# Science & Bull$#@!

## What We Can’t Not Know in the “Age of AI”

Phil Henrickson, PhD
AE Business Solutions

# a disclaimer

# sources and recommended reading

*Calling Bullshit: The Art of Skepticism in a Data -Driven World*
Carl Bergstrom and Jevin D. West

*The Alignment Problem: Machine Learning and Human Values*
Brian Christian

*Science Fictions: Exposing Fraud, Bias, Negligence and Hype in Science*
Stuart Ritchie

*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*
Stuart Ritchie

a brief survey

## Raise your hand if you have…

Created art with a generative AI tool?

Asked a question of ChatGPT?

Solved a problem with ChatGPT?

# Science
# &
# Bull$#@!

What We Can’t Not Know in the “Age of AI”

Advances in computing + increased availability of training data have led to breakthroughs in fields of data science (DS), machine learning (ML), and artificial intelligence (AI).

The “age of AI” is seemingly upon us.

A NEW ERA

# The Age of AI has begun

Artificial intelligence is as revolutionary as mobile phones and the Internet.

By **Bill Gates** | March 21, 2023 • 14 minute read

## 2. AI Technologies in Medical Image Analysis

Different medical imaging modalities have their unique characteristics and different responses to human body structure and organ tissue and can be used in different clinical purposes. The commonly used image modalities for diagnostic analysis in clinic include projection imaging (such as X-ray imaging), computed tomography (CT), ultrasound imaging, and magnetic resonance imaging (MRI). MRI sequences include T1, T1-w, T2, T2-w, diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC), and fluid attenuation inversion recovery (FLAIR). Figure 1 demonstrates a few examples of medical image modalities and their corresponding clinical applications.

**Figure 1** Examples of medical image modalities and their corresponding applications (original copy).

# Medical imaging

## Amazon Rekognition

Automate and lower the cost of your image recognition and video analysis with machine learning

Get Started with Amazon Rekognition

**Thousands of images and videos free per month**

for 12 months with the AWS Free Tier

Quickly add pre-trained or customizable computer vision APIs to your applications without building machine learning (ML) models and infrastructure from scratch.

Analyze millions of images, streaming, and stored videos within seconds, and augment human review tasks with artificial intelligence (AI).

Scale up and down based on your business needs with fully managed AI capabilities and pay only for the images and videos you analyze.

### Features

Amazon Rekognition offers pre-trained and customizable computer vision (CV) capabilities to extract information and insights from your images and videos.

**Face liveness**

Detect real users and deter bad actors using spoofs in seconds during facial verification.

Learn more »

**Face compare and search**

Determine the similarity of a face against another picture or from your private image repository.

Learn more »

**Face detection and analysis**

Detect faces appearing in images and videos and recognize attributes such as open eyes, glasses, and facial hair for each.

Learn more »

## Use Face ID on your iPhone or iPad Pro

Face ID lets you securely unlock your iPhone or iPad, authenticate purchases, sign in to apps, and more, all with just a glance.

The New York Times

## In Hong Kong Protests, Faces Become Weapons

A quest to identify protesters and police officers has people in both groups desperate to protect their anonymity. Some fear a turn toward China-style surveillance.

# Facial recognition

## Use Dictation on your iPhone, iPad, or iPod touch

You can use dictation instead of your keyboard to enter text with many apps and features that use the keyboard on your iPhone, iPad, or iPod touch.

On iPhone 6s or later, and iPad, you can use dictation without being connected to the Internet. Earlier models of iPhone and iPad require an Internet connection.[1]

### Dictate text

To enter text, tap the microphone button on your keyboard, then start speaking. As you speak, the text appears on the screen. To finish, stop speaking, then tap the keyboard button.

## Amazon Transcribe

Automatically convert speech to text

Get Started with Amazon Transcribe

Contact Sales

**60 minutes of speech-to-text for 12 months**
with the AWS Free Tier »

Extract key business insights from customer calls, video files, clinical conversations, and more.

Improve business outcomes with state of the art speech recognition models that are fully managed and continuously trained.

Enhance accuracy with custom models that understand your domain specific vocabulary.

Ensure customer privacy and safety by masking sensitive information.

Azure Explore Products Solutions Pricing Partners Resources

## Speech to text

A Speech service feature that accurately transcribes spoken audio to text.

Try Speech to text free

Create a pay-as-you-go account

Google Cloud Overview Solutions Products Pricing Resources

## Speech-to-Text

Accurately convert speech into text with an API powered by the best of Google's AI research and technology.

New customers get $300 in free credits to spend on Speech-to-Text. All customers get 60 minutes for transcribing and analyzing audio free per month, not charged against your credits.

Go to console

Contact sales

- Transcribe your content with accurate captions
- Enable the power of voice to create better user experiences
- Improve your service with insights from customer interactions
- Get started quickly with our in-console tutorial

# Speech to text

# Self-driving cars

# Stockfish 15.1

Strong open source chess engine

Download Stockfish

**Latest from the blog**

2022-12-04: Stockfish 15.1

2022-11-18: ChessBase GmbH and the Stockfish team reach an agreement and end their legal dispute

2022-06-22: Public court hearing soon!

# AlphaGo

Making history

AlphaGo is the first computer program to defeat a professional human Go player, the first to defeat a Go world champion, and is arguably the strongest Go player in history.

## Chess and Go

More recently, there have been remarkable technological achievements that we may not have even anticipated five years ago.

Creating art with NightCafé

&

OpenArt

*A big sales event for an IT company that loves Wisconsin, beer, and the Green Bay Packers*

*A big sales event for an IT company that loves Wisconsin, beer, and the Green Bay Packers*

*How Minnesota Vikings fans feel when their team loses in the playoffs again*

*How Minnesota Vikings fans feel when their team loses in the playoffs again*

# Generative text with ChatGPT

PH

Write the script for an action movie trailer about a data scientist named Phil trying to teach the world about the pitfalls of machine learning and a technology marketplace that exploits the public's misunderstanding of artificial intelligence.

PH

Write the script for an action movie trailer about a data scientist named Phil trying to teach the world about the pitfalls of machine learning and a technology marketplace that exploits the public's misunderstanding of artificial intelligence.

[Opening scene: A dimly lit lab where Phil, a data scientist, is hunched over his computer, deep in thought. The camera zooms in on the screen, where a complex algorithm is displayed.]

Narrator: In a world where technology has taken over, one data scientist is fighting to educate the masses about the dangers of machine learning.

[Cut to a crowded city street, where people are using their phones and other gadgets.]

Narrator: But in a marketplace where profits reign supreme, will Phil be able to get through to a society that's obsessed with artificial intelligence?

[Phil is shown speaking to a group of people, gesturing emphatically. A few nod, but most seem skeptical.]

Exorbitant claims are being made about the capabilities of AI and the impact it will have on everything from universities to government.

OPINION

THOMAS L. FRIEDMAN

# We Are Opening the Lids on Two Giant Pandora's Boxes

May 2, 2023

Doug Chayka

By Thomas L. Friedman
Opinion Columnist

**Sign up for the Opinion Today newsletter** Get expert analysis of the news and a guide to the big ideas shaping the world every weekday morning. Get it sent to your inbox.

Merriam-Webster notes that a "Pandora's box" can be "anything that looks ordinary but may produce unpredictable harmful results." I've been thinking a lot about Pandora's boxes lately, because we Homo sapiens are doing something we've never done before: lifting the lids on two giant Pandora's boxes at the same time, without any idea of what could come flying out.

One of these Pandora's boxes is labeled "artificial intelligence," and it is exemplified by the likes of ChatGPT, Bard and AlphaFold, which testify to humanity's ability for the first time to manufacture something in a godlike way that approaches general intelligence, far exceeding the brainpower with which we evolved naturally.

The Washington Post
Democracy Dies in Darkness

# Opinion | Type in your job to see how much AI will affect it

By Yan Wu and Sergio Peçanha

761

May 9 at 9:17 a.m.

Is AI coming for your job? If so, when?

Unfortunately, there are no clear-cut answers to these questions. Technology develops in unpredictable ways. But a paper published last month by three scholars — Princeton's Edward W. Felten, Manav Raj of the University of Pennsylvania and Robert Seamans of New York University — offered some helpful insight, at least in terms of artificial intelligence as we now know it.

## How long until machines take over?

Most researchers hesitate to predict the speed in which artificial intelligence will be implemented. NYU's Seamans, who is an associate professor of management and organizations, believes that it will still take a lot of time for the effect to grow large. "Whereas in 2019 [Elon] Musk was touting driverless fleets of Tesla by 2020 (still doesn't exist), Chris Urmson, who knows AVs (autonomous vehicles) very well, was saying 30-50 years," Seamans wrote in an email to us.

# *White House Pushes Tech C.E.O.s to Limit Risks of A.I.*

In the White House's first gathering of A.I. companies, Vice President Kamala Harris told the leaders of major tech companies they had a "moral" obligation to keep products safe.

Give this article 98

NEWS ANALYSIS

# *The Next Fear on A.I.: Hollywood's Killer Robots Become the Military's Tools*

U.S. national security officials are warning about the potential for the new technology to upend war, cyber conflict and — in the most extreme case — the use of nuclear weapons.

Give this article 230

"If we stop, guess who's not going to stop: potential adversaries overseas," the Pentagon's chief information officer, John Sherman, said on Wednesday. Haiyun Jiang/The New York Times

The technology landscape is now full of products branded with AI, offering AI capabilities or AI-driven insights.

“**If AI can do X, AI can do Y**”

# The power of a thousand analysts in your hand.

It’s not easy to know where to look in your data for insights, let alone find the time to build your own reports and dashboards. Using the power of AI, SpotIQ automatically uncovers personalized insights and delivers them to you in seconds.

# Insight Advisor, your intelligent AI assistant

Insight Advisor enhances nearly everything you do in Qlik Sense®. It auto-generates advanced analyses and insights, assists with analytics creation and data prep, and supports natural language interaction including search and conversational analytics.

**Watch Video**

**Insight Advisor Datasheet** ›

**Advanced insight generation** ›

**Search-based visual discovery** ›

**Conversational Analytics** ›

**Accelerated creation and data prep** ›

## Want to know how to get real value from AI-driven analytics?

# SMART, AI-POWERED COMPETITIVE ANALYSIS TOOLS

Analyze the sharpest insights using cutting-edge AI and create a sustainable CI program that's always-on, even when you're not

**Get a Demo**

## Always find what you're looking for—with ease

The press release that Competitor X published back in March? The product retirement that Competitor Y tried to sweep under the rug last week? The social post that Competitor Z promoted throughout Q4? Whatever it is you're looking for, we want you to find it with ease. That's why Crayon analyzes each and every data point and automatically assigns it to the appropriate categories.

# The fastest, most precise no-code AI tool ever.

Turn AI into ROI by going from raw data to industry leading predictive models in minutes, not months.

Model Deployment

## Get your AI models to production in just one click.

Forget spending months on DevOps to get a model in production. Instant web apps allow you to share a dynamic link to use the model in real time.

Model Monitoring

## Robotic Process Automation, now powered by AI.

Automatic model monitoring combined with if-this-than-that automations with Zapier. E.g. If prediction probability is 80%, automatically take this action.

Write

# AI Copywriter

We help you create marketing copy that is more effective faster thanks to our AI. It writes your copy for you. From Facebook ads to headlines, we create text that converts.

- ✓ Helps you write quality copy that converts.
- ✓ Lets you create more variations faster — so it's easier to find the "winning" piece of copy.
- ✓ Facebook Ads, Headlines, Idea Generator and other 40+ Skills

Get Started →

## The Talent Intelligence Platform

Powered by deep-learning AI, our unrivaled Talent Intelligence Platform shows you what you need, when you need it. Whether you're finding or developing talent, our skills-driven approach backed by our unparalleled talent insights delivers the outcomes you need to stay ahead.

AI Resume Writer

# Automate your CV creation with our AI resume builder.

Some people simply don't enjoy writing resumes as much as we do.
That's okay. Our AI Resume Writer is for you.

Try It For Free

## Larry Bear

**Date of birth:** 03/25/1980 **Email address:** hello@kickresume.com **Web:** www.kickresume.com

### RESUME OBJECTIVE

Hard-working and dedicated Mechanical Engineering student with a strong attention to detail and accuracy and experience in completing detailed technical drawings, diagnosing engineering systems, and conducting engineering reports. Offers well-developed analytical skills and excellent teamwork abilities. Daniel is presently seeking a Mechanical Engineering Intern position with a modern firm.

