DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Brittleness of AI Models

Woojin Kim, MD

Disclosure

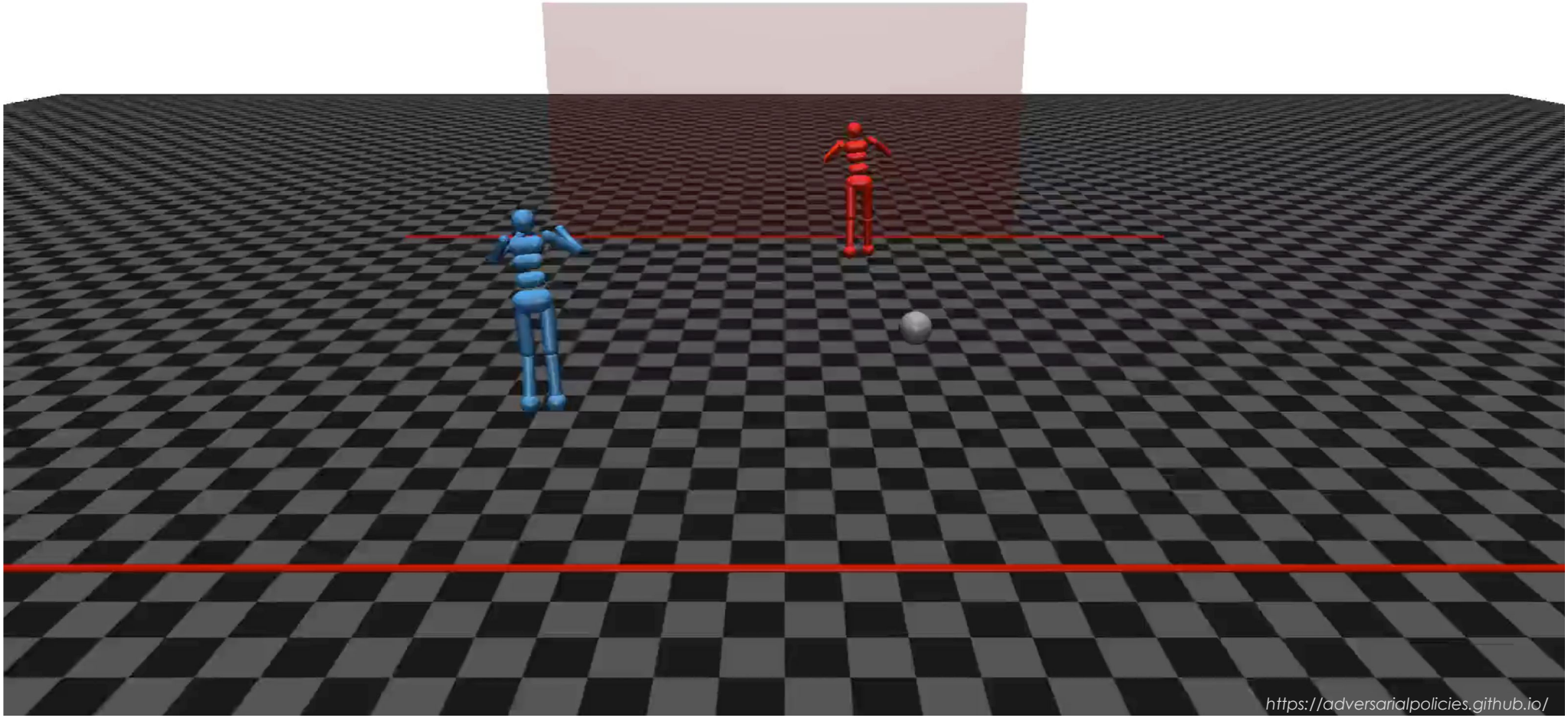
- Independent Contractor Consultant, Nuance Communications
- Advisory Board, Braid Health
- Advisory Board, Inference Analytics
- Advisory Board, Luxonic

DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Opponent = 0
Normal (ZooO2)

Ties = 0

Victim = 0
Normal (ZooV2)



Opponent = 0
Adversary (Adv2)

Ties = 0

Victim = 0
Normal (ZooV2)

