perception of regimentation.

Review Paper ■

Overriding of Drug Safety Alerts in Computerized Physician Order Entry

Heleen van der Sijs, MSc, RPh, Jos Aarts, PhD, Arnold Vulto, MSc, RPh, PhD, Marc Berg, MD, MA, PhD

Abstract Many computerized physician order entry (CPOE) systems have integrated drug safety alerts. The authors reviewed the literature on physician response to drug safety alerts and interpreted the results using Reason's framework of accident causation. In total, 17 papers met the inclusion criteria. Drug safety alerts are overridden by clinicians in 49% to 96% of cases. Alert overriding may often be justified and adverse drug events due to overridden alerts are not always preventable. A distinction between appropriate and useful alerts should be made. The alerting system may contain error-producing conditions like low specificity, low sensitivity, unclear information content, unnecessary workflow disruptions, and unsafe and inefficient handling. These may result in active failures of the physician, like ignoring alerts, misinterpretation, and incorrect handling. Efforts to improve patient safety by increasing correct handling of drug safety alerts should focus on the error-producing conditions in software and organization. Studies on cognitive processes playing a role in overriding drug safety alerts are lacking.

■ **J Am Med Inform Assoc.** 2006;13:138–147. DOI 10.1197/jamia.M1809.

✔ Technology immaturity

✘ Physician resistance

✘ Application “unfriendliness”

✔ Technology immaturity

✘ Physician resistance

✘ Application “unfriendliness”

✔ Technology immaturity

✘ Physician resistance

✘ Application “unfriendliness”

? Lack of financial incentives for adoption

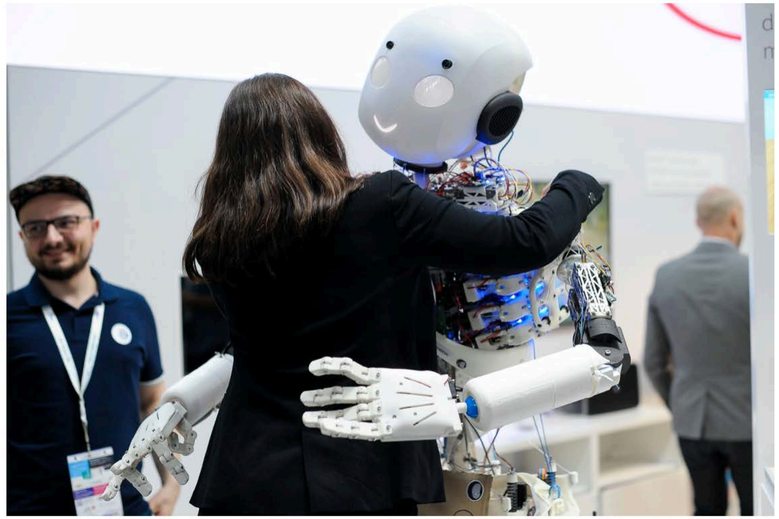
Some New, Known Obstacles in the A.I. 2.0 Era

nytimes.com

The New York Times

SUBSCRIBE NOW

Warnings of a Dark Side to A.I. in Health Care



Scientists worry that with just tiny tweaks to data, neural networks can be fooled into committing "adversarial attacks" that mislead rather than help.
Joan Cros/NurPhoto, via Getty Images

By Cade Metz and Craig S. Smith

March 21, 2019

healthitanalytics.com

HealthITAnalytics

Could Artificial Intelligence Do More Harm Than Good in Healthcare?

The benefits of using artificial intelligence in healthcare are well-documented, but the industry will also have to consider the possible pitfalls.

Source: Getty Images

By Jessica Kent

June 25, 2019 - In recent years, artificial intelligence has rapidly become the chief topic of conversation among healthcare executives, vendors, and

Michigan Medicine - PKUHSC
JOINT INSTITUTE
for Translational and Clinical Research

Unexplainable A.I. (the “Greek Oracle” problem again)

Incomplete or inaccurate training data (garbage in, garbage out)

Systematic biases in training data

Transformability

Over-control and over-reliance

Unexplainable A.I. (the “Greek Oracle” problem again)

Incomplete or inaccurate training data (garbage in, garbage out)

Systematic biases in training data

Transformability

Over-control and over-reliance

Research Paper ■

Comparison of Methodologies for Calculating Quality Measures Based on Administrative Data versus Clinical Data from an Electronic Health Record System: Implications for Performance Measures

PAUL C. TANG, MD, MS, MARY RALSTON, PHD, MICHELLE FERNANDEZ ARRIGOTTI, MPH, LUBNA QURESHI, MS, JUSTIN GRAHAM, MD, MS

Abstract New reimbursement policies and pay-for-performance programs to reward providers for producing better outcomes are proliferating. Although electronic health record (EHR) systems could provide essential clinical data upon which to base quality measures, most metrics in use were derived from administrative claims data. We compared commonly used quality measures calculated from administrative data to those derived from clinical data in an EHR based on a random sample of 125 charts of Medicare patients with diabetes. Using standard definitions based on administrative data (which require two visits with an encounter diagnosis of diabetes during the measurement period), only 75% of diabetics determined by manually reviewing the EHR (the gold standard) were identified. In contrast, 97% of diabetics were identified using coded information in the EHR.