### WORK EXPERIENCE

**Director, Managing Director and Consultant**
**Excel Insurance Brokers Limited**
01/2015 – present KINGSTON, JAMAICA

- Collaborated with colleagues on the design of a new wiring system, produced detailed technical drawings, and supported multiple Senior Mechanical Engineers with various tasks as required.
- Assisted in the development of new processes and procedures to reduce the production costs and increase overall production efficiency.
- Performed detailed design analysis, completed production checklists, and conducted reports on the progress of the assigned projects.
- Diagnosed engineering systems to identify any failures and problems and maintained and co-managed large databases and records.
- Awarded Intern of the Month for finishing all assigned projects in a timely manner.

### EDUCATION

**Mechanical Engineering**
**University of Glasgow**
08/2006 – 12/2008 GLASGOW, UNITED KINGDOM
Graduated with Distinction (Grade 1 - A/excellent equivalent in all 4 subjects)

The problem:

The problem: many of these claims are **bull$#@!**

The problem: many of these claims are **misrepresentations/ misapplications of machine/statistical learning/natural language processing and do not convey the underlying uncertainty inherent in the scientific process**

# Science
# &
## Misrepresentations/misapplications of machine/statistical learning/natural language processing and do not convey the underlying uncertainty inherent in the scientific process

Phil Henrickson, PhD
AE Business Solutions

Much of the discourse on AI at the minute **starts from the premise that “AI” technologies are so good** at solving problems they will replace jobs/disrupt industries/wipe out humanity.

Much of the discourse on AI at the minute **starts from the premise that “AI” technologies are so good** at solving problems they will replace jobs/disrupt industries/wipe out humanity.

Are they?

# The Fallacy of AI Functionality

INIOLUWA DEBORAH RAJI*, University of California, Berkeley, USA
I. ELIZABETH KUMAR*, Brown University, USA
AARON HOROWITZ, American Civil Liberties Union, USA
ANDREW D. SELBST, University of California, Los Angeles, USA

## The Fallacy of AI Functionality

INIOLUWA DEBORAH RAJI*, University of California, Berkeley, USA
I. ELIZABETH KUMAR*, Brown University, USA
AARON HOROWITZ, American Civil Liberties Union, USA
ANDREW D. SELBST, University of California, Los Angeles, USA

"Despite the current public fervor over the great potential of AI, **many deployed algorithmic products do not work**.

**https://arxiv.org/abs/2206.09511**

## The Fallacy of AI Functionality

INIOLUWA DEBORAH RAJI*, University of California, Berkeley, USA
I. ELIZABETH KUMAR*, Brown University, USA
AARON HOROWITZ, American Civil Liberties Union, USA
ANDREW D. SELBST, University of California, Los Angeles, USA

"Despite the current public fervor over the great potential of AI, **many deployed algorithmic products do not work**.

If a product works, we can weigh its costs and benefits. But **if the product does not work**, the judgment is no longer a matter of pros and cons, but a much simpler calculation, exposing that this product does not deserve its spot on the market.

**https://arxiv.org/abs/2206.09511**

## The Fallacy of AI Functionality

INIOLUWA DEBORAH RAJI*, University of California, Berkeley, USA
I. ELIZABETH KUMAR*, Brown University, USA
AARON HOROWITZ, American Civil Liberties Union, USA
ANDREW D. SELBST, University of California, Los Angeles, USA

"Despite the current public fervor over the great potential of AI, **many deployed algorithmic products do not work**.

If a product works, we can weigh its costs and benefits. But **if the product does not work**, the judgment is no longer a matter of pros and cons, but a much simpler calculation, exposing that this product does not deserve its spot on the market.

**The repercussions of AI-related functionality failures in high stakes scenarios cannot be overstated**, and the impact reverberates in real lives for weeks, months and even years."

**https://arxiv.org/abs/2206.09511**

a non-exhaustive list of algorithmic failures

NEWS | COMPUTING

# Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold

›A case study into how to automate false accusations of fraud for more than 34,000 unemployed people

BY ROBERT N. CHARETTE | 24 JAN 2018 | 5 MIN READ |

ILLUSTRATION: ISTOCKPHOTO/IEEE SPECTRUM

IEEE Spectrum FOR THE TECHNOLOGY INSIDER

Type to search

NEWS | COMPUTING

# Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold

›A case study into how to automate false accusations of fraud for more than 34,000 unemployed people

BY ROBERT N. CHARETTE | 24 JAN 2018 | 5 MIN READ

ILLUSTRATION: ISTOCKPHOTO/IEEE SPECTRUM

"A large number of **[unemployment] fraud accusations were being generated algorithmically by MiDAS, with no human intervention or review** of the accusation possible.

MiDAS also **accused some people of fraud even though they had never received any unemployment**. MiDAS was apparently basing some of its findings on missing or corrupt data.

In effect, **MiDAS was built upon the assumption that anyone claiming unemployment insurance was trying to defraud the UIA**…"

The Information

AI

# Toronto Tapped Artificial Intelligence to Warn Swimmers. The Experiment Failed

By Paris Martineau | Nov. 4, 2022 6:00 AM PDT

Photo: Sunnyside Beach in Toronto. High levels of bacteria made the water unsafe for swimming on 50 days this past summer. Photo by John Vetterli/Flickr

Earlier this year, Toronto's public health department quietly flipped the switch on an experiment targeting the city's most pollution-prone beaches.

Instead of relying on day-old laboratory tests to ensure that people don't swim in unsafe water, the city tapped the magic of artificial intelligence, contracting with Cann Forecast, a Montreal-based startup whose predictive modeling products use AI and machine learning to forecast water quality. Beginning in June, officials used Cann's model to decide when two of Toronto's most frequently contaminated beaches should be open to the public and when the water was unsafe.

Almost immediately, the experiment began to go awry. The model regularly declared the beaches safe to swim on days when history or the water's appearance suggested it was not safe. City officials waved off

The Information

AI

# Toronto Tapped Artificial Intelligence to Warn Swimmers. The Experiment Failed

By Paris Martineau | Nov. 4, 2022 6:00 AM PDT

Photo: Sunnyside Beach in Toronto. High levels of bacteria made the water unsafe for swimming on 50 days this past summer. Photo by John Vetterli/Flickr

Earlier this year, Toronto's public health department quietly flipped the switch on an experiment targeting the city's most pollution-prone beaches.

Instead of relying on day-old laboratory tests to ensure that people don't swim in unsafe water, the city tapped the magic of artificial intelligence, contracting with Cann Forecast, a Montreal-based startup whose predictive modeling products use AI and machine learning to forecast water quality. Beginning in June, officials used Cann's model to decide when two of Toronto's most frequently contaminated beaches should be open to the public and when the water was unsafe.

Almost immediately, the experiment began to go awry. The model regularly declared the beaches safe to swim on days when history or the water's appearance suggested it was not safe. City officials waved off

"Instead of relying on day-old laboratory tests to ensure that people don't swim in unsafe water, **the city tapped the magic of artificial intelligence**, contracting with [a start up] whose predictive modeling products use AI and machine learning to forecast water quality.

…**[their] predictive model was less accurate than the city's previous method—and less accurate than a coin flip**.

In all, the model flagged a little more than one of every three days when the water was unsafe, leading to **nearly 50 instances this summer when beaches with dangerous bacteria levels were open to the public** for swimming."

RETAIL OCTOBER 10, 2018 / 6:04 PM / UPDATED 5 YEARS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

"Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."

But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.

RETAIL OCTOBER 10, 2018 / 6:04 PM / UPDATED 5 YEARS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

"Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."

But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.

By 2015, [Amazon] realized its new system was n**ot rating candidates for software developer jobs and other technical posts in a gender neutral way**.

**Amazon's system taught itself that male candidates were preferable**. It penalized resumes that included the word "women's", as in "women's chess club captain." And it downgraded graduates of two all-women's colleges, according to people familiar with the matter.

QUARTZ

# Companies are on the hook if their hiring algorithms are biased

Quartz

October 22, 2018

A U.S. Marine shakes hands at an employers booth at the Hiring Our Heroes job fair

Between 2014 and 2017 Amazon tried to build an algorithmic system to analyze resumes and suggest the best hires. An anonymous Amazon employee called it the “holy grail” if it actually worked.

But it didn’t. After the company trained the algorithm on 10 years of its own hiring data, the algorithm reportedly became biased against female applicants. The word “women,” like in women’s sports, would cause the algorithm to specifically rank applicants lower. After Amazon engineers attempted to fix that problem, the algorithm still wasn’t up to snuff and the project was ended.

yahoo!

Mail News Finance Sports Entertainment Life Shopping Yahoo Plus More...

QUARTZ

# Companies are on the hook if their hiring algorithms are biased

Quartz
October 22, 2018

A U.S. Marine shakes hands at an employers booth at the Hiring Our Heroes job fair

Between 2014 and 2017 Amazon tried to build an algorithmic system to analyze resumes and suggest the best hires. An anonymous Amazon employee called it the "holy grail" if it actually worked.

But it didn't. After the company trained the algorithm on 10 years of its own hiring data, the algorithm reportedly became biased against female applicants. The word "women," like in women's sports, would cause the algorithm to specifically rank applicants lower. After Amazon engineers attempted to fix that problem, the algorithm still wasn't up to snuff and the project was ended.

"Mark J. Girouard, an employment attorney at Nilan Johnson Lewis, says one of his clients was **vetting a company selling a resume screening tool**…

After an audit of the algorithm, the resume screening company found that the algorithm found **two factors to be most indicative of job performance:**

**1) their name was Jared**
**2) they played high school lacrosse**

Girouard's client did not use the tool."

THE WALL STREET JOURNAL.



U.S. | TRANSIT

# MTA's Initial Foray Into Facial Recognition at High Speed Is a Bust

Agency test of technology at RFK bridge fails to identify any faces, internal MTA email says

*By Paul Berger*

"MTA had proposed the use of facial recognition to spot criminals or terrorists on its roads, which accommodate 900,000 cars a day.

[A pilot study using] the facial recognition technology to track drivers on the Robert F. Kennedy Bridge **resulted in 0% of faces being identified "within acceptable parameters,"** according to the Wall Street Journal.

Despite the failed pilot, MTA is moving ahead with the technology."

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.

The Washington Post
Democracy Dies in Darkness

BUSINESS

# Apple Card algorithm sparks gender bias allegations against Goldman Sachs

Entrepreneur David Heinemeier Hansson says his credit limit was 20 times that of his wife, even though she has the higher credit score

By Taylor Telford

November 11, 2019 at 10:44 a.m. EST

engadget

# CNET is reviewing its AI-written articles after being notified of serious errors

The publication recently admitted that it had published around 75 AI-written articles since November.

KATERYNA KON/SCIENCE PHOTO LIBRARY via Getty Images

US MARKETS CLOSED
In the news

Dow Jones -0.03% Nasdaq -0.37% S&P 500 -0.16% META -0.17% TSLA -0.2% BABA +0.05%



# Princeton computer science professor says don't panic over 'bullshit generator' ChatGPT

Sindhu Sundar Jan 30, 2023, 10:54 AM CST

Another conversation with **ChatGPT**

PH

Write a paper summarizing the important role the state of Wisconsin played in the writing of the Declaration of Independence.

PH

Write a paper summarizing the important role the state of Wisconsin played in the writing of the Declaration of Independence.

Title: The Important Role of Wisconsin in the Writing of the Declaration of Independence

Abstract:
This paper explores the often overlooked but significant role played by the state of Wisconsin in the process of writing the Declaration of Independence. While Wisconsin did not exist as a state at the time, its influence on the individuals involved in drafting the Declaration, the ideas it contributed to the document, and its subsequent impact on the American Revolution are essential to understanding the broader context of this pivotal moment in American history.

1. Introduction:
   The Declaration of Independence, adopted on July 4, 1776, is a foundational document that proclaimed the United States' independence from Britain. This paper examines the role of the future state of Wisconsin in shaping the ideas and perspectives of the individuals involved in crafting the Declaration, highlighting its intellectual and geographical contributions.
2. Intellectual Influences:
   Wisconsin, like other colonial territories, was influenced by Enlightenment ideas that championed individual rights, natural law, and self-governance. Prominent figures from Wisconsin, such as Thomas Paine, whose influential pamphlet "Common Sense" argued for independence, had a profound impact on the minds of the Founding Fathers, including Thomas Jefferson, who drafted the Declaration.
3. Geographic Representation:

PH

Which Wisconsin counties start with the letter "M"? Can you list them?

There are two counties in Wisconsin that start with the letter "M." Here they are:

1. Manitowoc County
2. Marathon County

These counties are located in different regions of the state. Manitowoc County is situated along the eastern shore of Lake Michigan in eastern Wisconsin, while Marathon County is located in the central part of the state.

