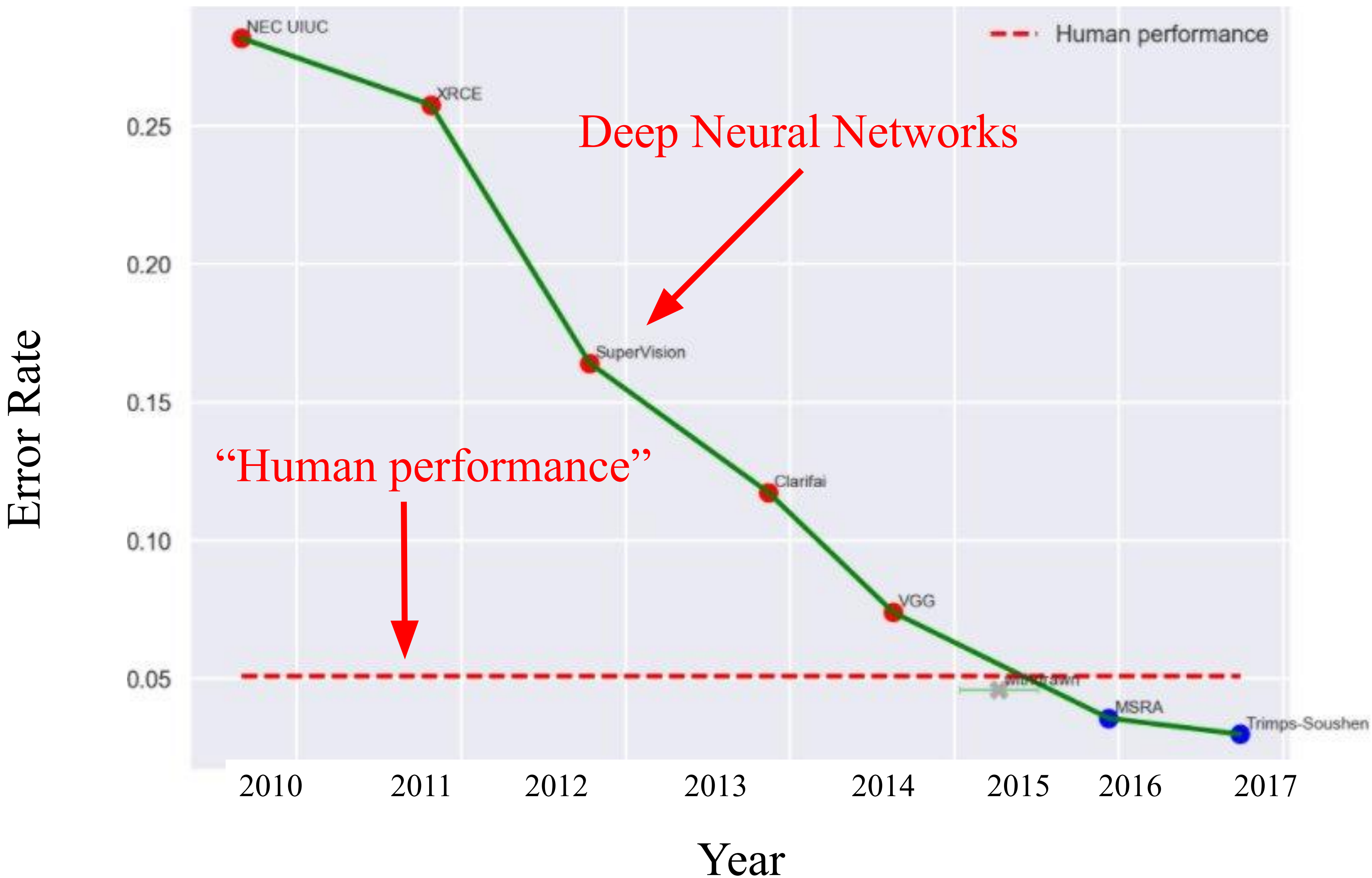
ImageNet Object Recognition:

Trained on 1.2 million images, tested on 500K images



The Guardian

Search | US edition

Computers now better than humans at recognising and sorting images

DIGITAL JOURNAL

Google's AI can now caption images almost as well as humans

BUSINESS INSIDER

TECH | FINANCE | POLITICS | STRATEGY | LIFE | ALL

10 million self-driving cars will be on the road by 2020

“Perhaps expectations are too high, and... this will eventually result in disaster…. [S]uppose that five years from now [funding] collapses miserably as autonomous vehicles fail to roll. Every startup company fails. And there's a big backlash so that you can't get money for anything connected with AI. Everybody hurriedly changes the names of their research projects to something else.

