

Should We Use AI to Evaluate Research?

Moshe Y. Vardi

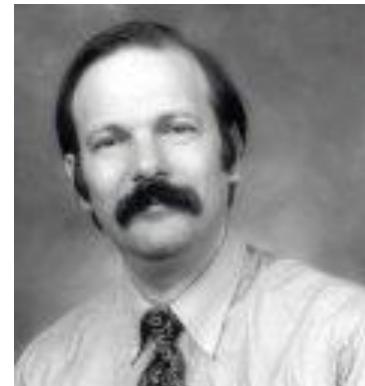
Rice University

vardi@cs.rice.edu

Follow me on Social Media

Who will win the Turing Award?

- ◆ Turing Award: "Nobel Prize in Computing"
- ◆ Moshe Vardi - h=101
- ◆ Gilles Brassard - h=58



The Age of Big Data

- ◆ Inside Higher Ed, 11/2/18: "Connecting Data Science to Almost Every Domain of Inquiry"
 - As reach of big data and AI grows, UC Berkeley and Massachusetts Institute of Technology unveil plans for major expansions.
 - Berkeley plans to form an entirely new division, Division of Data Science and Information, that will engage faculty members and students across the flagship UC campus.
 - MIT unveiled plans to build an entirely new \$1B-college devoted to AI, data science and related fields.

The Five V's of Big Data

- ◆ **Volume:** incredible amounts of data generated each second from social media, cell phones, cars, etc.
- ◆ **Velocity:** speed at which vast amounts of data are being generated, collected and analyzed.
- ◆ **Variety:** many different types of data, structured and unstructured
- ◆ **Veracity:** quality and trustworthiness of data
- ◆ **Value:** turning data into insight

AI vs Doctors

- ◆ Metastatic breast cancer:
 - one of the deadliest cancers, causing an estimated 90 percent of all breast cancer deaths worldwide.
 - small metastatic material can be missed up to 67 percent of the time when examinations happen under extreme time restrictions
- ◆ Google's Lymph Node Assistant algorithm was able to differentiate between metastatic cancer and non-cancer 99 percent of the time!

Restaurants and Big Data

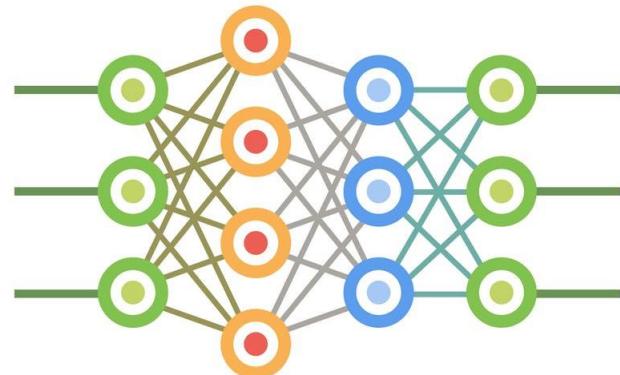
- ◆ Eater, 10/18: Why Restaurants Are So Hungry for Your Personal Data
 - Big chains and small businesses alike are increasingly collecting personal information from their customers.
 - By focusing on customer data collection and analysis and combining those insights with artificial intelligence, TGI Fridays has doubled its to-go business over the past 12 months.
 - Starbucks' wifi is and always has been free, but users must submit their full name, email address, and zip code before they're allowed online — information that the company then uses to send promotional offers.

Your CLV Score

- ◆ CLV= Customer Lifetime Value
 - a numeric representation of how valuable you are as a customer to a given company.
- ◆ Retailers, wireless carriers and others crunch data to determine what shoppers are worth for the long term—and how well to treat them.
- ◆ If you call customer support, your call is received according to your CLV.
- ◆ “There’s no free lunch. The more profitable you are, the better service you will get.”