PH

But what about Menominee county? Or Monroe county?

Apologies for the oversight. You are correct. There are indeed three counties in Wisconsin that start with the letter "M." Here is the complete list:

1. Manitowoc County
2. Marathon County
3. Menominee County
4. Monroe County

Manitowoc County is located in eastern Wisconsin along the shores of Lake Michigan. Marathon County is situated in central Wisconsin. Menominee County is located in northeastern Wisconsin and is the state's only county with a majority Native American population. Monroe County is situated in western Wisconsin. Thank you for pointing out the missing counties, and I apologize for any confusion caused.

PH

What are the cognitive benefits of playing board games? Cite your sources.

2 / 2

Sure, here are some sources that support the cognitive benefits of playing board games:

1. Mental Stimulation and Cognitive Benefits:
According to a study published in the Journal of Gerontology, playing board games was associated with a decreased risk of cognitive decline in older adults. The study found that board game playing was linked with enhanced cognitive functioning in domains such as attention, speed, and flexibility, among other benefits. [Source: Park DC, Bischof GN. The aging mind: neuroplasticity in response to cognitive training. Dialogues Clin Neurosci. 2013;15(1):109-119.]
2. Memory Improvement:
A study published in the Journal of Educational Psychology found that playing board games can improve working memory, the ability to store and manipulate information over short periods. The researchers found that board game playing was associated with significant improvement in working memory, specifically among children who played the game several times over a four-week period. [Source: Trammell JP, Schloss PJ, Alpern RM, et al. The effects of board games on the acquisition of spatial knowledge among elementary school students. J Educ Psychol. 2017;109(8):1122-1135.]
3. Cognitive Flexibility and Planning:
A study published in the journal PLOS ONE found that playing complex board games can enhance cognitive flexibility and planning abilities. The study examined the effects of playing the board game Settlers of Catan, and found that playing the game was associated with improved cognitive flexibility, planning, and decision-making. [Source: Herrewijn L, Poels K, T'Jonck K, De Grooff D. Board games as a complex intervention for improving cognitive and executive functions: A systematic review and meta-analysis. PLoS ONE. 2020;15(10):e0240459.]
4. Social Interaction and Communication:
According to research published in the Journal of Adolescent Health, playing board games can promote positive social interaction and communication skills among young adults. The study found that board game playing was associated with increased social bonding and social support, as well as improved communication skills and emotional regulation. [Source: Moberg DJ, DeMatteo D. A Review of Social Skills Training Groups for Youth with Asperger Syndrome and High Functioning Autism. J Autism Dev Disord. 2013;43(11):2273-2283.]

PH

Tell me about Phil Henrickson, a data scientist living in Madison, Wisconsin.

Phil Henrickson is a data scientist who resides in Madison, Wisconsin. He is known for his innovative research and practical applications of data-driven decision-making in various industries.

Phil holds a Bachelor's degree in Computer Science and a Master's degree in Political Science from the University of Wisconsin-Madison. During his graduate studies, he became interested in the use of data science in political science research and applied his skills in data analysis and statistical modeling to study voting patterns and election outcomes.

After completing his Master's degree, Phil joined a startup company in Madison as a data scientist. He worked on developing algorithms and models to optimize business processes and improve decision-making, quickly gaining a reputation for his expertise in data science.

Phil's contributions to the field of data science are not limited to his work in the private sector. He has also been an active member of the academic community, publishing research articles in leading political science journals. His research has focused on a wide range of topics, including election forecasting, campaign finance, and public opinion polling.

In addition to his work as a data scientist and researcher, Phil is also an active member of the Madison community. He volunteers his time and expertise to various non-profit organizations and community initiatives. He is passionate about using data to address social issues and improve the quality of life for all members of the community.

PH

Tell me about Phil Henrickson, a data scientist living in Madison, Wisconsin.

Phil Henrickson is a data scientist who resides in Madison, Wisconsin. He is known for his innovative research and practical applications of data-driven decision-making in various industries.

Phil holds a Bachelor's degree in Computer Science and a Master's degree in Political Science from the University of Wisconsin-Madison. During his graduate studies, he became interested in the use of data science in political science research and applied his skills in data analysis and statistical modeling to study voting patterns and election outcomes.

After completing his Master's degree, Phil joined a startup company in Madison as a data scientist. He worked on developing algorithms and models to optimize business processes and improve decision-making, quickly gaining a reputation for his expertise in data science.

Phil's contributions to the field of data science are not limited to his work in the private sector. He has also been an active member of the academic community, publishing research articles in leading political science journals. His research has focused on a wide range of topics, including election forecasting, campaign finance, and public opinion polling.

In addition to his work as a data scientist and researcher, Phil is also an active member of the Madison community. He volunteers his time and expertise to various non-profit organizations and community initiatives. He is passionate about using data to address social issues and improve the quality of life for all members of the community.

Phil's work has received several awards and recognition for his contributions to the field of data science. In 2020, he was named one of Forbes' "30 Under 30" in the field of science, and in 2021, he was awarded the "Young Data Scientist of the Year" award by the International Association of Data Scientists.

Overall, Phil Henrickson is a highly accomplished data scientist who has made significant contributions to the field of data science through his innovative research and practical applications of data-driven decision-making.

Phil's work has received several awards and recognition for his contributions to the field of data science. In 2020, he was named one of Forbes' "30 Under 30" in the field of science, and in 2021, he was awarded the "Young Data Scientist of the Year" award by the International Association of Data Scientists.

Overall, Phil Henrickson is a highly accomplished data scientist who has made significant contributions to the field of data science through his innovative research and practical applications of data-driven decision-making.

“If AI can be wrong about X, AI can be wrong Y”

To quote a famous philosopher:

To quote a famous philosopher:

“Why is there so much bullshit?”

In 2019, a professor of Computer Science at Princeton gave a talk on the emergence of snake oil in AI.

**How to recognize AI snake oil**

Arvind Narayanan

Associate Professor of Computer Science

@random_walker

PRINCETON UNIVERSITY

CITP
CENTER FOR INFORMATION TECHNOLOGY POLICY

## Why is there so much AI snake oil?

AI is an umbrella term for a set of related technologies

Some of those technologies have made genuine, remarkable, widely-publicized progress

Companies exploit public confusion, slap the “AI” label on whatever they’re selling

https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

# Assessing personality & job suitability from 30-second video

https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

# Incomplete & crude but useful breakdown

**Genuine, rapid progress**

- Shazam, reverse img search
- Face recognition
- Med. diagnosis from scans
- Speech to text
- Deepfakes

**Imperfect but improving**

- Spam detection
- Copyright violation
- Automated essay grading
- Hate speech detection
- Content recommendation

**Fundamentally dubious**

- Predicting recidivism
- Predicting job success
- Predictive policing
- Predicting terrorist risk
- Predicting at-risk kids

Perception

Automating judgment

Predicting social outcomes

https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

CALL
Carl T. Bergstrom & Jevin D. West
ING
Calling Bullshit
The Art of Skepticism
in a Data-Driven
World
BULL
SHIT

**New-school bullshit** uses the language of math and science and statistics to create the impression of rigor and accuracy…

**New-school bullshit** uses the language of math and science and statistics to create the impression of rigor and accuracy…

**New-school bullshit** can be particularly effective because many of us don't feel qualified to challenge information that is presented in quantitative form.

That is exactly what **new-school bullshiters** are counting on.

To fight back, one must learn when and how to question such statements.

**DS/ML can be tremendously valuable** for organizations.

But the technology landscape is **full of snake oil/ bull$#@!** that exploits a lack of understanding of their capabilities.

Almost every organization we speak to is trying to **advance their capabilities in data science**.

Vendors, products, firms
in DS/ML are offering:

Science **Bull$#@!**

| | |
|---|---|
| | |

Vendors, products, firms
in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
| --- | --- | --- |
| Use | | |
| Reject | | |

Vendors, products, firms
in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
|---|---|---|
| Use | Learn.<br>Improve decisions.<br>Automate tasks with human oversight. | |
| Reject | Fail to learn.<br>Miss opportunities to improve.<br>Fall behind competition. | |

Vendors, products, firms
in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
|---|---|---|
| Use | | Waste $ and time on flawed projects. Make (lots of) bad, automated decisions. |
| Reject | | Save $ and time. |

Vendors, products, firms in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
|---|---|---|
| Use | Learn.<br>Improve decisions.<br>Automate tasks with human oversight. | Waste $ and time on flawed projects.<br>Make (lots of) bad, automated decisions. |
| Reject | Fail to learn.<br>Miss opportunities to improve.<br>Fall behind competition. | Save $ and time. |

The ability to distinguish between Science and **Bull$#@!** is more important than ever in data/analytics.

**You do not need to be a data scientist to spot the difference**; we are all capable of this.

Data
Data
Data
Data
Statistical procedures or data science algorithms
Output
Interpretation

Any black box used to generate bullshit has to take in data and spit results out.

Most often, bullshit arises because:

1) **There are biases in the data that get fed into the black box**
2) **There are obvious problems with the results that come out**

You usually don't have to open the analytic black box to spot problems coming out of it.

We do not all need to be experts in statistical/ machine learning.

But **there are some things we can't not know** in the "age of AI".

Data
Data
Data
Data
Statistical procedures or data science algorithms
Output
Interpretation

**Part 1**

**Part 2**

Statistical procedures or data science algorithms

I shall begin at the natural starting point for a cool, hip, talk about “AI” at a data/analytics conference:

I shall begin at the natural starting point for a cool, hip, talk about “AI” at a data/analytics conference:

**Cambridge, England in the 1920s and a lady tasting tea**

# 1) The Lady Tasting Tea and the Garden of Forking Paths

Or, the problem with automated insights

Ronald Fisher, an academic, was having lunch with colleagues when a lady made a curious claim.

Ronald Fisher, an academic, was having lunch with colleagues when a lady made a curious claim.

By sipping a cup of tea, she claimed she could tell which had been poured first: **milk** or **tea**.

Ronald Fisher, an academic, was having lunch with colleagues when a lady made a curious claim.

By sipping a cup of tea, she claimed she could tell which had been poured first: **milk** or **tea**.

Ronald Fisher, an academic, was having lunch with colleagues when a lady made a curious claim.

By sipping a cup of tea, she claimed she could tell which had been poured first: **milk** or **tea**.

**Ronald Fisher, being an academic, decided to test the validity of this claim**; how could we determine whether she's telling the truth?

# The Design of Experiments

By

R. A. Fisher, Sc.D., F.R.S.

Formerly Fellow of Gonville and Caius College, Cambridge
Honorary Member, American Statistical Association
and American Academy of Arts and Sciences
Galton Professor, University of London

Oliver and Boyd
Edinburgh: Tweeddale Court
London: 33 Paternoster Row, E.C.
1935

II

## THE PRINCIPLES OF EXPERIMENTATION, ILLUSTRATED BY A PSYCHO-PHYSICAL EXPERIMENT

### 5. Statement of Experiment

A LADY declares that by tasting a cup of tea made with milk she can discriminate whether the milk or the tea infusion was first added to the cup. We will consider the problem of designing an experiment by means of which this assertion can be tested. For this purpose let us first lay down a simple form of experiment with a view to studying its limitations and its characteristics, both those which appear to be essential to the experimental method, when well developed, and those which are not essential but auxiliary.

Our experiment consists in mixing eight cups of tea, four in one way and four in the other, and presenting them to the subject for judgment in a random order. The subject has been told in advance of what the test will consist, namely that she will be asked to taste eight cups, that these shall be four of each kind, and that they shall be presented to her in a random order, that is in an order not determined arbitrarily by human choice, but by the actual manipulation of the physical apparatus used in games of chance, cards, dice, roulettes, etc., or, more

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

But, by random guessing, she has a 50/50 chance of getting every cup right.

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

But, by random guessing, she has a 50/50 chance of getting every cup right.

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

But, by random guessing, she has a 50/50 chance of getting every cup right.

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

But, by random guessing, she has a 50/50 chance of getting every cup right.

The answer seems obvious: just have her sip a few cups of tea and see if she gets them right.

But, by random guessing, she has a 50/50 chance of getting every cup right.

How many cups does she need to sip?
How many does she need to correctly identify?
What evidence could support her claim?

Fisher settled on eight cups, four milk and four tea.

Fisher settled on eight cups, four milk and four tea.

His idea was simple: **if someone was randomly guessing, how many cups would they typically get correct** out of eight?