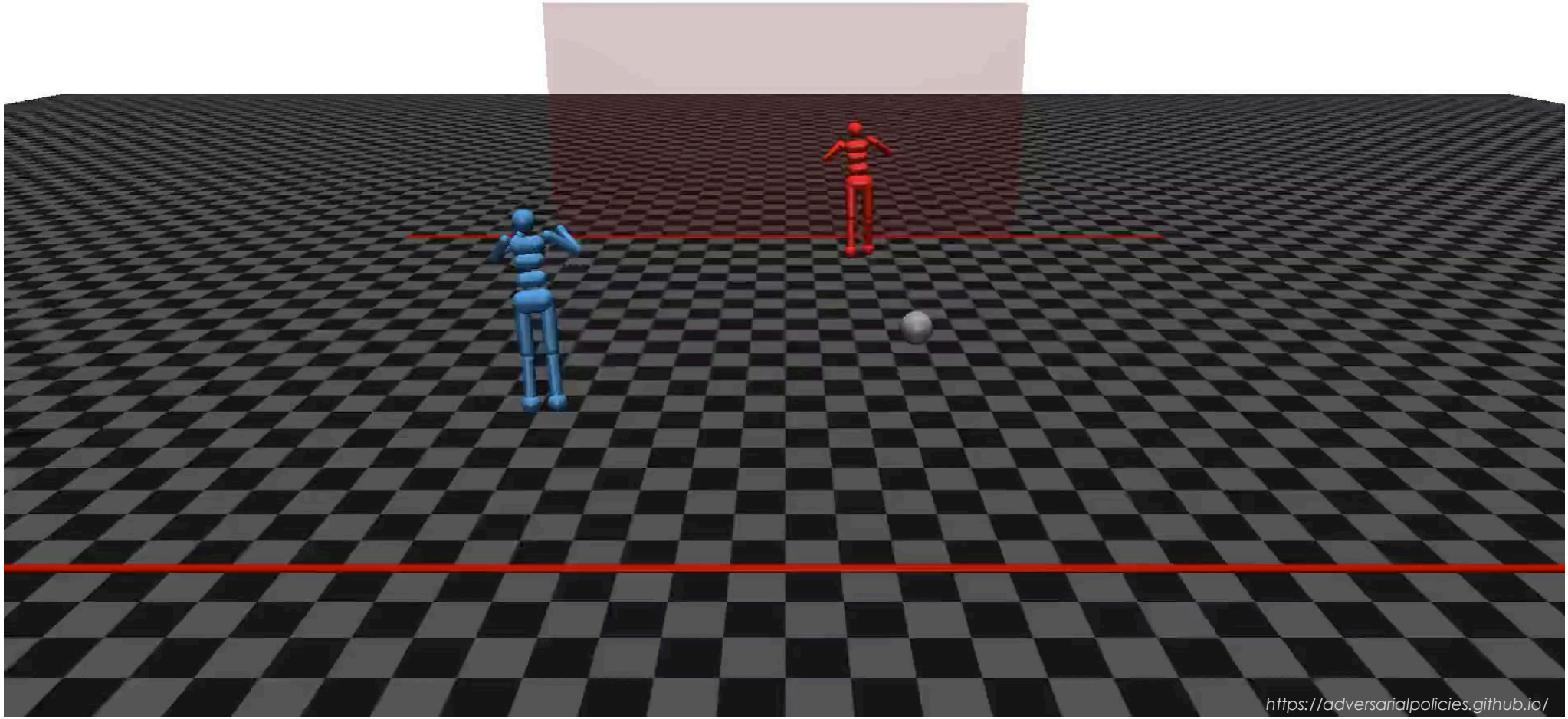
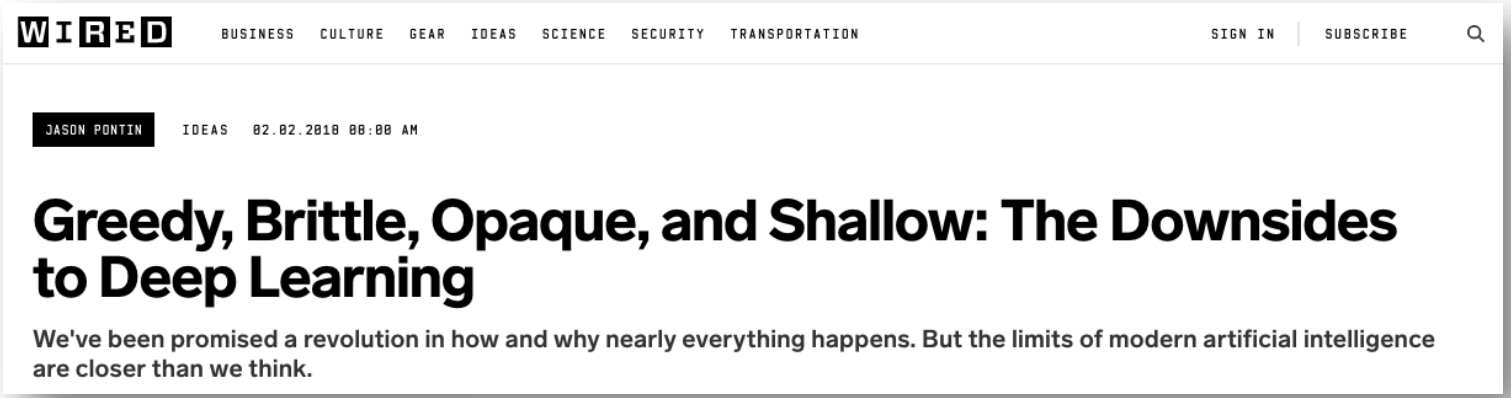
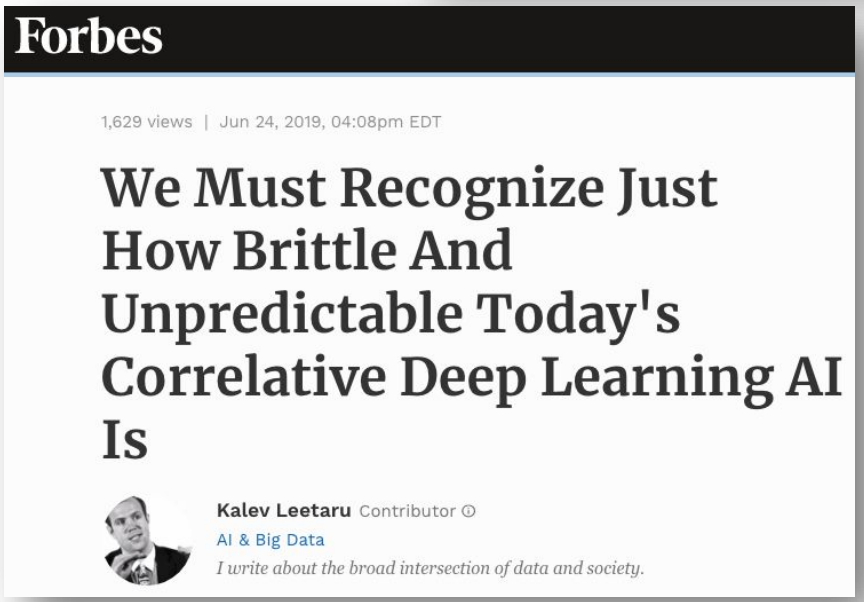
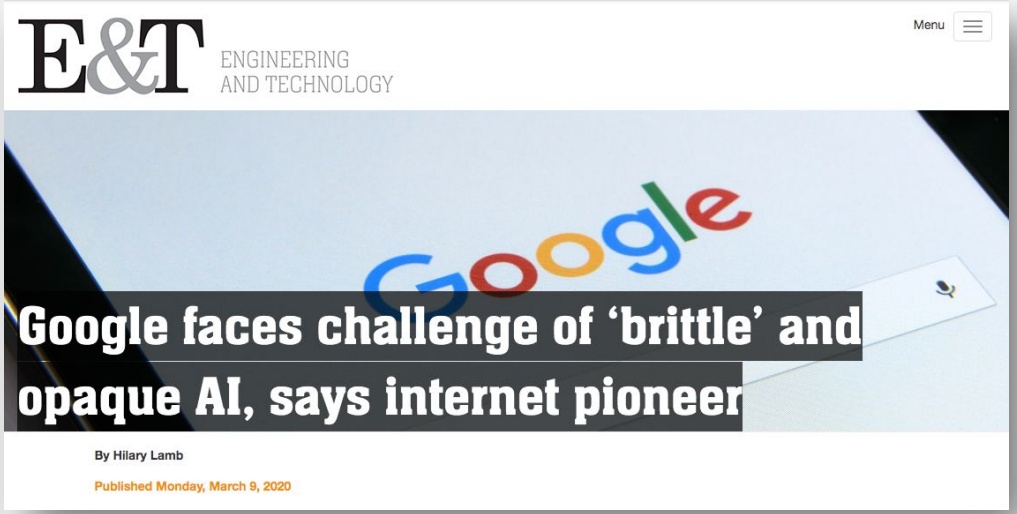
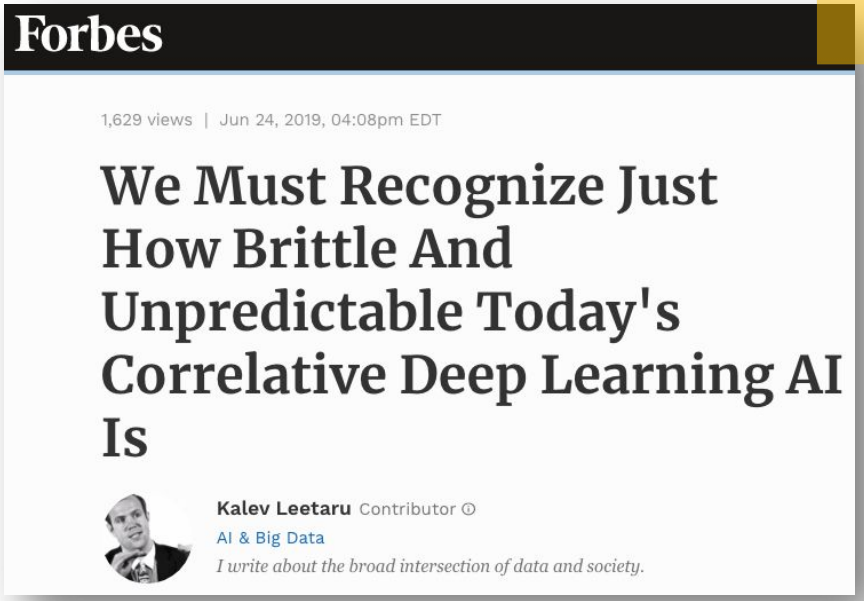