Neural Nets

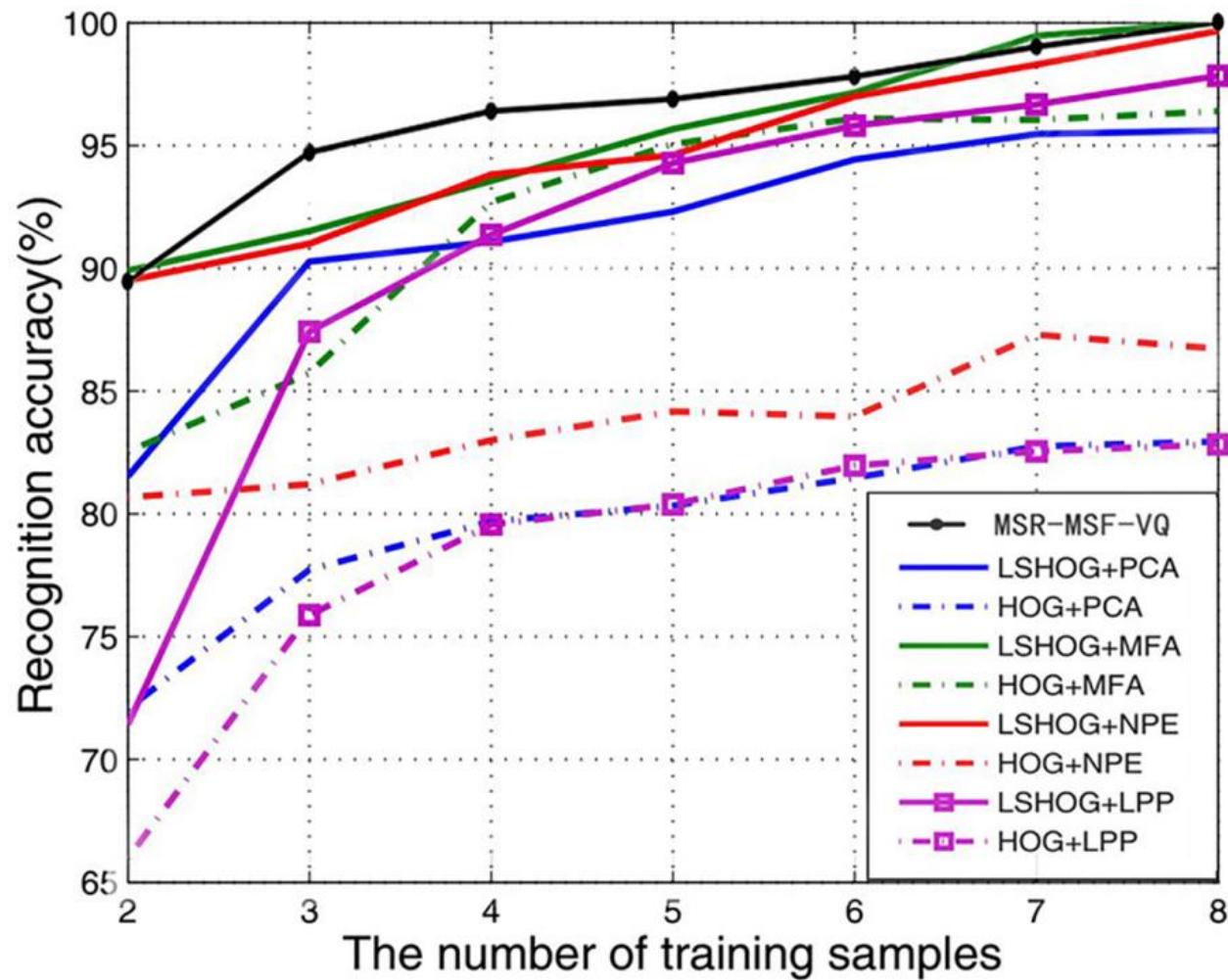
- ◆ Neural Nets: Computational model for pattern recognition based on the visual cortex
 - Key: Strength of connections between nodes



The Deep-Learning Revolution

- ◆ For 70 years artificial neural nets failed to deliver results!
- ◆ Until this decade!
 - Better training algorithms
 - More powerful computers
 - More training data
 - E.g., from 1Ks of images to 1Ms of images