How many possible sequences of eight cups (four milk, four tea) are there?

tea tea tea tea milk milk milk milk
milk milk tea milk milk tea tea tea
milk tea tea tea milk milk milk tea
milk milk tea milk tea tea milk tea
milk tea milk milk tea milk tea tea
tea milk milk tea milk tea milk tea
tea milk tea milk tea milk tea milk
milk milk tea tea milk tea milk tea
milk milk milk milk tea tea tea tea
milk milk milk tea tea milk tea tea
tea milk tea tea milk tea milk milk
milk tea milk milk milk tea tea tea
tea tea tea milk milk tea milk milk
tea tea milk tea milk tea milk milk
tea milk tea milk tea tea milk milk
tea milk tea tea tea milk milk milk
tea tea tea milk milk milk tea milk
tea milk milk milk tea tea tea milk
milk tea tea tea milk milk tea milk
tea tea milk tea tea milk milk milk
tea milk tea milk milk tea tea milk
tea milk milk milk milk tea tea tea
tea milk milk tea tea milk milk tea
tea milk milk milk tea tea milk tea
tea tea milk milk milk milk tea tea
tea milk milk tea milk milk tea tea
tea tea milk milk tea milk milk tea
tea milk milk tea tea milk tea milk
tea tea tea milk tea milk milk milk
milk tea milk milk tea tea tea milk
milk milk tea milk tea tea tea milk
tea tea milk milk tea milk tea milk
tea milk milk tea milk tea tea milk
milk tea tea milk tea milk tea milk
milk milk milk tea milk tea tea tea

milk tea milk milk tea tea milk tea
tea tea milk milk tea tea milk milk
milk tea tea tea milk tea milk milk
tea tea milk tea milk milk milk tea
milk tea milk tea tea milk milk tea
tea tea milk milk milk tea milk tea
tea milk milk tea tea tea milk milk
milk milk tea tea tea milk milk tea
milk milk tea tea tea tea milk milk
tea tea milk milk milk tea tea milk
milk tea tea tea tea milk milk milk
tea milk tea tea milk milk tea milk
milk milk milk tea tea tea milk tea
milk tea tea milk milk tea tea milk
milk milk milk tea tea tea tea milk
tea milk tea milk tea milk milk tea
milk tea milk tea milk tea tea milk
milk milk tea milk tea milk tea tea
tea milk tea milk milk tea milk tea
milk tea milk milk tea milk tea tea
milk tea tea milk tea tea milk milk
milk tea tea milk tea tea milk milk
milk milk tea tea milk tea tea milk
milk tea milk tea milk milk tea tea
tea tea tea milk milk milk milk tea
milk tea milk tea tea milk tea milk
tea milk tea tea milk milk milk tea
milk tea milk tea milk tea milk tea
milk tea milk tea tea tea milk milk
tea tea milk tea milk milk tea milk
milk tea tea milk milk milk tea tea
tea milk tea milk milk milk tea tea

There are 70 possible sequences of eight cups with four milk and four tea.

| | |
|---|---|
| tea tea tea tea milk milk milk milk | milk tea milk milk tea tea milk tea |
| milk milk tea milk milk tea tea tea | tea tea milk milk tea tea milk milk |
| milk tea tea tea milk milk milk tea | milk tea tea tea milk tea milk milk |
| milk milk tea milk tea tea milk tea | tea tea milk tea milk milk milk tea |
| milk tea milk milk tea milk tea tea | milk tea milk tea tea milk milk tea |
| tea milk milk tea milk tea milk tea | tea tea milk milk milk tea milk tea |
| tea milk tea milk tea milk tea milk | tea milk milk tea tea tea milk milk |
| milk milk tea tea milk tea milk tea | milk milk tea tea tea milk milk tea |
| milk milk milk milk tea tea tea tea | milk milk tea tea tea tea milk milk |
| milk milk milk tea tea milk tea tea | tea tea milk milk milk tea tea milk |
| tea milk tea tea milk tea milk milk | milk tea tea tea tea milk milk milk |
| milk tea milk milk milk tea tea tea | tea milk tea tea milk milk tea milk |
| tea tea tea milk milk tea milk milk | milk milk milk tea tea tea milk tea |
| tea tea milk tea milk tea milk milk | milk tea tea milk milk tea tea milk |
| tea milk tea milk tea tea milk milk | milk milk milk tea tea tea tea milk |
| tea milk tea tea tea milk milk milk | tea milk tea milk tea milk milk tea |
| tea tea tea milk milk milk tea milk | milk tea milk tea milk tea tea milk |
| tea milk milk milk tea tea tea milk | milk milk tea milk tea milk tea tea |
| milk tea tea tea milk milk tea milk | tea milk tea milk milk tea milk tea |
| tea tea milk tea tea milk milk milk | milk tea tea milk milk tea milk tea |
| tea milk tea milk milk tea tea milk | milk milk tea tea milk tea tea milk |
| tea milk milk milk milk tea tea tea | milk tea tea milk tea tea milk milk |
| tea milk milk tea tea milk milk tea | tea milk milk milk tea milk tea tea |
| tea milk milk milk tea tea milk tea | milk milk tea tea tea milk tea milk |
| tea tea milk milk milk milk tea tea | milk tea tea milk tea milk milk tea |
| tea milk milk tea milk milk tea tea | milk milk tea tea milk milk tea tea |
| tea tea milk milk tea milk milk tea | milk tea milk tea milk milk tea tea |
| tea milk milk tea tea milk tea milk | tea tea tea milk milk milk milk tea |
| tea tea tea milk tea milk milk milk | milk tea milk tea tea milk tea milk |
| milk tea milk milk tea tea tea milk | tea milk tea tea milk milk milk tea |
| milk milk tea milk tea tea tea milk | milk tea milk tea milk tea milk tea |
| tea tea milk milk tea milk tea milk | milk tea milk tea tea tea milk milk |
| tea milk milk tea milk tea tea milk | tea tea milk tea milk milk tea milk |
| milk tea tea milk tea milk tea milk | milk tea tea milk milk milk tea tea |
| milk milk milk tea milk tea tea tea | tea milk tea milk milk milk tea tea |

**tea tea milk milk milk tea milk tea** — the actual sequence

There are 70 possible sequences of eight cups with four milk and four tea.

| | |
|---|---|
| tea tea tea tea milk milk milk milk | milk tea milk milk tea tea milk tea |
| milk milk tea milk milk tea tea tea | tea tea milk milk tea tea milk milk |
| milk tea tea tea milk milk milk tea | milk tea tea tea milk tea milk milk |
| milk milk tea milk tea tea milk tea | tea tea milk tea milk milk milk tea |
| milk tea milk milk tea milk tea tea | milk tea milk tea tea milk milk tea |
| tea milk milk tea milk tea milk tea | tea tea milk milk milk tea milk tea |
| tea milk tea milk tea milk tea milk | tea milk milk tea tea tea milk milk |
| milk milk tea tea milk tea milk tea | milk milk tea tea tea milk milk tea |
| milk milk milk milk tea tea tea tea | milk milk tea tea tea tea milk milk |
| milk milk milk tea tea milk tea tea | tea tea milk milk milk tea tea milk |
| tea milk tea tea milk tea milk milk | milk tea tea tea tea milk milk milk |
| milk tea milk milk milk tea tea tea | tea milk tea tea milk milk tea milk |
| tea tea tea milk milk tea milk milk | milk milk milk tea tea tea milk tea |
| tea tea milk tea milk tea milk milk | milk tea tea milk milk tea tea milk |
| tea milk tea milk tea tea milk milk | milk milk milk tea tea tea tea milk |
| tea milk tea tea tea milk milk milk | tea milk tea milk tea milk milk tea |
| tea tea tea milk milk milk tea milk | milk tea milk tea milk tea tea milk |
| tea milk milk milk tea tea tea milk | milk milk tea milk tea milk tea tea |
| milk tea tea tea milk milk tea milk | tea milk tea milk milk tea milk tea |
| tea tea milk tea tea milk milk milk | milk tea tea milk milk tea milk tea |
| tea milk tea milk milk tea tea milk | milk milk tea tea milk tea tea milk |
| tea milk milk milk milk tea tea tea | milk tea tea milk tea tea milk milk |
| tea milk milk tea tea milk milk tea | tea milk milk milk tea milk tea tea |
| tea milk milk milk tea tea milk tea | milk milk tea tea tea milk tea milk |
| tea tea milk milk milk milk tea tea | milk tea tea milk tea milk milk tea |
| tea milk milk tea milk milk tea tea | milk milk tea tea milk milk tea tea |
| tea tea milk milk tea milk milk tea | milk tea milk tea milk milk tea tea |
| tea milk milk tea tea milk tea milk | tea tea tea milk milk milk milk tea |
| tea tea tea milk tea milk milk milk | milk tea milk tea tea milk tea milk |
| milk tea milk milk tea tea tea milk | tea milk tea tea milk milk milk tea |
| milk milk tea milk tea tea tea milk | milk tea milk tea milk tea milk tea |
| tea tea milk milk tea milk tea milk | milk tea milk tea tea tea milk milk |
| tea milk milk tea milk tea tea milk | tea tea milk tea milk milk tea milk |
| milk tea tea milk tea milk tea milk | milk tea tea milk milk milk tea tea |
| milk milk milk tea milk tea tea tea | tea milk tea milk milk milk tea tea |

The probability of getting eight correct: 1/70, or 0.014

tea tea tea tea milk milk milk milk
milk milk tea milk milk tea tea tea
milk tea tea tea milk milk milk tea
milk milk tea milk tea tea milk tea
milk tea milk milk tea milk tea tea
tea milk milk tea milk tea milk tea
tea milk tea milk tea milk tea milk
milk milk tea tea milk tea milk tea
milk milk milk milk tea tea tea tea
milk milk milk tea tea milk tea tea
tea milk tea tea milk tea milk milk
milk tea milk milk milk tea tea tea
tea tea tea milk milk tea milk milk
tea tea milk tea milk tea milk milk
tea milk tea milk tea tea milk milk
tea milk tea tea tea milk milk milk
tea tea tea milk milk milk tea milk
tea milk milk milk tea tea tea milk
milk tea tea tea milk milk tea milk
tea tea milk tea tea milk milk milk
tea milk tea milk milk tea tea milk
tea milk milk milk milk tea tea tea
tea milk milk tea tea milk milk tea
tea milk milk milk tea tea milk tea
tea tea milk milk milk milk tea tea
tea milk milk tea milk milk tea tea
tea tea milk milk tea milk milk tea
tea milk milk tea tea milk tea milk
tea tea tea milk tea milk milk milk
milk tea milk milk tea tea tea milk
milk milk tea milk tea tea tea milk
tea tea milk milk tea milk tea milk
tea milk milk tea milk tea tea milk
milk tea tea milk tea milk tea milk
milk milk milk tea milk tea tea tea

milk tea milk milk tea tea milk tea
tea tea milk milk tea tea milk milk
milk tea tea tea milk tea milk milk
tea tea milk tea milk milk milk tea
milk tea milk tea tea milk milk tea
tea tea milk milk milk tea milk tea
tea milk milk tea tea tea milk milk
milk milk tea tea tea milk milk tea
milk milk tea tea tea tea milk milk
tea tea milk milk milk tea tea milk
milk tea tea tea tea milk milk milk
tea milk tea tea milk milk tea milk
milk milk milk tea tea tea milk tea
milk tea tea milk milk tea tea milk
milk milk milk tea tea tea tea milk
tea milk tea milk milk milk tea milk
milk tea milk tea milk tea tea milk
milk milk tea milk tea milk tea tea
tea milk tea milk milk tea milk tea
milk tea tea milk milk tea milk tea
milk milk tea tea milk tea tea milk
tea milk milk milk tea milk tea tea
milk milk tea tea tea milk tea milk
milk tea milk tea milk milk tea tea
milk tea milk milk tea tea milk tea
milk tea milk tea milk milk tea tea
milk milk tea tea milk milk tea tea
milk tea milk tea milk milk tea tea
tea tea tea milk milk milk milk tea
milk tea milk tea tea milk tea milk
tea milk tea tea milk milk milk tea
milk tea milk tea milk tea milk tea
milk tea milk tea tea tea milk milk
tea tea milk tea milk milk tea milk
milk tea tea milk milk milk tea tea
tea milk tea milk milk milk tea tea