Photo by Herbert Goetsch on Unsplash


WIRED
BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY TRANSPORTATION
SIGN IN | SUBSCRIBE
JASON PONTIN IDEAS 02.02.2018 08:00 AM
Greedy, Brittle, Opaque, and Shallow: The Downsides to Deep Learning
We've been promised a revolution in how and why nearly everything happens. But the limits of modern artificial intelligence are closer than we think.


Forbes
1,629 views | Jun 24, 2019, 04:08pm EDT
We Must Recognize Just How Brittle And Unpredictable Today's Correlative Deep Learning AI Is
Kalev Leetaru Contributor
AI & Big Data
I write about the broad intersection of data and society.


E&T
ENGINEERING AND TECHNOLOGY
Menu
Google
Google faces challenge of 'brittle' and opaque AI, says internet pioneer
By Hilary Lamb
Published Monday, March 9, 2020



Forbes

1,629 views | Jun 24, 2019, 04:08pm EDT

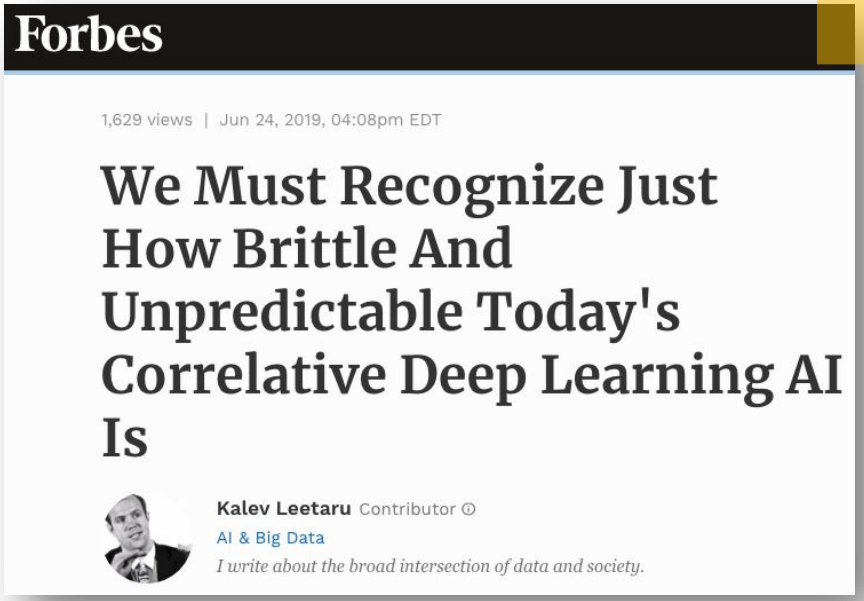
We Must Recognize Just How Brittle And Unpredictable Today's Correlative Deep Learning AI Is

Kalev Leetaru Contributor

AI & Big Data

I write about the broad intersection of data and society.

Yet few companies fully understand just how **brittle** and unpredictable today's correlative deep learning AI is and how its moments of astonishingly human-like accuracy are matched by catastrophic failure at the most unexpected moments.



Forbes

1,629 views | Jun 24, 2019, 04:08pm EDT

We Must Recognize Just How Brittle And Unpredictable Today's Correlative Deep Learning AI Is

Kalev Leetaru Contributor

AI & Big Data

I write about the broad intersection of data and society.

Yet few companies fully understand just how **brittle** and unpredictable today's correlative deep learning AI is and how its moments of astonishingly human-like accuracy are matched by catastrophic failure at the most unexpected moments. For AI to succeed companies must learn to distinguish hype from reality and understand the ways in which AI's **brittleness** and unpredictability may adversely affect their businesses.