The discrepancies in identified patients resulted in statistically significant differences in the quality measures for frequency of HbA1c testing, control of blood pressure, frequency of testing for urine protein, and frequency of eye exams for diabetic patients. New development of standardized quality measures should shift from claims-based measures to clinically based measures that can be derived from coded information in an EHR. Using data from EHRs will also leverage their clinical content without adding burden to the care process.

■ **J Am Med Inform Assoc.** 2007;14:10–15. DOI 10.1197/jamia.M2198.

Background

Since the Institute of Medicine's (IOM's) 2001 call to "cross the quality chasm,"[1] many major health systems have put in place programs to improve the quality of health care in America.[2] To support these quality improvement efforts, there has been a major push to promote the adoption and use of electronic health record (EHR) systems by clinicians,[3] and personal health record systems by patients.[4] Larger physician practices have made significant strides in deploying EHR systems, but penetration of EHR use among smaller practices lags behind.[5]

Major payers are using payment incentives to motivate providers to demonstrate that they have achieved improved quality.[6] Increasingly, providers are asked to submit process and outcomes data—using different data definitions and different reporting formats—to be used in pay-for-performance programs, quality improvement initiatives, and other public-reporting endeavors. At the heart of any improvement activity must be accurate, reliable, standardized, and cost-effective means for measuring current performance and for setting desired performance goals. While the number of quality measures in use has increased substantially over recent years, debate is surfacing as to whether these diverse measures of quality have actually led to improvements.[6] Some assert that only modest improvements have been achieved since the IOM's *Quality Chasm* report was released.[7]

To be useful, a quality measure should be precisely defined, tied causally to an outcome, and affected by processes that the providers and/or the patients control. The accuracy and validity of the data used to calculate a measure's value are primarily determined by the match between the purpose for which the data was entered (whether on paper or in an electronic system) and the meaning ascribed to that data element when generating a report. The meaning of the data

Affiliations of the authors: Palo Alto Medical Foundation (PCT, LQ), Palo Alto, CA; Lumetra (MR, MFA, JG), San Francisco, CA

Funding for the study was provided through Lumetra, under contract with CMS.

The analyses upon which this publication is based were performed under contract number 500-02-CA02, funded by the Centers for Medicare & Medicaid Services, an agency of the U.S. Department of Health and Human Services. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. The authors assume full responsibility for the accuracy and completeness of the ideas presented.

The authors thank Charles Young for his consultation on data extraction from the EHR, Laurel Trujillo and Tomas Moran for sharing the PAMF internal quality data, and Catherine Coleman, Carlos Gonzales, Susan Lasota, and Laura Stewart for conducting the expert review of the medical records.

Correspondence and reprints: Paul C. Tang, MD, Palo Alto Medical Foundation, 795 El Camino Real, Palo Alto, CA 94301; Tel: (650) 853-5775; Fax: (650) 853-6050; E-mail: pctang@pacbell.net.

Received for review: 07/09/06; accepted for publication: 10/11/06.

Discrepancies between administrative data and clinical data are substantial.

Research Paper ■

Comparison of Information Content of Structured and Narrative Text Data Sources on the Example of Medication Intensification

Alexander Turchin, MD, MS, Maria Shubina, DSc, Eugene Breydo, PhD, Merri L. Pendergrass, MD, PhD, Jonathan S. Einbinder, MD, MPH

Abstract **Objective:** To compare information obtained from narrative and structured electronic sources using anti-hypertensive medication intensification as an example clinical issue of interest.

Design: A retrospective cohort study of 5,634 hypertensive patients with diabetes from 2000 to 2005.

Measurements: The authors determined the fraction of medication intensification events documented in both narrative and structured data in the electronic medical record. The authors analyzed the relationship between provider characteristics and concordance between intensifications in narrative and structured data. As there is no gold standard data source for medication information, the authors clinically validated medication intensification information by assessing the relationship between documented medication intensification and the patients' blood pressure in univariate and multivariate models.

Results: Overall, 5,627 (30.9%) of 18,185 medication intensification events were documented in both sources. For a medication intensification event documented in narrative notes the probability of a concordant entry in structured records increased by 11% for each study year ($p < 0.0001$) and decreased by 19% for each decade of provider age ($p = 0.035$). In a multivariate model that adjusted for patient demographics and intraphysician correlations, an increase of one medication intensification per month documented in either narrative or structured data were associated with a 5–8 mm Hg monthly decrease in systolic and 1.5–4 mm Hg decrease in diastolic blood pressure ($p < 0.0001$ for all).