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

| | |
|---|---|
| tea tea tea tea milk milk milk milk | milk tea milk milk tea tea milk tea |
| milk milk tea milk milk tea tea tea | tea tea milk milk tea tea milk milk |
| milk tea tea tea milk milk milk tea | milk tea tea tea milk tea milk milk |
| milk milk tea milk tea tea milk tea | tea tea milk tea milk milk milk tea |
| milk tea milk milk tea milk tea tea | milk tea milk tea tea milk milk tea |
| tea milk milk tea milk tea milk tea | tea tea milk milk milk tea milk tea |
| tea milk tea milk tea milk tea milk | tea milk milk tea tea tea milk milk |
| milk milk tea tea milk tea milk tea | milk milk tea tea tea milk milk tea |
| milk milk milk milk tea tea tea tea | milk milk tea tea tea tea milk milk |
| milk milk milk tea tea milk tea tea | tea tea milk milk milk tea tea milk |
| tea milk tea tea milk tea milk milk | milk tea tea tea tea milk milk milk |
| milk tea milk milk milk tea tea tea | tea milk tea tea milk milk tea milk |
| tea tea tea milk milk tea milk milk | milk milk milk tea tea tea milk tea |
| tea tea milk tea milk tea milk milk | milk tea tea milk milk tea tea milk |
| tea milk tea milk tea tea milk milk | milk milk milk tea tea tea tea milk |
| tea milk tea tea tea milk milk milk | tea milk tea milk tea milk milk tea |
| tea tea tea milk milk milk tea milk | milk tea milk tea milk tea tea milk |
| tea milk milk milk tea tea tea milk | milk milk tea milk tea milk tea tea |
| milk tea tea tea milk milk tea milk | tea milk tea milk milk tea milk tea |
| tea tea milk tea tea milk milk milk | milk tea tea milk milk tea milk tea |
| tea milk tea milk milk tea tea milk | milk milk tea tea milk tea tea milk |
| tea milk milk milk milk tea tea tea | milk tea tea milk tea tea milk milk |
| tea milk milk tea tea milk milk tea | tea milk milk milk tea milk tea tea |
| tea milk milk milk tea tea milk tea | milk milk tea tea tea milk tea milk |
| tea tea milk milk milk milk tea tea | milk tea tea milk tea milk milk tea |
| tea milk milk tea milk milk tea tea | milk milk tea tea milk milk tea tea |
| tea tea milk milk tea milk milk tea | milk tea milk tea milk milk tea tea |
| tea milk milk tea tea milk tea milk | tea tea tea milk milk milk milk tea |
| tea tea tea milk tea milk milk milk | milk tea milk tea tea milk tea milk |
| milk tea milk milk tea tea tea milk | tea milk tea tea milk milk milk tea |
| milk milk tea milk tea tea tea milk | milk tea milk tea milk tea milk tea |
| tea tea milk milk tea milk tea milk | milk tea milk tea tea tea milk milk |
| tea milk milk tea milk tea tea milk | tea tea milk tea milk milk tea milk |
| milk tea tea milk tea milk tea milk | milk tea tea milk milk milk tea tea |
| milk milk milk tea milk tea tea tea | tea milk tea milk milk milk tea tea |

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

The probability of getting four correct: 36/70, or 0.514

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

The probability of getting four correct: 36/70, or 0.514

Fisher's test was this:

It is improbable (~ 0.01) that someone would randomly go 8/8.

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

The probability of getting four correct: 36/70, or 0.514

Fisher's test was this:

It is improbable (~ 0.01) that someone would randomly go 8/8.

**If she goes 8/8, we have evidence supporting her claim; we conclude that she is capable of telling the difference.**

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

The probability of getting four correct: 36/70, or 0.514

Fisher's test was this:

It is improbable (~ 0.01) that someone would randomly go 8/8.

If she goes 8/8, we have evidence supporting her claim; we conclude that she is capable of telling the difference.

If she does not go 8/8, we do not have sufficient evidence to support her claim.

The probability of getting eight correct: 1/70, or 0.014

The probability of getting six correct: 17/70, or 0.229

The probability of getting four correct: 36/70, or 0.514

Fisher’s ideas form the foundation of hypothesis testing and experimental design that have been in use for the last ~90 years.

Fisher's ideas form the foundation of hypothesis testing and experimental design that have been in use for the last ~90 years.

**"What evidence would we need to support this claim?"**
**"Is the data consistent with what we would expect from random chance?"**

CORONAVIRUS

# The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic

Source: Michael Worobey et al. • The New York Times

These patterns pointed to the market as the origin of the outbreak, Dr. Worobey and his colleagues concluded. The researchers ran tests that showed it was extremely unlikely that such a pattern could be produced merely by chance.

RESEARCH SUMMARY

# Efficacy and Safety of mRNA-1273 SARS-CoV-2 Vaccine

L.R. Baden, et al. DOI: 10.1056/NEJMoa2035389

## CLINICAL PROBLEM

The Covid-19 pandemic continues and expands. Additional data regarding vaccines to prevent symptomatic severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection are needed. The mRNA-1273 vaccine is a lipid-encapsulated mRNA vaccine encoding the prefusion stabilized spike protein of SARS-CoV-2.

## CLINICAL TRIAL

A randomized, double-blind trial to evaluate the efficacy and safety of mRNA-1273.

30,420 participants ≥18 years old were assigned to receive either the vaccine or placebo in two intramuscular injections 28 days apart. Participants were followed for safety and the development of laboratory-confirmed, symptomatic Covid-19 over a median of 2 months after the second dose.

## RESULTS

**Safety:**
Vaccine recipients had higher rates of local reactions (e.g., pain, erythema, swelling) and systemic reactions (e.g., headache, fatigue, myalgia) than placebo recipients. Most reactions were mild to moderate and resolved over 1–3 days.

**Efficacy:**
The incidence of Covid-19 was lower among vaccine recipients than among placebo recipients as early as 14 days after the first dose. Protection in the vaccine group persisted for the period of follow-up.

## LIMITATIONS AND REMAINING QUESTIONS

Further study is required to understand the following:

- Safety and efficacy over a longer period of time, in a larger population, and in pregnant women and children.
- Whether the vaccine protects against asymptomatic infection and transmission to unvaccinated persons.
- How to care for those who miss the second vaccine dose.

**Links:** Full article | NEJM Quick Take | Editorial

| | mRNA-1273 Vaccine N=14,550 | Placebo N=14,598 |
|---|---|---|
| Symptomatic Covid-19 | 11 | 185 |
| Severe Covid-19 | 0 | 30 |

**Vaccine efficacy of 94.1% (95% CI, 89.3–96.8%; P<0.001)**

## CONCLUSIONS

Two doses of a SARS-CoV-2 mRNA-based vaccine were safe and provided 94% efficacy against symptomatic Covid-19 in persons 18 or older.

## STATISTICAL ANALYSIS

For analysis of the primary end point, the trial was designed for the null hypothesis that the efficacy of the mRNA-1273 vaccine is 30% or less. A total of 151 cases of Covid-19 would provide 90% power to detect a 60% reduction in the hazard rate (i.e., 60% vaccine efficacy), with two planned interim analyses at approximately 35% and 70% of the target total number of cases (151) and with a one-sided O'Brien–Fleming boundary for efficacy and an overall one-sided error rate of 0.025. The efficacy of the mRNA-1273 vaccine could be demonstrated at either the interim or the primary analysis, performed when the target total number of cases had been observed. The Lan–DeMets alpha-spending function was used for calculating efficacy boundaries at each analysis. At the first interim analysis on November 15, 2020, vaccine efficacy had been demonstrated in accordance with the prespecified statistical criteria. The vaccine efficacy esti-

Fisher's ideas form the foundation of hypothesis testing and experimental design that have been in use for the last ~90 years.

**"What evidence would we need to support this claim?"**
**"Is the data consistent with what we would expect from random chance?"**

Pretty much every statistical test under the sun operates in this way.

Source: http://allendowney.blogspot.com/2016/06/there-is-still-only-one-test.html

Why the history lesson?

Why the history lesson?

1) Testing hypotheses is critical to how we learn from data.
2) It’s very easy to misuse hypothesis testing.

Why the history lesson?

1) Testing hypotheses is critical to how we learn from data.
2) It’s very easy to misuse hypothesis testing.
3) Fisher makes no mention of actually running the experiment in his textbook. He did run the experiment.

Why the history lesson?

1) Testing hypotheses is critical to how we learn from data.
2) It's very easy to misuse hypothesis testing.
3) Fisher makes no mention of actually running the experiment in his textbook. He did run the experiment.

Muriel Bristol,
the lady tasting tea

Why the history lesson?

1) Testing hypotheses is critical to how we learn from data.
2) It's very easy to misuse hypothesis testing.
3) Fisher makes no mention of actually running the experiment in his textbook. He did run the experiment.
4) **She went 8/8.**

Muriel Bristol,
the lady tasting tea

Why the history lesson?

1) **Testing hypotheses is critical to how we learn from data.**

“What evidence would we need to support this claim?”
“Is the data consistent with what we would expect from random chance?”

The Information

AI

# Toronto Tapped Artificial Intelligence to Warn Swimmers. The Experiment Failed

By Paris Martineau | Nov. 4, 2022 6:00 AM PDT

Photo: Sunnyside Beach in Toronto. High levels of bacteria made the water unsafe for swimming on 50 days this past summer. Photo by John Vetterli/Flickr

Earlier this year, Toronto's public health department quietly flipped the switch on an experiment targeting the city's most pollution-prone beaches.

Instead of relying on day-old laboratory tests to ensure that people don't swim in unsafe water, the city tapped the magic of artificial intelligence, contracting with Cann Forecast, a Montreal-based startup whose predictive modeling products use AI and machine learning to forecast water quality. Beginning in June, officials used Cann's model to decide when two of Toronto's most frequently contaminated beaches should be open to the public and when the water was unsafe.

Almost immediately, the experiment began to go awry. The model regularly declared the beaches safe to swim on days when history or the water's appearance suggested it was not safe. City officials waved off

"What evidence do we have supporting your claims of predictive accuracy?"

"Does your model outperform a coin flip?"

"When would your model outperform our current solution?"

Assessing personality & job suitability from 30-second video

“What’s your evidence?”

“How do these scores compare to a random number generator?”

Vendors, products, firms in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
|---|---|---|
| Use | Learn.<br>Improve decisions.<br>Automate tasks with human oversight. | Waste $ and time on flawed projects.<br>Make (lots of) bad, automated decisions. |
| Reject | Fail to learn.<br>Miss opportunities to improve.<br>Fall behind competition. | Save $ and time. |

Why the history lesson?

1) Testing hypotheses is critical to how we learn from data.
2) **It's very easy to misuse hypothesis testing.**

# The power of a thousand analysts in your hand.

It’s not easy to know where to look in your data for insights, let alone find the time to build your own reports and dashboards. Using the power of AI, SpotIQ automatically uncovers personalized insights and delivers them to you in seconds.

# Insight Advisor, your intelligent AI assistant

Insight Advisor enhances nearly everything you do in Qlik Sense®. It auto-generates advanced analyses and insights, assists with analytics creation and data prep, and supports natural language interaction including search and conversational analytics.

**Watch Video**

**Insight Advisor Datasheet** ›

**Advanced insight generation** ›

**Search-based visual discovery** ›

**Conversational Analytics** ›

**Accelerated creation and data prep** ›

## Want to know how to get real value from AI-driven analytics?

**Does wearing a jersey on game day have any effect on my hockey team?**

VICTORY GREEN
12

I have for years avoided wearing jerseys on the days my team, the **Dallas Stars**, plays its games.

This superstition has its roots in my wildly subjective feeling that my wearing a jersey (in Wisconsin) somehow affects my team's performance (usually in Texas).

I have for years avoided wearing jerseys on the days my team, the **Dallas Stars**, plays its games.

This superstition has its roots in my wildly subjective feeling that my wearing a jersey (in Wisconsin) somehow affects my team's performance (usually in Texas).

For the 2022-2023 season, I decided to examine the data.

Stars Win?

no yes

2022
October

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 30 | 31 | | | | | |

2022
November

| | | | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| 27 | 28 | 29 | 30 | | | |

2022
December

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 25 | 26 | 27 | 28 | 29 | 30 | 31 |

2023
January

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | | | | |

Sun Mon Tue Wed Thu Fri Sat

2023
February

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | | | | |

2023
March

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 26 | 27 | 28 | 29 | 30 | 31 | |

2023
April

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | 1 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| 23 | 24 | 25 | 26 | 27 | 28 | 29 |

Sun Mon Tue Wed Thu Fri Sat

# Dallas Stars 2022-2023 season schedule and results

# Dallas Stars 2022-2023 season schedule and results

Game days that I wore a jersey

How did the **Dallas Stars** do when I was wearing a Stars jersey?

How did the **Dallas Stars** do when I was wearing a Stars jersey?

Before I show you any results, **we all agree that my wearing a jersey has absolutely no effect** on how the **Dallas Stars** play, correct?

How did the **Dallas Stars** do when I was wearing a Stars jersey?

Before I show you any results, **we all agree that my wearing a jersey has absolutely no effect** on how the **Dallas Stars** play, correct?