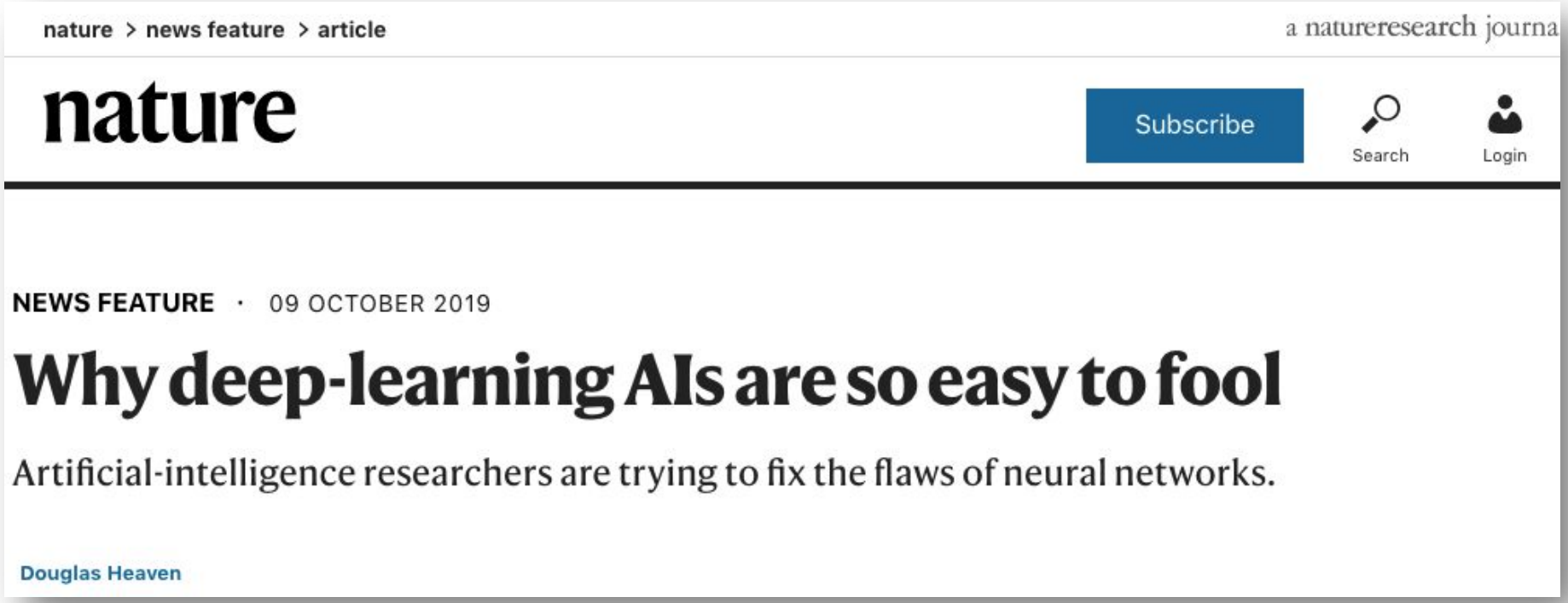
Slight Street Sign Modifications Can Completely Fool Machine Learning Algorithms

By Evan Ackerman
Posted 4 Aug 2017 | 18:00 GMT

Images: Evtimov et al

Camouflage graffiti and art stickers cause a neural network to misclassify stop signs as speed limit 45 signs or yield signs.

https://spectrum.ieee.org/cars-that-think/transportation/sensors/slight-street-sign-modifications-can-fool-machine-learning-algorithms



https://www.nature.com/articles/d41586-019-03013-5

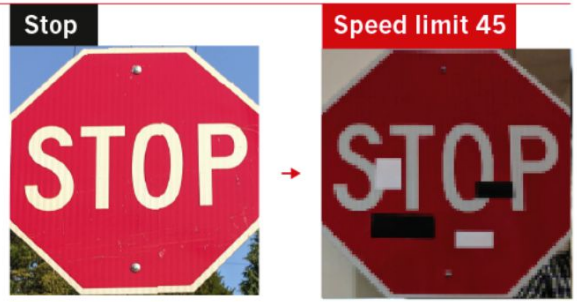
DNNs are fundamentally **brittle**: brilliant at what they do until, taken into unfamiliar territory, they break in unpredictable ways.

https://www.nature.com/articles/d41586-019-03013-5

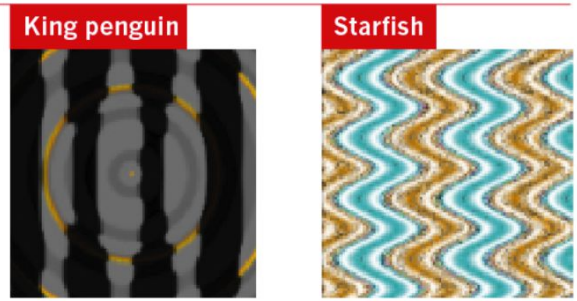
FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.

These stickers made an artificial-intelligence system read this stop sign as 'speed limit 45'.



Scientists have evolved images that look like abstract patterns — but which DNNs see as familiar objects.



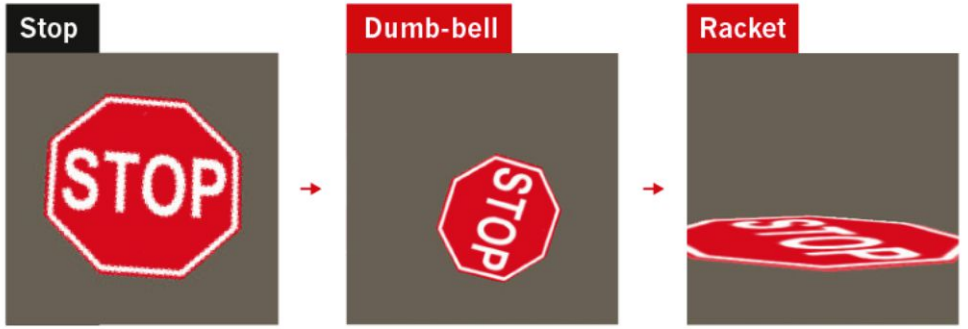
©nature

Sources: Stop sign: Ref. 1; Penguin: Ref. 5

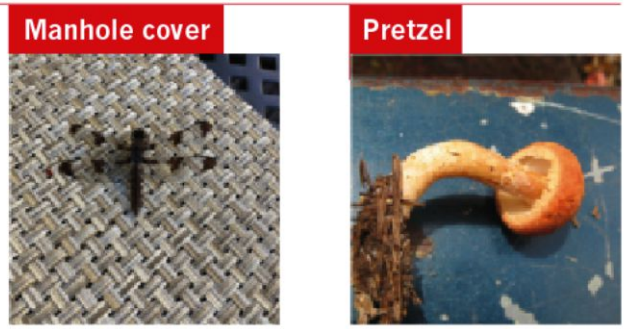
https://www.nature.com/articles/d41586-019-03013-5

LATEST TRICKS

Rotating objects in an image confuses DNNs, probably because they are too different from the types of image used to train the network.