Conclusion: Narrative and structured electronic data sources provide complementary information on anti-hypertensive medication intensification. Clinical validity of information in both sources was demonstrated by correlation with changes in blood pressure.

■ **J Am Med Inform Assoc.** 2009;16:362–370. DOI 10.1197/jamia.M2777.

Introduction

A large fraction of medical data are contained in narrative documents.[1] As electronic medical record (EMR) systems grow more prevalent,[2] narrative information is increasingly being entered in digital format and thus becomes amenable to computational extraction. Since the late 1990's, a large number of tools have been successfully developed for this purpose.[3–9]

Electronic medical record (EMR) systems employ increasingly rich data models that offer a large variety of options for structured data entry.[10] Data available from the EMR systems frequently includes electronic prescribing information problem and allergy lists, structured note templates, inpatient and outpatient orders, and laboratory results, among others. These data sets have great potential for use in clinical research and/or quality of care surveillance.[11,12]

Not surprisingly, the information in the narrative and structured data sources in the EMR frequently overlap. Physicians typically document all facts pertinent to patient care in narrative notes; at the same time, many of these facts are also entered into the structured data fields in the EMR. It is not known how the data from narrative and structured electronic information compare, to what extent they are overlapping or complementary, and which one better represents reality.

Background

Elevated blood pressure is the most common treatable cardiovascular risk factor[13] and is one of the major risk factors for macro- and micro-vascular complications in patients with diabetes.[14–16] Nevertheless, a majority of diabetic patients with hypertension do not have their blood pressure under control.[17,18] The reasons for poor blood pressure control are not completely understood but lack of appropriate intensification of anti-hypertensive medications is thought

Affiliations of the authors: Clinical Informatics Research and Development, Partners HealthCare System (AT, EB, JSE), Wellesley, MA; Division of Endocrinology (AT, MLP), Center for Clinical Investigation (MS), Division of General Internal Medicine (JSE), Brigham and Women's Hospital, Boston, MA; Harvard Medical School (AT, MLP, JSE), Boston, MA.

Supported in part by grants from Diabetes Action Research and Education Foundation and Agency for Healthcare Research and Quality (R18 HS017030). The authors thank Dr. Alla Keselman for her helpful comments.

Correspondence: Alexander Turchin, MD, MS, Clinical Informatics Research and Development, Suite 201, 93 Worcester St, Wellesley MA, 02481; e-mail: <aturchin@partners.org>.

Received for review: 02/28/08; accepted for publication: 02/09/09.

Discrepancies between narrative text data and structured data in EHRs are also substantial.

Unexplainable A.I. (the “Greek Oracle” problem again)

Incomplete or inaccurate training data (garbage in, garbage out)

Systematic biases in training data

Transformability

Over-control and over-reliance

nytimes.com

The New York Times

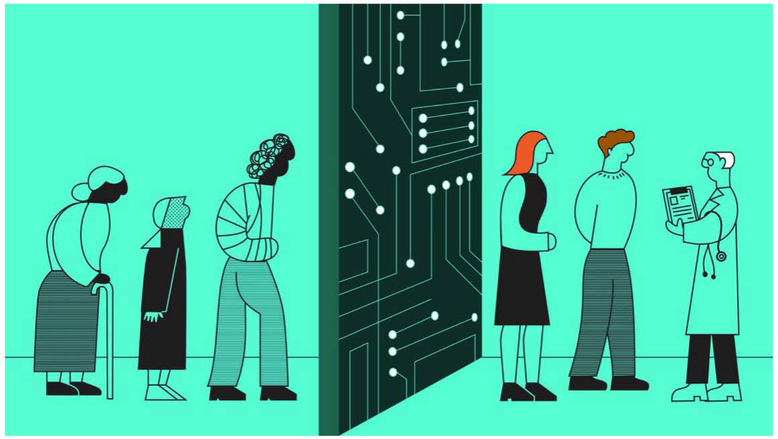
Opinion

A.I. Could Worsen Health Disparities

By Dhruv Khullar

Dr. Khullar is an assistant professor of health care policy and research.