**We all agree my wearing a jersey has absolutely no effect on how the Dallas Stars play.**

Turns out the Stars actually won more games while I was wearing a jersey!

The Stars **won 64% of games** in which I wore a jersey, **compared to 51% of games** in which I didn't.

But it’s analyzing game results based on **which jersey I wore** where I start to really find insights.

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 0.588 | 0.090 | 6.500 | 0.000 |
| Seguin Home | 0.359 | 0.196 | 1.833 | 0.071 |
| Pavelski Away | 0.166 | 0.176 | 0.942 | 0.349 |
| Benn Home Blackout | 0.183 | 0.210 | 0.872 | 0.386 |
| Robertson Home RR | 0.189 | 0.152 | 1.242 | 0.218 |
| Heiskanen Away RR | -0.588 | 0.256 | -2.296 | 0.024 |
| Team Home | -0.157 | 0.109 | -1.436 | 0.155 |

My **Tyler Seguin jersey** had a positive and statistically significant effect on my team's proportion of wins, compared to wearing no jersey.

My **Miro Heiskanen jersey** had a negative and statistically significant effect on my team's proportion of wins, compared to no jersey.

| term | estimate | std.error | statistic | p.value |
|---|---|---|---|---|
| (Intercept) | 32.647 | 3.407 | 9.582 | 0.000 |
| Seguin Home | -2.846 | 2.473 | -1.151 | 0.254 |
| Pavelski Away | 3.864 | 2.223 | 1.738 | 0.087 |
| Benn Home Blackout | -4.647 | 2.563 | -1.813 | 0.074 |
| Robertson Home RR | 1.143 | 2.012 | 0.568 | 0.572 |
| Heiskanen Away RR | -2.706 | 3.125 | -0.866 | 0.390 |
| Team Home | 1.030 | 1.345 | 0.766 | 0.446 |

My **Pavelski jersey** was associated with an **increase of 3.86 shots on goal** per game [90% CI: 0.2, 7.4]

My **Benn jersey** was associated with a **decrease of 4.64 shots on goal** per game [90% CI: -1.68,-7.5]

My **Seguin, Robertson, and Pavelski jerseys** reduced the opponent's goal scoring in specific periods.

My **Benn jersey** increased the Stars' goal scoring in first periods.

My Robertson jersey **helped the opponent and hurt the Stars' shot generation in third periods**!

# INSIGHTS

- Never touch my Heiskanen jersey again.
- Wear Benn jersey in the first period to increase our goal scoring.
- Put Seguin or Pavelski jersey on once we have the lead in the first or second period.
- Put back on Benn jersey to shut down all shot generation.
- Never wear Robertson jersey in third period.
- Also we score significantly more goals on Tuesdays so contact league to get all our games played on Tuesdays.

**Does wearing a jersey on game day have any effect on my hockey team?**

**Does wearing a jersey on game day have any effect on my hockey team?**

**No.**

**Does wearing a jersey on game day have any effect on my hockey team?**

**No.**
This is bullshit via **the garden of forking paths**.

The Stars won 64% of games in which I wore a jersey, compared to 51% of games in which I didn't.

No evidence this difference is meaningful - we would randomly find a result as extreme as this ~30% of the time.

My Robertson jersey **both helped the opponent and hurt my team's shot generation in third periods**!

```r
add_formula = function(data, formula) {

        data %>%
                mutate(mod = deparse1(formula)) %>%
                mutate(outcome = sub(" .*", "", mod))

}


fit_poisson = function(formula, data) {

        glm(as.formula(formula),
            family = 'poisson',
            data) %>%
                tidy() %>%
                mutate_if(is.numeric, round, 3) %>%
                mutate(method = 'poisson') %>%
                add_formula(formula)

}

fit_lm = function(formula, data) {

        lm(as.formula(formula),
           data) %>%
                tidy(conf.int = T, conf.level = .8) %>%
                mutate_if(is.numeric, round, 3) %>%
                mutate(method = 'lm') %>%
                add_formula(formula)

}
```

| term | estimate | std.error |
|---|---|---|
| (Intercept) | 1.644 | 0.271 |
| Seguin Home | -0.093 | 0.238 |
| Pavelski Away | -0.121 | 0.219 |
| Benn Home Blackout | -0.003 | 0.247 |
| Robertson Home RR | 0.189 | 0.178 |
| Heiskanen Away RR | -0.269 | 0.326 |
| Team Home | -0.052 | 0.126 |
| Mon | -0.415 | 0.311 |
| Tue | -0.639 | 0.315 |
| Wed | -0.395 | 0.322 |
| Thu | -0.257 | 0.306 |
| Fri | -0.405 | 0.342 |
| Sat | -0.347 | 0.297 |

I ran hundreds of tests.
Of course there were a few times I found things that that would occur randomly 5-10% of the time.

# The Garden of Forking Paths

## by Jorge Luis Borges

# The Garden of Forking Paths

# The Garden of Forking Paths

# The Garden of Forking Paths

*Open access, freely available online*

Essay

# Why Most Published Research Findings Are False

**John P. A. Ioannidis**

## Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

## Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

**It can be proven that most claimed research findings are false.**

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful.

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1-\beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, $\alpha$. Assuming that $c$ relationships are being probed in the field, the expected values of the $2 \times 2$ table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2

**Data** → **Insight!**

It is very easy to, without even realizing it, select only the evidence that supports a story.

**Data** → **Insight!**

It is very easy to, without even realizing it, select only the evidence that supports a story.

Science demands transparency.

Science is not about proving yourself right; it is about doing everything you can to prove yourself wrong.

**Data** → **Insight!**

It is very easy to, without even realizing it, select only the evidence that supports a story.

Science demands transparency.

Science is not about proving yourself right; it is about doing everything you can to prove yourself wrong.

**“What could prove you wrong?”**
**“What else could explain these results?”**

Data
Data
Data
Data
Statistical procedures or data science algorithms
Output
Interpretation

**Part 1**

**Part 2**

Statistical procedures or data science algorithms

Let’s now jump forward about fifty years, to the early days of machine learning and a **man sitting in his office surrounded by pictures of aircraft radar profiles.**

# 2) Unidentified Aircraft and Faces

Or, when should we trust a model?

Leo Breiman, an academic turned consultant, was trying to help the government identify aircraft based on their radar profiles.

Incident
Wave front
radar
scatterers
Range window
Intensity
Range Profile
Range

Fig. 1. Range profile of aircraft.

Leo Breiman, an academic turned consultant, was trying to help the government identify aircraft based on their radar profiles.

**This is the appropriate setting for training a predictive model:** can we identify an aircraft based on its radar range profile?

An-26
Normalized Amplitude
0.8
0.6
0.4
0.2
0
50
100
150
200
250
Range Cell

Cessna Citation S/II
Normalized Amplitude
0.8
0.6
0.4
0.2
0
50
100
150
200
250
Range Cell

TU Delft
PH-LAB

Yark-42
Normalized Amplitude
0.8
0.6
0.4
0.2
0
50
100
150
200
250
Range Cell

Training Data
Learning Algorithm
Model
Test Data
Model
Output

Known Aircraft
Learning Algorithm
Model
Unknown Aircraft
Model
Predicted Aircraft

Features

Label

Training Data

Features

Label

Test Data

?

?

Known Aircraft
Learning Algorithm
Model
Unknown Aircraft
Model
Predicted Aircraft

What happens here?
How does a model learn?
What is a model?

Breiman’s idea:

Print out pictures of radar profiles and put them on the walls of his office.

Breiman’s idea:

Print out pictures of radar profiles and put them on the walls of his office.

Look at the pictures.

Breiman’s idea:

Print out pictures of radar profiles and put them on the walls of his office.

**Look at the pictures. Think.**

Breiman’s idea:

Print out pictures of radar profiles and put them on the walls of his office.

**Look at the pictures. Think.**

In his own words:

“The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship’s bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

“The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship’s bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

Here’s what we do: manufacture a whole list of yes/no questions:

“The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship’s bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

Here’s what we do: manufacture a whole list of yes/no questions:

- Do you have a bump in the interval [0, 0.33]?

“The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship’s bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

Here’s what we do: manufacture a whole list of yes/no questions:

- Do you have a bump in the interval [0, 0.33]?
- Do you have a bump in the interval [.712, 0.785]?
- And so on on

"The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship's bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

Here's what we do: manufacture a whole list of yes/no questions:

- Do you have a bump in the interval [0, 0.33]?
- Do you have a bump in the interval [.712, 0.785]?
- …

"The data was unorthodox… non-constant dimensionality. One thing I did realize: the ship's bumps persisted at the same length proportional to the length of the ship.

I thought about it for many weeks, and **then finally a bolt from the blue hit me**.

Here's what we do: manufacture a whole list of yes/no questions:

- Do you have a bump in the interval [0, 0.33]?
- Do you have a bump in the interval [.712, 0.785]?
- …

We had many hundreds, thousands of questions with questions like this. We [take] all of these profiles and run them down this list of questions…

This was the first **[decision] tree** I ever grew."

https://www.youtube.com/watch?v=JIXeMj1zwVU

Known Aircraft
Learning Algorithm
Do you have a bump in the interval [0, 0.33]?
Yes
No
Do you have a bump in the interval [0.712, 0.785]?
Do you have a bump in the interval [0.31, 0.56]?
Yes
No
Yes
No
...
...
...
Do you have a bump in the interval [0.95 0.99]?
Yes
No
...
...
Unknown Aircraft
Do you have a bump in the interval [0, 0.33]?
Yes
No
Do you have a bump in the interval [0.712, 0.785]?
Do you have a bump in the interval [0.31, 0.56]?
Yes
No
Yes
No
...
...
...
Do you have a bump in the interval [0.95 0.99]?
Yes
No
...
...
Predict Aircraft

Breiman’s work led to the development of **decision trees**, a highly flexible learning algorithm that is foundational in machine learning.

Why the history lesson?

1) **Models learn patterns from data.**

# Model Trained to Predict **Board Game Ratings** on BoardGameGeek

# Model Trained to Predict **Board Games Phil Has in His Collection**

$$\log\left[\frac{P(\widehat{\text{PhilCollection}} = \text{yes})}{1 - P(\widehat{\text{PhilCollection}} = \text{yes})}\right] = -7.06 + 0.06(\text{MechActionRetrieval}) + 0.03(\text{MechBias}) +$$
$$0.05(\text{MechCardPlayConflictResolution}) + 0.04(\text{MechClosedDrafting}) + 0.14(\text{MechCommunicationLimits}) +$$
$$0.08(\text{MechConnections}) - 0.01(\text{MechContracts}) + 0.05(\text{MechCriticalHitsAndFailures}) +$$
$$0.02(\text{MechDeckBagAndPoolBuilding}) - 0.02(\text{MechDeduction}) + 0.09(\text{MechEndGameBonuses}) +$$
$$0.15(\text{MechFlicking}) + 0.14(\text{MechGridCoverage}) + 0.18(\text{MechHandManagement}) +$$
$$0.05(\text{MechHiddenRoles}) + 0.06(\text{MechLayering}) + 0.02(\text{MechMapAddition}) +$$
$$0.09(\text{MechMarket}) - 0.01(\text{MechOncePergameAbilities}) + 0.12(\text{MechRace}) +$$
$$0.05(\text{MechRondel}) + 0.08(\text{MechScenarioMissionCampaignGam}) + 0.02(\text{MechSemiCooperativeGame}) +$$
$$0.06(\text{MechSimultaneousActionSelectin}) + 0.12(\text{MechSoloSolitaireGame}) + 0.02(\text{MechTechTreesTechTracks}) +$$
$$0.07(\text{MechTrackMovement}) + 0.05(\text{MechTraitorGame}) + 0(\text{MechTurnOrderClaimAction}) +$$
$$0.01(\text{MechTurnOrderProgressive}) + 0.02(\text{MechTurnOrderStatBased}) + 0.07(\text{MechVariableSetUp}) +$$
$$0.09(\text{MechVictoryPointsAsAResource}) + 0.07(\text{CatCityBuilding}) + 0.24(\text{CatDeduction}) +$$
$$0.11(\text{CatIndustryManufacturing}) + 0.11(\text{CatMurderMystery}) + 0.08(\text{CatNovelBased}) +$$
$$0.1(\text{CatPolitical}) + 0.06(\text{CatSpaceExploration}) + 0.07(\text{ComponentsDgtlHybrdAppWbstRqrd}) +$$
$$0.14(\text{ComponentsGemsCrystals}) + 0.14(\text{ComponentsMapContinentlNtnlScl}) + 0.08(\text{ComponentsWoodenPiecesBoards}) +$$
$$0.09(\text{FamCountryEngland}) + 0.13(\text{FamCountryFrance}) + 0.11(\text{FamCreaturesAliensExtrtrrstrls}) +$$
$$0.09(\text{FamMechanismCampaignGames}) + 0.16(\text{FamMoviesStarWars}) - 0.01(\text{FamPlayersGamesWithSolitairRls}) +$$
$$0.08(\text{FamPlayersOneVersusMany}) + 0.16(\text{PubAlderacEntertainmentGroup}) + 0.25(\text{PubAsmodee}) +$$
$$0.13(\text{PubEagleGryphonGames}) + 0.2(\text{PubFantasyFlightGames}) + 0.09(\text{PubMatagot}) +$$
$$0.01(\text{PubPegasusSpiele}) + 0.15(\text{PubRenegadeGameStudios}) + 0.13(\text{PubSpaceCowboys}) +$$
$$0.09(\text{PubStrongholdGames}) + 0.09(\text{PubZManGames}) + 0.1(\text{DesAlanRMoon}) +$$
$$0.15(\text{DesJacquesZeimet}) + 0.08(\text{DesMichaelKiesling}) + 0.11(\text{DesPhilWalkerHarding}) +$$
$$0.05(\text{DesRichardGarfield}) + 0.08(\text{DesTedAlspach}) + 0.07(\text{DesUweRosenberg}) +$$
$$0.06(\text{DesVlaadaChvátil}) + 0.1(\text{ArtChrisQuilliams}) + 0.11(\text{ArtFioreGmbh}) +$$
$$0.12(\text{ArtFranzVohwinkel}) + 0.07(\text{ArtMichaelaKienle}) + 0.06(\text{ArtPeterWocken}) +$$
$$0.11(\text{FamilyUnmatchedSeries}) + 0.1(\text{WordCount}) - 0.09(\text{YearNs1}) -$$
$$0.24(\text{YearNs5}) + 0.55(\text{NumberMechanicsNs3}) + 0.16(\text{EstAverageWeightNs3})$$