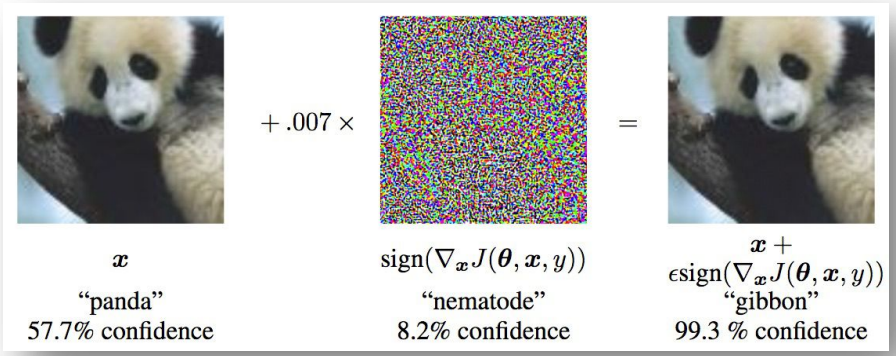


Even natural images can fool a DNN, because it might focus on the picture's colour, texture or background rather than picking out the salient features a human would recognize.



Share Image

https://www.nature.com/articles/d41586-019-03013-5



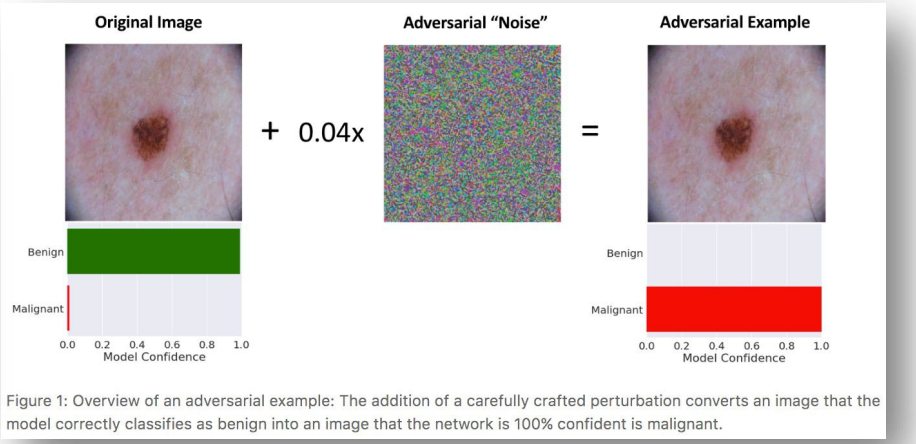
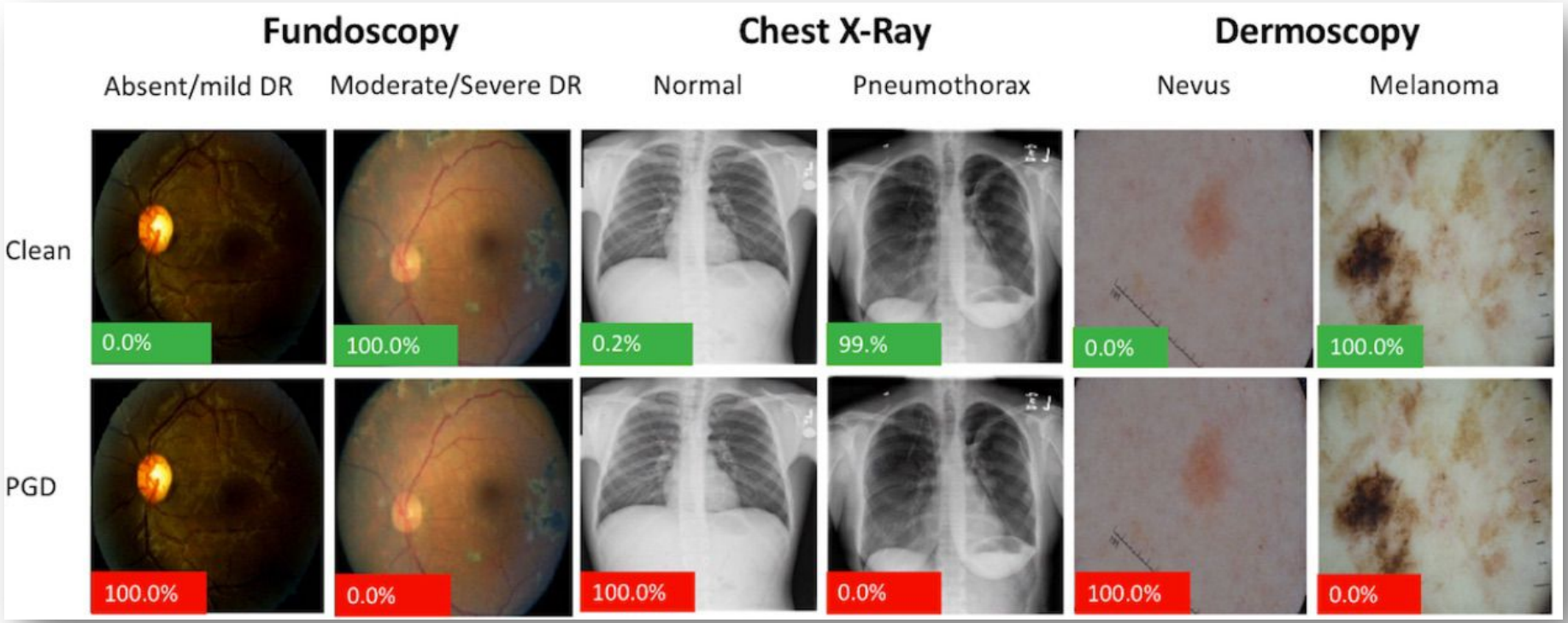


Figure 1: Overview of an adversarial example: The addition of a carefully crafted perturbation converts an image that the model correctly classifies as benign into an image that the network is 100% confident is malignant.

https://arxiv.org/pdf/1412.6572.pdf

https://www.arxiv-vanity.com/papers/1804.05296/



https://arxiv.org/pdf/1804.05296.pdf

Beyond the AI Hype

By Whitney J. Palmer

July 22, 2019

AHRA, Automation, Technology

https://www.diagnosticimaging.com/ahra/beyond-ai-hype

Harvard Business Review

Coronavirus Magazine Popular Topics Podcasts Video Store The Big Idea

INNOVATION

What AI "App Stores" Will Mean for Radiology

by Woojin Kim and Karen Holzberger

June 17, 2019

https://hbr.org/2019/06/what-ai-app-stores-will-mean-for-radiology


SIIM
MGH 1811
LABORATORY OF MEDICAL IMAGING AND COMPUTATION
VE RI TAS
Real-World Performance of Deep-Learning-Based System for Intracranial Hemorrhage Detection
Sehyo Yune, MD MPH MBA
Hyunkwang Lee, Stuart Pomerantz, Javier Romero, Shahmir Kamalian, Ramon Gonzalez, Michael Lev, Synho Do
Department of Radiology
Massachusetts General Hospital
#CMIMI18