Jan. 31, 2019



Jenice Kim

Artificial intelligence is beginning to meet (and sometimes exceed) assessments by doctors in various clinical situations. A.I. can now

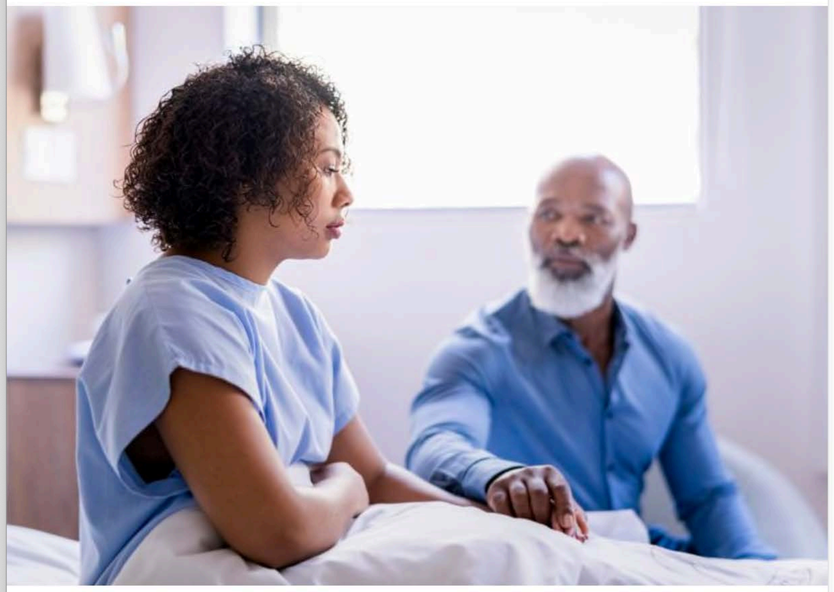
washingtonpost.com

The Washington Post

Democracy Dies in Darkness

Health

Racial bias in a medical algorithm favors white patients over sicker black patients



Scientists discovered racial bias in a widely used medical algorithm that predicts which patients will have complex health needs. (iStock)

By **Carolyn Y. Johnson**

Michigan Medicine - PKUHSC

JOINT INSTITUTE

for Translational and Clinical Research

Unexplainable A.I. (the “Greek Oracle” problem again)

Incomplete or inaccurate training data (garbage in, garbage out)

Systematic biases in training data

Transformability

Over-control and over-reliance

MI: Michigan or Myocardial Infarction

CA: California or Cancer

Unexplainable A.I. (the “Greek Oracle” problem again)

Incomplete or inaccurate training data (garbage in, garbage out)

Systematic biases in training data

Transformability

Over-control and over-reliance

In A.I.
We Trust ?


In A.I.
We
Trust?

Cell

Resource

Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning

Daniel S. Kermany,[1,2,14] Michael Goldbaum,[2,14] Wenjia Cai,[2,14] Carolina C.S. Valentim,[2,14] Huiying Liang,[1,14] Sally L. Baxter,[2,14] Alex McKeown,[3] Ge Yang,[2] Xiaokang Wu,[4] Fangbing Yan,[4] Justin Dong,[1] Made K. Prasadha,[2] Jacqueline Pei,[1,2] Magdalene Y.L. Ting,[2] Jie Zhu,[1,5] Christina Li,[2] Sierra Hewett,[1,2] Jason Dong,[1] Ian Ziyar,[2] Alexander Shi,[2] Runze Zhang,[2] Lianghong Zheng,[6] Rui Hou,[5] William Shi,[2] Xin Fu,[1,2] Yaou Duan,[2] Viet A.N. Huu,[1,2] Cindy Wen,[2] Edward D. Zhang,[1,2] Charlotte L. Zhang,[1,2] Oulan Li,[1,2] Xiaobo Wang,[7] Michael A. Singer,[8] Xiaodong Sun,[9] Jie Xu,[10] Ali Tafreshi,[3] M. Anthony Lewis,[11] Huimin Xia,[1] and Kang Zhang[1,2,4,12,13,15,*]

[1]Guangzhou Women and Children's Medical Center, Guangzhou Medical University, 510005 Guangzhou, China
[2]Shiley Eye Institute, Institute for Engineering in Medicine, Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA
[3]Heidelberg Engineering, Heidelberg, Germany
[4]Molecular Medicine Research Center, State Key Laboratory of Biotherapy, The National Clinical Research Center of Senile Disease, West China Hospital, Sichuan University, Chengdu, China
[5]Guangzhou KangRui Biological Pharmaceutical Technology Company, 510005 Guangzhou, China
[6]YouHealth AI, 510005 Guangzhou, China
[7]Beihai Hospital, Dalian, 116021, China
[8]Department of Ophthalmology, University of Texas Health Science Center, San Antonio, TX 78229, USA
[9]Shanghai Key Laboratory of Ocular Fundus Diseases, Shanghai General Hospital, Shanghai JiaoTong University, 200080 Shanghai, China
[10]Beijing Instute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing, China
[11]Qualcomm, San Diego, CA 92121, USA
[12]Guangzhou Regenerative Medicine and Health Guangdong Laboratory, 510005 Guangzhou, China
[13]Veterans Administration Healthcare System, San Diego, CA 92037, USA
[14]These authors contributed equally
[15]Lead Contact
*Correspondence: kang.zhang@gmail.com
https://doi.org/10.1016/j.cell.2018.02.010