Face Recognition

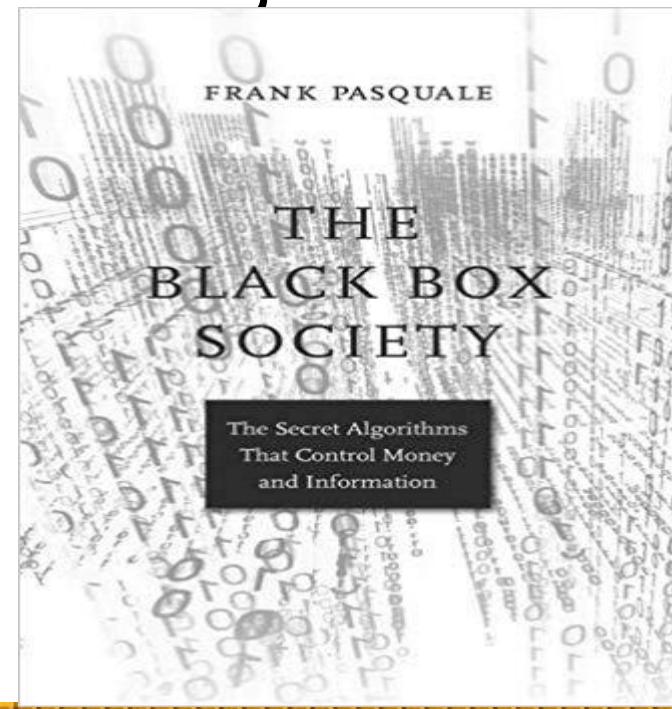


AI in Research Evaluation

- ◆ **Big Question:** How should AI and big-data approaches might impact the area of research policy making?
 - Could AI and related methods offer new insights that transgress the current limitations of available quantitative methods in research evaluation?
 - What might be the benefits and risks of such approaches?

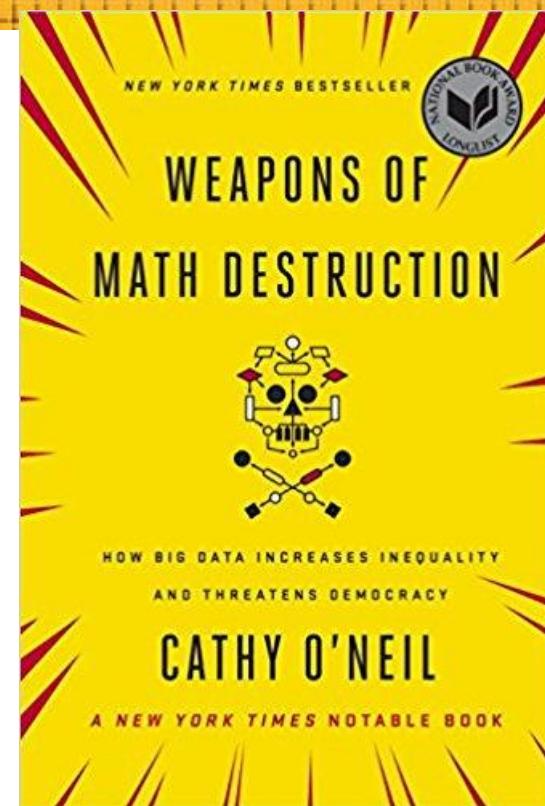
“The Black-Box Society”, ‘15

“Every day, corporations are connecting the dots about our personal behavior—silently scrutinizing clues left behind by our work habits and Internet use. But who connects the dots about what firms are doing with all this information?”



“Weapons of Math Destruction”, ‘16

“The models being used today are opaque, unregulated, and uncontestable, even when they’re wrong. Models are propping up the lucky and punishing the downtrodden. Welcome to the dark side of Big Data.”



“Automating Inequality”, ‘18

- ◆ “Since the dawn of the digital age, decision-making in finance, employment, politics, health and human services has undergone revolutionary change. Today, automated systems—rather than humans—control which neighborhoods get policed, which families attain needed resources, and who is investigated for fraud. While we all live under this new regime of data, the most invasive and punitive systems are aimed at the poor.”