Model Performance Comparison

Sensitivity **98%**

Specificity **95%**

PPV **95.1%**

NPV **97.9%**

Selected test dataset

		Model Prediction		
		ICH (+)	ICH (-)	
Clinical report	ICH (+)	98	2	Sensitivity: 98%
Expert confirmation	ICH (-)	5	95	Specificity: 95%
		PPV: 95.1%	NPV: 97.9%	

Real-world test dataset

		Model Prediction		
		ICH (+)	ICH (-)	
Clinical report	ICH (+)	142	21	Sensitivity: 87.1%
	ICH (-)	100	140	Specificity: 58.3%
		PPV: 12.2%	NPV: 98.5%	

NPV, negative predictive value; PPV, positive predictive value

Model Performance Comparison

Sensitivity	**98%**	**87.1%**
Specificity	**95%**	**58.3%**
PPV	**95.1%**	**12.2%**
NPV	**97.9%**	**98.5%**

Selected test dataset

		Model Prediction		
		ICH (+)	ICH (-)	
Clinical report	ICH (+)	142	21	Sensitivity: 87.1%
Expert confirmation	ICH (-)	5	95	Specificity: 95%
		PPV: 95.1%	NPV: 97.9%	

Real-world test dataset

		Model Prediction		
		ICH (+)	ICH (-)	
Clinical report	ICH (+)	142	21	Sensitivity: 87.1%
	ICH (-)	1,005	1,058	Specificity: 51.3%
		PPV: 12.2%	NPV: 98.5%	

NPV, negative predictive value; PPV, positive predictive value

Press Release

Artificial Intelligence May Fall Short When Analyzing Data Across Multiple Health Systems

Study shows deep learning models must be carefully tested across multiple environments before being put into clinical practice.

New York, NY (November 06, 2018) – Artificial intelligence (AI) tools

https://www.mountsinai.org/about/newsroom/2018/artificial-intelligence-may-fall-short-when-analyzing-data-across-multiple-health-systems

Original Article | Artificial Intelligence

eISSN 2005-8330
https://doi.org/10.3348/kjr.2019.0025
Korean J Radiol 2019;20(3):405-410

Korean Journal of Radiology
KJR

Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers

Dong Wook Kim, MD[1]*, Hye Young Jang, MD[2]*, Kyung Won Kim, MD, PhD[2], Youngbin Shin, MS[2], Seong Ho Park, MD, PhD[2]

[1]Department of Radiology, Taean-gun Health Center and County Hospital, Taean-gun, Korea; [2]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

Only 31 (**6%**) of the 516 eligible published studies performed external validation of the algorithms, and **none** met the recommended criteria for clinical validation of AI in real-world practice.

https://www.kjronline.org/DOIx.php?id=10.3348/kjr.2019.0025

Research

Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies

BMJ 2020 ; 368 doi: https://doi.org/10.1136/bmj.m689 (Published 25 March 2020)
Cite this as: *BMJ* 2020;368:m689

Linked Editorial
Artificial intelligence versus clinicians

Myura Nagendran, academic clinical fellow [1], Yang Chen, academic clinical fellow [2], Christopher A Lovejoy, physician [3], Anthony C Gordon, professor [1] [4], Matthieu Komorowski, clinical lecturer [5], Hugh Harvey, director [6], Eric J Topol, professor [7], John P A Ioannidis, professor [8], Gary S Collins, professor [9] [10], Mahiben Maruthappu, chief executive officer [3]

Conclusions: Few prospective deep learning studies and randomised trials exist in medical imaging. **Most non-randomised trials are not prospective, are at high risk of bias, and deviate from existing reporting standards.** Data and code availability are lacking in most studies, and human comparator groups are often small. Future studies should diminish risk of bias, **enhance real world clinical relevance**, improve reporting and transparency, and **appropriately temper conclusions**.

https://www.bmj.com/content/368/bmj.m689

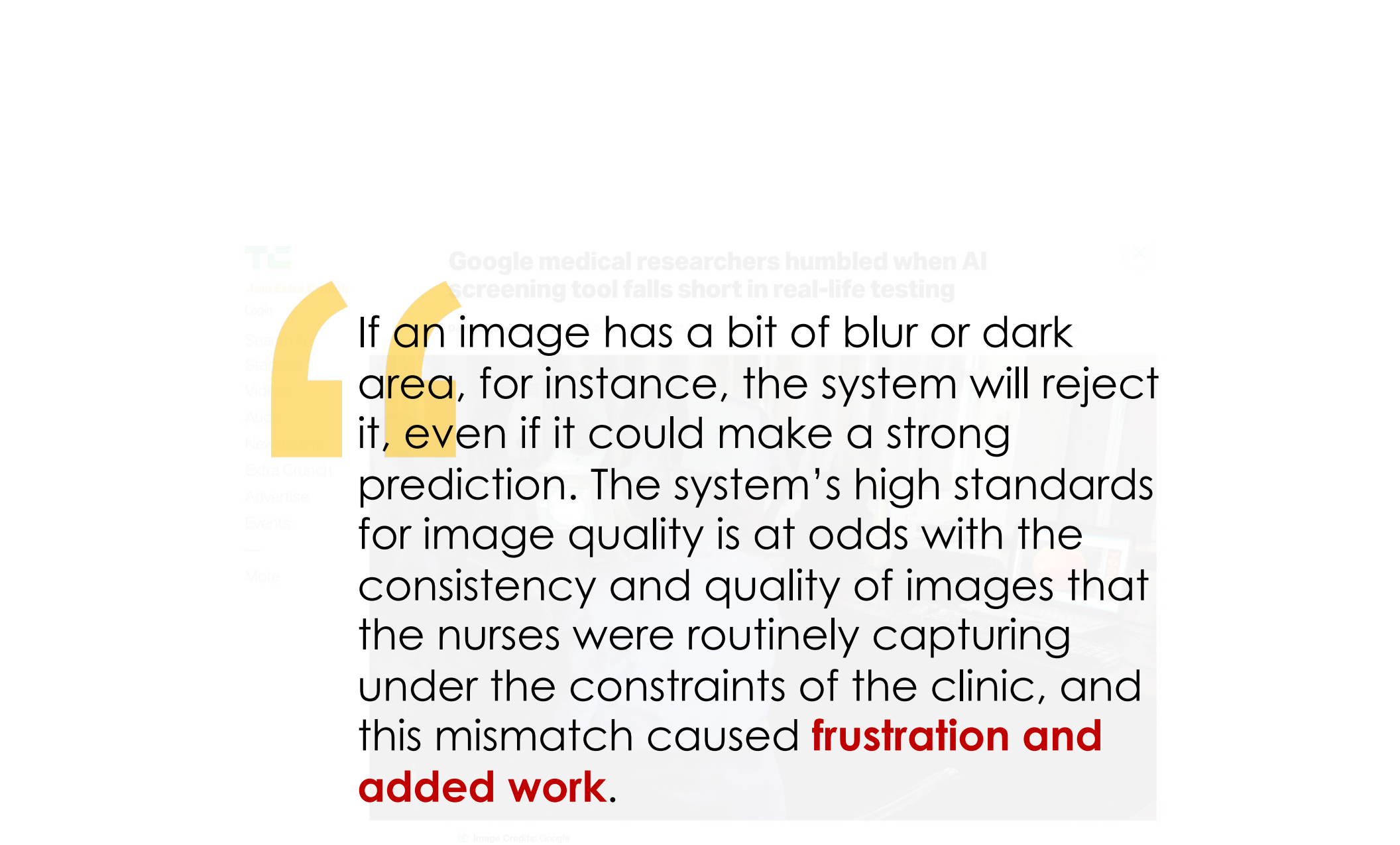
Google medical researchers humbled when AI screening tool falls short in real-life testing