# Model Trained to Predict **Board Games Phil's Brother Has in His Collection**

$$
\begin{aligned}
\log\left[\frac{P(\widehat{\text{JimCollection}} = \text{yes})}{1 - P(\widehat{\text{JimCollection}} = \text{yes})}\right] = {} & -8.02 + 0.01(\text{MechActionRetrieval}) + 0.21(\text{MechAreaImpulse}) + \\
& 0.09(\text{MechCardPlayConflictResolution}) + 0.06(\text{MechClosedDrafting}) + 0.19(\text{MechCommoditySpeculation}) + \\
& 0.19(\text{MechCommunicationLimits}) + 0.07(\text{MechConnections}) + 0.06(\text{MechCriticalHitsAndFailures}) + \\
& 0.09(\text{MechDeckBagAndPoolBuilding}) + 0.05(\text{MechDeckConstruction}) + 0.2(\text{MechDeduction}) + \\
& 0.14(\text{MechEndGameBonuses}) + 0.09(\text{MechGridCoverage}) + 0.48(\text{MechHandManagement}) + \\
& 0(\text{MechMapAddition}) + 0.06(\text{MechOncePergameAbilities}) + 0.23(\text{MechRace}) + \\
& 0.08(\text{MechRondel}) + 0.04(\text{MechSoloSolitaireGame}) + 0.05(\text{MechTurnOrderProgressive}) + \\
& 0.11(\text{MechTurnOrderTimeTrack}) + 0.13(\text{MechVariableSetUp}) + 0.16(\text{MechVoting}) - \\
& 1.41(\text{MechWorkerPlcmntDffrntWrkrTyps}) + 0.18(\text{CatBluffing}) + 0.17(\text{CatCollectibleComponents}) + \\
& 0.09(\text{CatSpaceExploration}) + 0.13(\text{CatSpiesSecretAgents}) + 0.17(\text{FamAnimalsCats}) + \\
& 0.06(\text{ComponentsDiceWithIcons}) + 0.13(\text{ComponentsGemsCrystals}) + 0.07(\text{ComponentsMiniatures}) + \\
& 0.01(\text{ComponentsMultiUseCards}) + 0.15(\text{ComponentsSandTimers}) + 0.1(\text{ComponentsWoodenPiecesBoards}) + \\
& 0.07(\text{FamCreaturesAliensExtrtrrstrls}) + 0.19(\text{FamMoviesStarWars}) + 0.12(\text{FamPlayersGamesWithSolitairRls}) + \\
& 0.1(\text{PubAlderacEntertainmentGroup}) + 0.19(\text{PubAsmodee}) + 0.05(\text{PubCMONGlobalLimited}) + \\
& 0.16(\text{PubFantasyFlightGames}) + 0.13(\text{PubMatagot}) + 0.13(\text{PubSpaceCowboys}) + \\
& 0.08(\text{DesAlanRMoon}) + 0.02(\text{DesChristianTPetersen}) + 0.05(\text{DesCoreyKonieczka}) + \\
& 0.03(\text{DesEricMLang}) + 0.15(\text{DesJacquesZeimet}) + 0.1(\text{DesPhilWalkerHarding}) + \\
& 0.08(\text{DesRichardGarfield}) + 0.07(\text{DesRichardLaunius}) + 0.1(\text{DesSeijiKanai}) + \\
& 0.09(\text{DesUweRosenberg}) + 0.04(\text{DesVlaadaChvátil}) + 0.17(\text{ArtChrisQuilliams}) + \\
& 0.15(\text{ArtChristofTisch}) + 0.03(\text{ArtHenningLudvigsen}) + 0.06(\text{ArtJulienDelval}) + \\
& 0.05(\text{ArtKlemensFranz}) + 0.12(\text{FamilyUnmatchedSeries}) + 0.19(\text{PublishedBefore1900}) + \\
& 0.12(\text{WordCount}) - 0.55(\text{YearNs5}) + 0.49(\text{NumberMechanicsNs3})
\end{aligned}
$$

# Large Language Models

**Source: https://lilianweng.github.io/posts/2019-01-31-lm/**

Why the history lesson?

1) Models learn patterns from data.
2) **The patterns they learn come from the data on which they are trained.**

Machine learning and artificial intelligence live and die by the data they employ.

With good data, you can engineer remarkable effective algorithms for translating one language into another…

But **there's no magical algorithm that can spin flax into gold**.

**You can't compensate for bad data**. If someone tells you otherwise, they are bullshiting.

When it comes to predictive modeling and machine learning, the first place to ask questions is right here.

When it comes to predictive modeling and machine learning, the first place to ask questions is right here.

**“What was the data on which the model was trained?”**
**“What was it trained to predict?”**

RETAIL OCTOBER 10, 2018 / 6:04 PM / UPDATED 5 YEARS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

"Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."

But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.

RETAIL OCTOBER 10, 2018 / 6:04 PM / UPDATED 5 YEARS AGO

# Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.

Automation has been key to Amazon's e-commerce dominance, be it inside warehouses or driving pricing decisions. The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars - much like shoppers rate products on Amazon, some of the people said.

"Everyone wanted this holy grail," one of the people said. "They literally wanted it to be an engine where I'm going to give you 100 resumes, it will spit out the top five, and we'll hire those."

But by 2015, the company realized its new system was not rating candidates for software developer jobs and other technical posts in a gender-neutral way.

## Dominated by men

Top U.S. tech companies have yet to close the gender gap in hiring, a disparity most pronounced among technical staff such as software developers where men far outnumber women. Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.

### GLOBAL HEADCOUNT

### EMPLOYEES IN TECHNICAL ROLES

yahoo!

Mail News Finance Sports Entertainment Life Shopping Yahoo Plus More...

QUARTZ

# Companies are on the hook if their hiring algorithms are biased

Quartz
October 22, 2018

A U.S. Marine shakes hands at an employers booth at the Hiring Our Heroes job fair

Between 2014 and 2017 Amazon tried to build an algorithmic system to analyze resumes and suggest the best hires. An anonymous Amazon employee called it the "holy grail" if it actually worked.

But it didn't. After the company trained the algorithm on 10 years of its own hiring data, the algorithm reportedly became biased against female applicants. The word "women," like in women's sports, would cause the algorithm to specifically rank applicants lower. After Amazon engineers attempted to fix that problem, the algorithm still wasn't up to snuff and the project was ended.

"Mark J. Girouard, an employment attorney at Nilan Johnson Lewis, says one of his clients was **vetting a company selling a resume screening tool**…

After an audit of the algorithm, the resume screening company found that the algorithm found **two factors to be most indicative of job performance:**

**1) their name was Jared**
**2) they played high school lacrosse**

Girouard's client did not use the tool."

“What was the data on which the model was trained?”
“What was it trained to predict?”

“When Joy Buolamwini was a computer science undergrad at Georgia Tech in the early 2010s, she was given an assignment to program a robot to play peekaboo.

The programming part was easy, but there was one issue: **the robot wouldn’t recognize Buolamwini’s face**.

"When Joy Buolamwini was a computer science undergrad at Georgia Tech in the early 2010s, she was given an assignment to program a robot to play peekaboo.

The programming part was easy, but there was one issue: **the robot wouldn't recognize Buolamwini's face**.

Later in her undergraduate studies, she traveled to Hong Kong for an entrepreneurship competition. A local startup was giving a demo of one of its "social robots".

The demo worked on everyone in the tour group… **except Buolamwini**."

Joy Buolamwini, PhD

**The patterns models learn come from the data on which they are trained.**

“One of the more popular public-domain database of pictures of faces… [was] assembled in 2007 from online news articles and image captions…

The Alignment Problem, p.31

“One of the more popular public-domain database of pictures of faces… [was] assembled in 2007 from online news articles and image captions…

In 2014 Michigan State [researchers] determined **it was more than 77% male and more than 83% white**.

In fact, **there are more than twice as many images of George W. Bush in the dataset as there are of all Black women, combined**.”

The Alignment Problem, p.31

Creating art with OpenArt

*A collage of human faces smiling at the camera…*

*A collage of human faces smiling at the camera…*

…turns out to mostly be white dudes

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.

Why the history lesson?

1) Models learn patterns from data.
2) The patterns they learn come from the data on which they are trained.
3) **It's important to figure out what patterns they learned.**

Looking at the output of a model is critical to understanding how it reached its predictions.

Sometimes, it takes awhile to notice it learned a troublesome pattern.

**“What did the model learn?”**
**“How did the model reach this prediction?”**

# Skin Cancer Detection and Tracking using Data Synthesis and Deep Learning


**Yunzhu Li***
Peking University
leo.liyunzhu@pku.edu.cn

**Andre Esteva***
Stanford University
esteva@cs.stanford.edu

**Brett Kuprel**
Stanford University
brkuprel@gmail.com

**Rob Novoa**
Stanford University
ranovoa@gmail.com

**Justin Ko**
Stanford University
jmko@stanford.edu

**Sebastian Thrun**
Stanford University
thrun@stanford.edu

Figure 1: Key factors for skin cancer care include early detection and tracking over time.

# Skin Cancer Detection and Tracking using Data Synthesis and Deep Learning


**Yunzhu Li***
Peking University
leo.liyunzhu@pku.edu.cn

**Andre Esteva***
Stanford University
esteva@cs.stanford.edu

**Brett Kuprel**
Stanford University
brkuprel@gmail.com

**Rob Novoa**
Stanford University
ranovoa@gmail.com

**Justin Ko**
Stanford University
jmko@stanford.edu

**Sebastian Thrun**
Stanford University
thrun@stanford.edu

Figure 1: Key factors for skin cancer care include early detection and tracking over time.

Stanford Medicine / News / Algorithm can identify skin cancer

# Deep-learning algorithm matches dermatologists' ability to identify skin cancer

In the hope of creating better access to medical care, Stanford researchers have trained an algorithm to diagnose skin cancer.