ML and the Justice System, I

Shirley A. Jackson, President, RPI, April 2017:
"Can you foresee a day when smart machines,
driven with artificial intelligences, will assist
with courtroom fact-finding or, more
controversially even, judicial decision-making?"

John G. Roberts, Chief Justice: "It's a day that's
here, and it's putting a significant strain on
how the judiciary goes about doing things."

ML and the Justice System, II

Many applications:

- ◆ Bail
- ◆ Sentencing
- ◆ Parole
- ◆ Separating children from parents
- ◆ ...

Proprietary Algorithms



Advancing Justice.

Embracing Community.



Machine Bias

ProPublica, May 2016: There's software used across the country to predict future criminals. And it's biased against blacks!



Accuracy and Fairness

The accuracy, fairness, and limits of predicting recidivism, Science Advances, Jan. 2018:

"We show that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features."

The Alchemy of AI

Ali Rahimi, Google:

Dec. 2017: "Machine learning algorithms have become a form of 'alchemy.' Researchers do not know why some algorithms work and others don't, nor do they have rigorous criteria for choosing one AI architecture over another."

April 2018: Paper documents examples of the alchemy problem and offers prescriptions for bolstering AI's rigor.

AI Faces Reproducibility Crisis

Matthew Hutson, *Science*, Feb. 2018: "The booming field of artificial intelligence (AI) is grappling with a replication crisis. Just because algorithms are based on code doesn't mean experiments are easily replicated. Far from it. Unpublished codes and a sensitivity to training conditions have made it difficult for AI researchers to reproduce many key results."

“My Dog is a Very Well Trained Dog”

- ◆ My dog Fluffy has been trained to detect risk of recidivism.
- ◆ Its accuracy has even been tested!
- ◆ But it is a “black box” and does not provide explanations



Question: Would you allow Fluffy to make parole decisions?

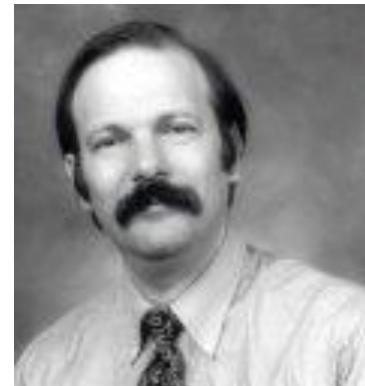
A New Right: *A right for reasonable inference!*

Big Data in Research Evaluation

- ◆ Citeseer, 9/11/2001: "Autonomous citation indexing and literature browsing using citation context"
- ◆ Google Scholar, 11/2004: A freely accessible web search engine that indexes the scholarly literature and its citations.
- ◆ H-index, 2005: author-level metric to measure productivity and citation impact of publications of scholars.
- ◆ Semantic Scholar, 2015: Focuses on "significant" citations.

Who will win the Turing Award?

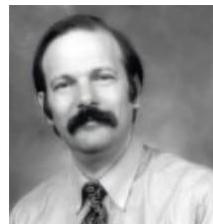
- ◆ Turing Award: "Nobel Prize in Computing"
- ◆ Moshe Vardi - h=101
- ◆ Gilles Brassard - h=58



Vardi's Google Scholar Profile

11/5/2018

Moshe Y. Vardi - Google Scholar Citations



Moshe Y. Vardi

Rice University
Logic and Computation

EIGENES PROFIL ERSTELLEN

Alle Seit 2013

| | | |
|-----------|-------|-------|
| Zitate | 45367 | 13886 |
| h-index | 101 | 56 |
| i10-index | 373 | 260 |