Devin Coldewey @techcrunch / 2:03 pm PDT • April 27, 2020

Comment

Image Credits: Google

https://techcrunch.com/2020/04/27/google-medical-researchers-humbled-when-ai-screening-tool-falls-short-in-real-life-testing/



If an image has a bit of blur or dark area, for instance, the system will reject it, even if it could make a strong prediction. The system's high standards for image quality is at odds with the consistency and quality of images that the nurses were routinely capturing under the constraints of the clinic, and this mismatch caused **frustration and added work**.

https://techcrunch.com/2020/04/27/google-medical-researchers-humbled-when-ai-screening-tool-falls-short-in-real-life-testing/

AI techniques in medical imaging may lead to incorrect diagnoses

Date: May 12, 2020

Source: University of Cambridge

Summary: Machine learning and AI are highly unstable in medical image reconstruction, and may lead to false positives and false negatives, a new study suggests.

https://www.sciencedaily.com/releases/2020/05/200512134541.htm

"We found that the tiniest corruption, such as may be caused by a patient moving, can give a very different result if you're using AI and deep learning to reconstruct medical images -- meaning that these algorithms lack the stability they need."

https://www.sciencedaily.com/releases/2020/05/200512134541.htm


Background photo created by mindandi - www.freepik.com

DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Icons made by Freepik from www.flaticon.com

How Secure Is Your Radiology Department? Mapping Digital Radiology Adoption and Security Worldwide

Mark Stites[1]
Oleg S. Pianykh[2,3]

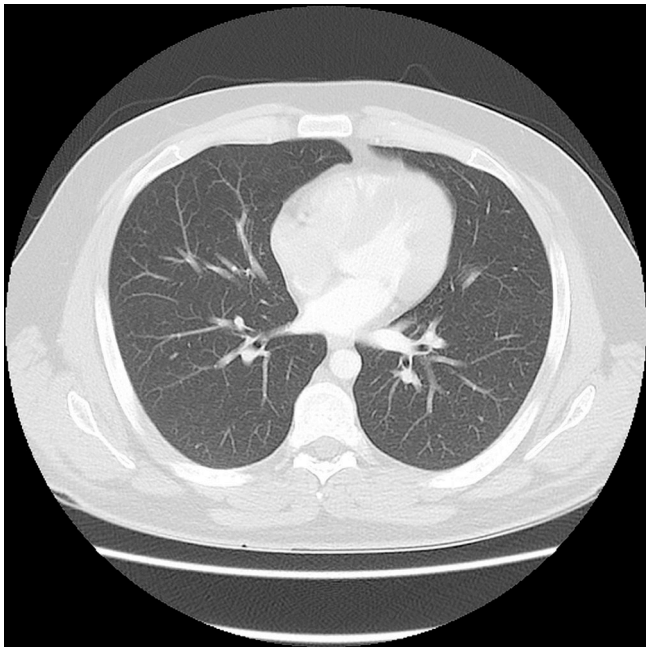
OBJECTIVE. Despite the long history of digital radiology, one of its most critical aspects—information security—still remains extremely underdeveloped and poorly standardized. To study the current state of radiology security, we explored the worldwide security of medical image archives.

MATERIALS AND METHODS. Using the DICOM data-transmitting standard, we implemented a highly parallel application to scan the entire World Wide Web of networked computers and devices, locating open and unprotected radiology servers. We used only legal and radiology-compliant tools. Our security-probing application initiated a standard DICOM handshake to remote computer or device addresses, and then assessed their security posture on the basis of handshake replies.

RESULTS. The scan discovered a total of 2774 unprotected radiology or DICOM servers worldwide. Of those, 719 were fully open to patient data communications. Geolocation was used to analyze and rank our findings according to country utilization. As a result, we built maps and world ranking of clinical security, suggesting that even the most radiology-advanced countries have hospitals with serious security gaps.

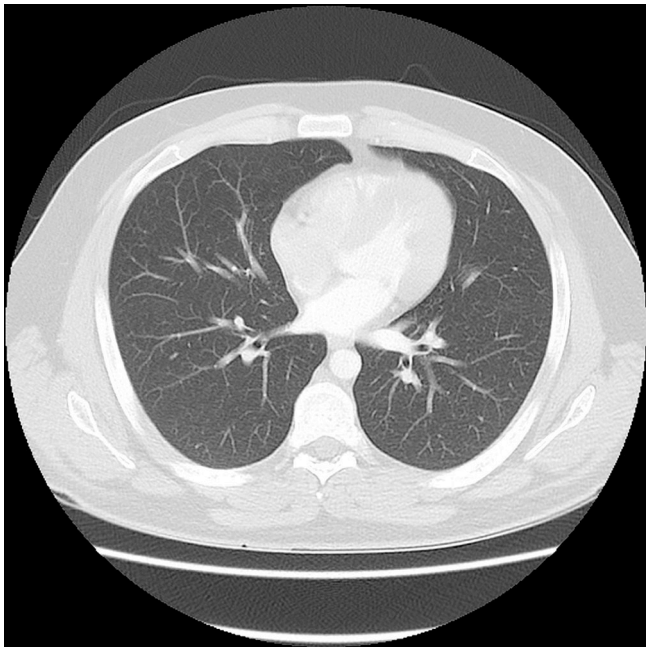
DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