**February 10, 2017** - By Taylor Kubota

A dermatologist uses a dermatoscope, a type of handheld microscope, to look at skin. Computer scientists at Stanford have created an artificially intelligent diagnosis algorithm for skin cancer that matched the performance of board-certified dermatologists.
*Matt Young*

“What did the model learn?”
“What increases the risk of skin cancer?”

inch
1
2
3
4
5
6
cm
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

COMMENTARY

*See related letters on pgs 2275 and 2277*

# Automated Classification of Skin Lesions: From Pixels to Practice

Akhila Narla[1], Brett Kuprel[2], Kavita Sarin[3], Roberto Novoa[4,5] and Justin Ko[3,5]

**The letters "Interpretation of the Outputs of Deep Learning Model trained with Skin Cancer Dataset" and "Automated Dermatological Diagnosis: Hype or Reality?" highlight the opportunities, hurdles, and possible pitfalls with the development of tools that allow for automated skin lesion classification. The potential clinical impact of these advances relies on their scalability, accuracy, and generalizability across a range of diagnostic scenarios.**

*Journal of Investigative Dermatology* (2018) **138,** 2108–2110. doi:10.1016/j.jid.2018.06.175

As researchers and clinicians delve into the medical applications of artificial intelligence (AI) and develop deep learning-based tools, dermatology's visually oriented tasks stand out as ripe for innovation. Both providers and patients have ready access to the tissue of interest, and with their smartphones, they possess the imaging devices needed to collect data at scale. We have seen a number of recent advances, including the work of Han et al. (2018), on the automated skin lesion classification tool, "ModelDerm." The dermatological applications of AI hold both opportunities and pitfalls as we cross from "pixels to practice," deploying these tools across diverse patient populations.

**Contextual learning in lesion**

the body. Without multilesion change detection and classification capability, consumer-facing technology runs the risk of reassuring a hypothetical patient about the lentigo on her arm, while missing the melanoma on her leg. Lesion classification can also benefit from multimodal inputs such as age, gender, race, location on the body, or examples of other lesions on the body.

A one-system model may be capable of answering a number of clinical questions across a breadth of dermatological diseases, beyond the binary classification of benign versus malignant (Esteva et al., 2017), whereas from a logistical and usability perspective, it may be suboptimal to have a different model for each skin type or clinical

between multiple tasks and transfers abilities to new tasks using prior experience with similar tasks would also be the most clinically functional and streamlined solution. Although there may be fewer data for some skin types, we hypothesize that the benefits of training a joint model with sufficient data may outweigh its limitations.

**Nonstandardized and standardized input in AI classification**

Making an artificially intelligent system robust enough to handle the variation inherent in image input also poses a hurdle, yet supports the technology's potential for scalability. Dermatology images are the easiest to capture of all medical images, but also the least standardized. Standardization of images is difficult, even with dermoscopic images, as shown in Figure 1. Variability must be incorporated into training algorithms to create capacity to handle noisy data. This includes multiple camera angles, different orientations, blurry photos, multiple skin backgrounds, pen markings or rulers included in the photo, or variations in lighting. Otherwise, the algorithm will use features of nonstandardized photos to guide decision making. For instance, in our work, we noted that the algorithm appeared more likely to interpret images with rulers as malignant. Why? In our dataset, images with rulers were more likely to be malignant; thus the algorithm inadvertently "learned" that rulers are malignant. These biases in AI models are inherent unless specific

"the algorithm appeared more likely to **interpret images with rulers as malignant**. Why?

In our dataset, images with rulers were more likely to be malignant;

thus **the algorithm inadvertently "learned" that rulers are malignant."**

"the algorithm appeared more likely to **interpret images with rulers as malignant**. Why?

In our dataset, images with rulers were more likely to be malignant;

thus **the algorithm inadvertently "learned" that rulers are malignant.**"

Looking at the output of a model is critical to understanding how it reached its predictions.

Sometimes, **it will be obvious that something is wrong**.

“What did the model learn?”
“How did the model reach this prediction?”

# Automated Inference on Criminality using Face Images

Xiaolin Wu
McMaster University
Shanghai Jiao Tong University
xwu510@gmail.com

Xi Zhang
Shanghai Jiao Tong University
zhangxi_19930818@sjtu.edu.cn

arXiv:1611.04135v2 [cs.CV] 21 Nov 2016

## Abstract

*We study, for the first time, automated inference on criminality based solely on still face images, which is free of any biases of subjective judgments of human observers. Via supervised machine learning, we build four classifiers (logistic regression, KNN, SVM, CNN) using facial images of 1856 real persons controlled for race, gender, age and facial expressions, nearly half of whom were convicted criminals, for discriminating between criminals and non-criminals. All four classifiers perform consistently well and empirically establish the validity of automated face-induced inference on criminality, despite the historical controversy surrounding this line of enquiry. Also, some discriminating structural features for predicting criminality have been found by machine learning. Above all, the most important discovery of this research is that criminal and non-criminal face images populate two quite distinctive manifolds. The variation among criminal faces is significantly greater than that of the non-criminal faces. The two manifolds consisting of criminal and non-criminal faces appear to be concentric, with the non-criminal manifold lying in the kernel with a smaller span, exhibiting a law of "normality" for faces of non-criminals. In other words, the faces of general law-biding public have a greater degree of resemblance compared with the faces of criminals, or criminals have a higher degree of dissimilarity in facial appearance than non-criminals.*

management science, criminology, etc.

In all cultures and all periods of recorded human history, people share the belief that the face alone suffices to reveal innate traits of a person. Aristotle in his famous work Prior Analytics asserted, "It is possible to infer character from features, if it is granted that the body and the soul are changed together by the natural affections". Psychologists have known, for as long as a millennium, the human tendency of inferring innate traits and social attributes (e.g., the trustworthiness, dominance) of a person from his/her facial appearance, and a robust consensus of individuals' inferences . These are the facts found through numerous studies [3, 39, 5, 6, 10, 26, 27, 34, 32].

Independent of the validity of pedestrian belief in the (pseudo)science of physiognomy, a tantalizing question naturally arises: what facial features influence average Joes' impulsive and yet consensual judgments on social attributes of a non-acquaintance member of their own specie? Attempting to answer the question, Todorov and Oosterhof proposed a data-driven statistical modeling method to find visual determinants of social attributes by asking human subjects to score four percepts: dominance, attractiveness, trustworthiness, and extroversion, based on first impression of static face images [33]. This method can synthesize a representative (average) face image for a set of input face images scored closely on any of the four aforementioned social percepts. The ranking of these synthesized face images by subjective scores (e.g., from least to most trustworthy looking) apparently agrees with the intuition of most

Known
"Criminals"
Learning
Algorithm
Model
Unknown
"Criminals"
Model
Predicted
"Criminals"

(a) Three samples in criminal ID photo set $S_c$.

(b) Three samples in non-criminal ID photo set $S_n$

"**Unlike a human examiner/judge, a computer vision algorithm or classifier has absolutely no subjective baggages [sic], having no emotions, no biases whatsoever** due to past experience, race, religion, political doctrine, gender, age, etc., no mental fatigue, no preconditioning of a bad sleep or meal.

The **automated inference on criminality eliminates the variable of meta-accuracy (the competence of the human judge/ examiner) all together**."

https://confilegal.com/wp-content/uploads/2016/11/ESTUDIO-UNIVERSIDAD-DE-JIAO-TONG-SHANGHAI.pdf

Figure 2. Accuracy of all four classifiers in all thirteen cases.

| Classifiers | CNN | SVM | KNN | LR |
|---|---|---|---|---|
| AUC | 0.9540 | 0.9303 | 0.8838 | 0.8666 |

Table 1. The AUC results for the four tested face classifiers on criminality.

**"As expected, the state-of-the-art [convolutional neural networks] (CNN) classifier performs the best, achieving 89.51% accuracy.**

The relatively high accuracy of CNN is also paralleled by all other three classifiers which are only few percentage points behind in the success rate of classification.

These highly consistent results are **evidences for the validity of automated face-induced inference on criminality**, despite the historical controversy surrounding the topic."

https://confilegal.com/wp-content/uploads/2016/11/ESTUDIO-UNIVERSIDAD-DE-JIAO-TONG-SHANGHAI.pdf

“What did the model learn?”
“How did the model reach this prediction?”

(a) Three samples in criminal ID photo set $S_c$.

(b) Three samples in non-criminal ID photo set $S_n$

"**What features of a human face betray its owner's propensity for crimes?**

We try to answer the question in the most mechanical and scientific way allowed by the available tools and data.

The approach is to **let a machine learning method explore the data and reveal the most discriminating facial features that tell apart criminals and non-criminals**."

https://confilegal.com/wp-content/uploads/2016/11/ESTUDIO-UNIVERSIDAD-DE-JIAO-TONG-SHANGHAI.pdf

(a) (b)

Figure 8. (a) FGM results; (b) Three discriminative features $\rho$, $d$ and $\theta$.

"The **angle θ from nose tip to two mouth corners is on average 19.6% smaller for criminals than for non-criminals** and has a larger variance.

Also, **the upper lip curvature ρ is on average 23.4% larger for criminals** than for non- criminals.

On the other hand, the distance d between two eye inner corners for criminals is slightly narrower (5.6%) than for non-criminals."

**https://confilegal.com/wp-content/uploads/2016/11/ESTUDIO-UNIVERSIDAD-DE-JIAO-TONG-SHANGHAI.pdf**

The algorithm finds that criminals have shorter distances d between the inner corners of the eyes, smaller angles θ between the nose and the corners of the mouth, and higher curvature ρ to the upper lip.

Why would this possibly be?

https://www.callingbullshit.org/case_studies/case_study_criminal_machine_learning.html

The algorithm finds that criminals have shorter distances d between the inner corners of the eyes, smaller angles θ between the nose and the corners of the mouth, and higher curvature ρ to the upper lip.

Why would this possibly be?

There's a glaringly obvious explanation for the nose-mouth angle θ and the lip curvature ρ. As one smiles, the corners of the mouth spread out and the upper lip straightens. Try it yourself in the mirror.

https://www.callingbullshit.org/case_studies/case_study_criminal_machine_learning.html

(a) Three samples in criminal ID photo set $S_c$.

(b) Three samples in non-criminal ID photo set $S_n$

(a) (b) (c) (d)

Figure 10. (a) and (b) are ”average” faces for criminals and non-criminals generated by averaging of eigenface representations ; (c) and (d) are ”average” faces for criminals and non-criminals generated by averaging of landmark points and image warping.

https://confilegal.com/wp-content/uploads/2016/11/ESTUDIO-UNIVERSIDAD-DE-JIAO-TONG-SHANGHAI.pdf

Data
Data
Data
Data
Statistical procedures or data science algorithms
Output

Human Face
Are they smiling?
Criminal
Not a Criminal

A NEW ERA

# The Age of AI has begun

Artificial intelligence is as revolutionary as mobile phones and the Internet.

By **Bill Gates** | March 21, 2023 • 14 minute read

# wrapping up

# Science
# &
# Bull$#@!

What We Can’t Not Know in the “Age of AI”

The ability to distinguish between Science and **Bull$#@!** is more important than ever in data/analytics.

**You do not need to be a data scientist to spot the difference**; we are all capable of this.

Vendors, products, firms in DS/ML are offering:

| Organizations deciding what to do with DS/ML offerings: | Science | Bull$#@! |
|---|---|---|
| Use | Learn.<br>Improve decisions.<br>Automate tasks with human oversight. | Waste $ and time on flawed projects.<br>Make (lots of) bad, automated decisions. |
| Reject | Fail to learn.<br>Miss opportunities to improve.<br>Fall behind competition. | Save $ and time. |

Questions that can help tell the difference:

“What evidence would we need to support this claim?”

“Is the data consistent with random chance?”

“What could prove you wrong?”

“What else could explain these results?”

“What data was used to train the model?

“What was it trained to predict?”

“What did the model learn?”

“How did the model reach this prediction?”

Model Matrix Elements
for arb. QN
HEARN(KA
901-581-

Richard Feynman

**There is an idea that we all hope you have learned in studying science in school** - we never explicitly say what this *is,* but just hope that you catch on by all the examples of scientific investigation.

**There is an idea that we all hope you have learned in studying science in school** - we never explicitly say what this *is,* but just hope that you catch on by all the examples of scientific investigation.

It's a kind of scientific integrity, **a principle of scientific thought that corresponds to a kind of utter honesty** - a kind of leaning over backwards.

**There is an idea that we all hope you have learned in studying science in school** - we never explicitly say what this *is,* but just hope that you catch on by all the examples of scientific investigation.

It's a kind of scientific integrity, **a principle of scientific thought that corresponds to a kind of utter honesty** - a kind of leaning over backwards.

Details that could throw doubt on your interpretation must be given, if you know them. **You must do the best you can - if you know anything at all wrong, or possibly wrong - to explain it**.

**There is an idea that we all hope you have learned in studying science in school** - we never explicitly say what this *is,* but just hope that you catch on by all the examples of scientific investigation.

It's a kind of scientific integrity, **a principle of scientific thought that corresponds to a kind of utter honesty** - a kind of leaning over backwards.

Details that could throw doubt on your interpretation must be given, if you know them. **You must do the best you can - if you know anything at all wrong, or possibly wrong - to explain it**.

**This is our responsibility as scientists**, certainly to other scientists, and I think to laymen.

thanks for listening