| TITEL | ZITIERT VON | JAHR |
|--|-------------|------|
| Reasoning about knowledge R Fagin, JY Halpern, Y Moses, M Vardi MIT press | 4829 | 2004 |
| An automata-theoretic approach to automatic program verification MY Vardi, P Wolper Proceedings of the First Symposium on Logic in Computer Science, 322-331 | 1963 | 1986 |
| The complexity of relational query languages MY Vardi Proceedings of the fourteenth annual ACM symposium on Theory of computing ... | 1647 | 1982 |
| Simple on-the-fly automatic verification of linear temporal logic R Gerth, D Peled, MY Vardi, P Wolper Protocol Specification, Testing and Verification XV, 3-18 | 994 | 1995 |
| Reasoning about infinite computations MY Vardi, P Wolper Information & Computation 115 (1), 1-37 | 987 | 1994 |
| Reasoning about infinite computations MY Vardi, P Wolper Information & Computation 115 (1), 1-37 | 971 | 1994 |

Brassard's Google Scholar Profile

11/5/2018

Gilles Brassard - Google Scholar Citations



Gilles Brassard

Professor of computer science,
Université de Montréal
quantum information
cryptography
foundations of physics

EIGENES PROFIL ERSTELLEN

Alle Seit 2013

| | | |
|-----------|-------|-------|
| Zitate | 45442 | 15079 |
| h-index | 58 | 37 |
| i10-index | 111 | 68 |

| TITEL | ZITIERT VON | JAHR |
|---|-------------|------|
| Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels CH Bennett, G Brassard, C Crépeau, R Jozsa, A Peres, WK Wootters Physical review letters 70 (13), 1895 | 12596 | 1993 |
| Quantum cryptography: Public key distribution and coin tossing CH Bennett, G Brassard 1984 International Conference on Computers, Systems & Signal Processing ... | 7693 * | 1984 |
| Purification of noisy entanglement and faithful teleportation via noisy channels CH Bennett, G Brassard, S Popescu, B Schumacher, JA Smolin, ... Physical review letters 76 (5), 722 | 2689 | 1996 |
| Experimental quantum cryptography CH Bennett, F Bessette, G Brassard, L Salvail, J Smolin Advances in Cryptology — Proceedings of EUROCRYPT '90, 253-265 | 2240 | 1990 |
| Experimental quantum cryptography CH Bennett, F Bessette, G Brassard, L Salvail, J Smolin | 2185 | 1992 |

Research and Art Evaluation

- ◆ **Claim:** Evaluating research is like evaluating art!
 - It cannot be reduced to numbers.
 - It is a matter of human judgement.
- ◆ **Why?** Because we do not know what to measure!
 - Number of papers?
 - Citations: h-index? g-index? Normalized?
 - Research dollars?
 - Analogous to measuring business by expenses rather than by profit!

The Citation “Game”

Dear ...,

How are you? My name is Zhang, I am from China. I am a businessman, my business is about Citing for Money. Please let me introduce the nature of this business. As you know, the IF(Impact Factor)of a journal is very important, if a journal has a high Impact Factor, the journal will be very significant and famous. So now I invite you to cooperate into this business, the cooperation method is as follows: The price is : 50 USD for each citation, that is, if you cite 1 paper in one of your papers, you will get 50USD, ...

“Free-Style” Research Evaluation

- ◆ A free-style chess player: one who plays the game by marrying human intuition, creativity and empathy with a computer's brute-force ability to remember and calculate a staggering number of chess moves, countermoves and outcomes.
 - Teaming humans and machines produces a force that plays better chess than either humans or computers can manage on their own.
- ◆ Computer-Aided Research evaluation:
Augment human judgement and intuition with the power of big data and AI.

Example: Test-of-Time Awards

- ◆ The LICS Test-of-Time Award recognizes a small number of papers from the LICS proceedings from 20 years prior that have best met the "test of time".
- ◆ Award committee identifies the top 10% of papers with respect to citations, and then makes a choice, based on "intellectual depth, novelty, and impact".

Scholarship of Research

- ◆ **Paradox:** Our science is evidence based, but the way we manage science is not!
 - Needed: *Scholarship of research*
- ◆ **Example:** What is the best way to allocate research funding?
 - Strength of research proposals?
 - Record of research proposer?
 - Lottery (after eliminating weak proposals)?
- ◆ **Bottom line:** Before we measure, we better know what to measure!