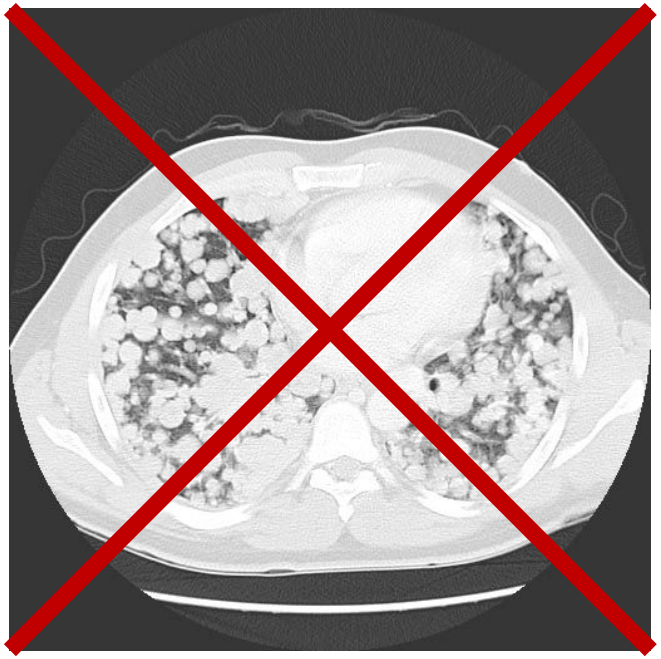
https://www.ajronline.org/doi/pdf/10.2214/AJR.15.15283



DATA SCIENCE INSTITUTE®
AMERICAN C

Case courtesy of Dr Andrew Dixon, Radiopaedia.org, rID: 36676
Case courtesy of Radswiki, Radiopaedia.org, rID: 11850
Icons made by Freepik from www.flaticon.com





0ae849d09909df9812612814

0ae721e17542eh1428424519

DATA SCIENCE INSTITUTE®
AMERICAN C

Case courtesy of Dr Andrew Dixon, Radiopaedia.org, rID: 36676
Case courtesy of Radswiki, Radiopaedia.org, rID: 11850
Icons made by Freepik from www.flaticon.com

DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY

Icons made by Freepik from www.flaticon.com

Image by Free-Photos from Pixabay

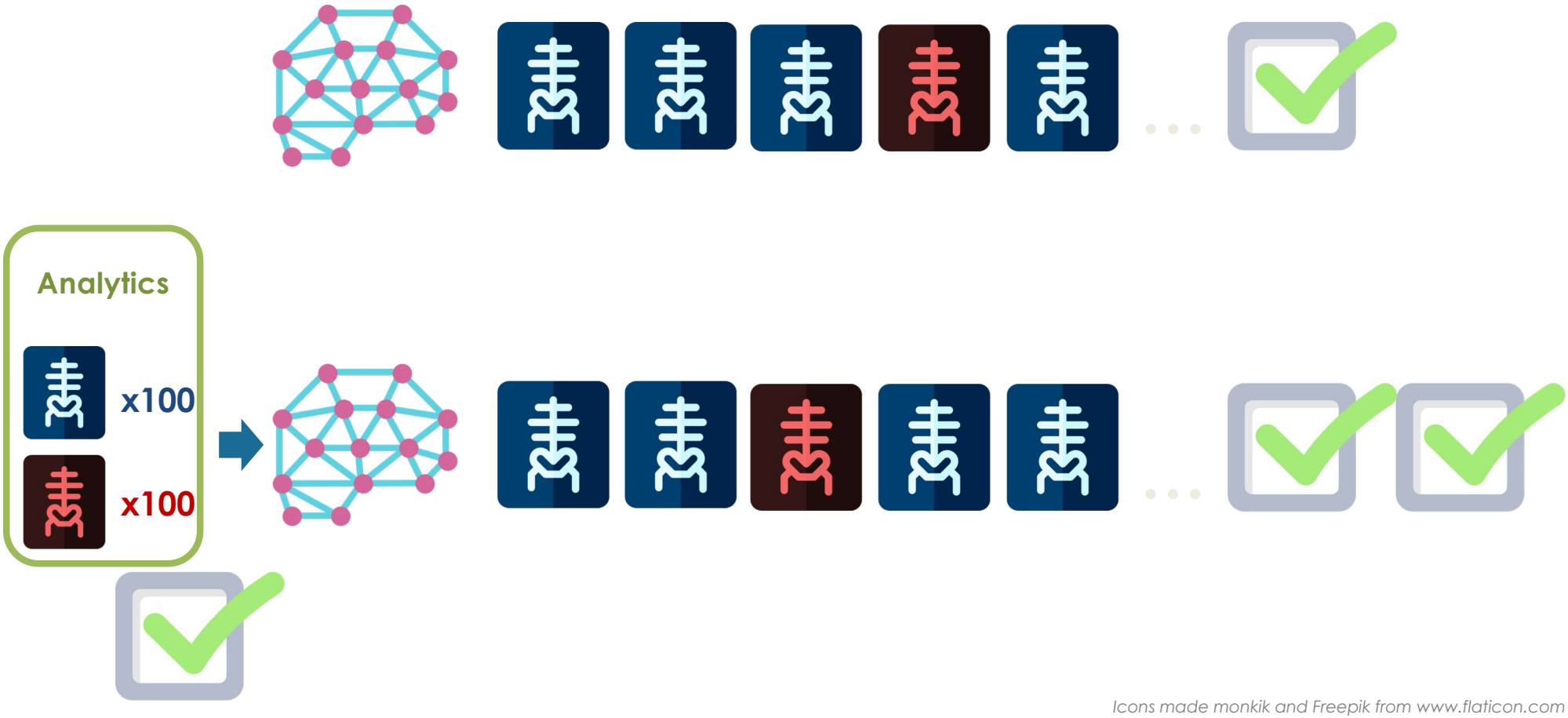
This fits 87.2% of the population without adjustments.

Photo by Artem Beliaikin from Pexels

Image by Free-Photos from Pixabay

This fits 93.4% of the population without adjustments.

Photo by Artem Beliaikin from Pexels



Icons made monkik and Freepik from www.flaticon.com

Conclusion

Banner vector created by starline - www.freepik.com

Thank you

woojinrad@gmail.com

@woojinrad

www.linkedin.com/woojinkim

DATA SCIENCE INSTITUTE®
AMERICAN COLLEGE OF RADIOLOGY