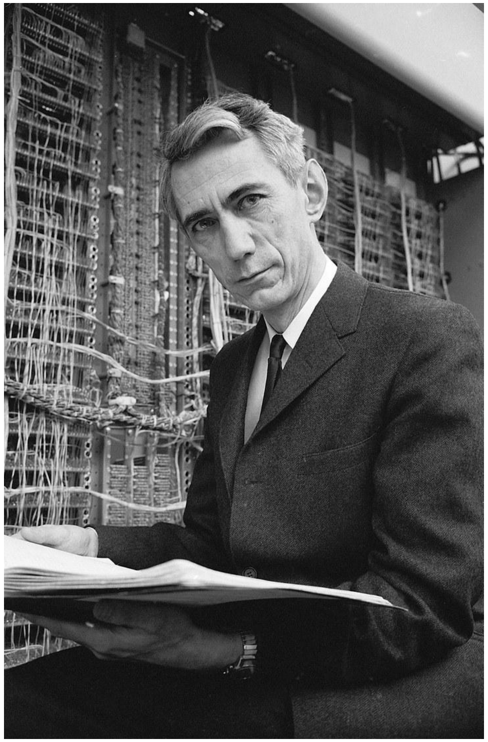
This condition [is] called the ‘AI Winter.’”

—Drew McDermott, 1984

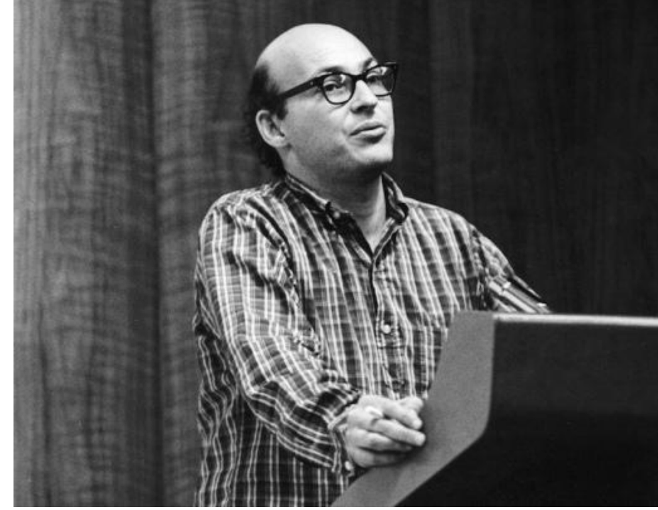
AI Winter Is Well On Its Way

POSTED 3 WEEKS AGO BY FILIP PIEKNIEWSKI

https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way/

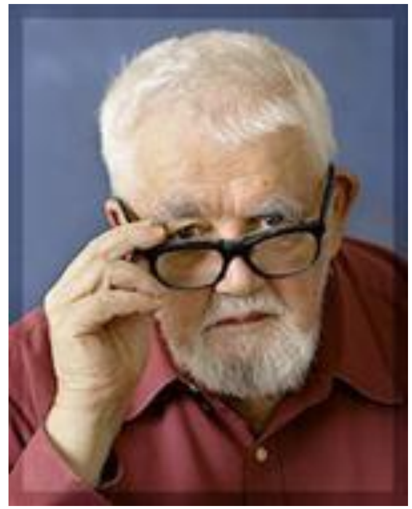


Machines will be capable, within twenty years, of doing any work that a man can do.
— Herbert Simon, 1965



Within a generation...the problem of creating ‘artificial intelligence’ will be substantially solved.
— Marvin Minsky, 1967

I confidently expect that within a matter of 10 or 15 years, something will emerge from the laboratory which is not too far from the robot of science fiction fame.
— Claude Shannon, 1961



“AI was harder than we thought.”
— John McCarthy, 2006

Human-level AI will be passed in the mid-2020s.

— Shane Legg, 2008

One of [Facebook's] goals for the next five to 10 years is to basically get better than human level at all of the primary human senses: vision, hearing, language, general cognition

— Mark Zukerberg, 2015

When will superintelligent AI arrive?...it [will] probably happen in the lifetime of my children.

(My timeline of, say, eighty years is considerably more conservative than that of the typical AI researcher.)

— Stuart Russell, 2019

Some limitations of state-of-the-art AI

- Shortcut learning
- Adversarial vulnerability
- Lack of “common sense”

Shortcut Learning

What Did My Machine Learn?

“Animal”

“No Animal”

What Did My Machine Learn?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).

school bus 1.0 garbage truck 0.99 punching bag 1.0 snowplow 0.92

What Did My Machine Learn?

Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." *arXiv preprint arXiv:1811.11553* (2018).


Tesla Totaled on 405
CULVER CITY
CBSLA.com

Shortcut Learning in Deep Neural Networks

Robert Geirhos[1,2,*,§], Jörn-Henrik Jacobsen[3,*], Claudio Michaelis[1,2,*],
Richard Zemel[†,3], Wieland Brendel[†,1], Matthias Bethge[†,1] & Felix A. Wichmann[†,1]

Article

Uncovering and Correcting Shortcut Learning in Machine Learning Models for Skin Cancer Diagnosis

Meike Nauta [1,2,*], Ricky Walsh [1,*], Adam Dubowski [1,*] and Christin Seifert [2,3,*]

Article | Open Access | Published: 11 March 2019

Unmasking Clever Hans predictors and assessing what machines really learn

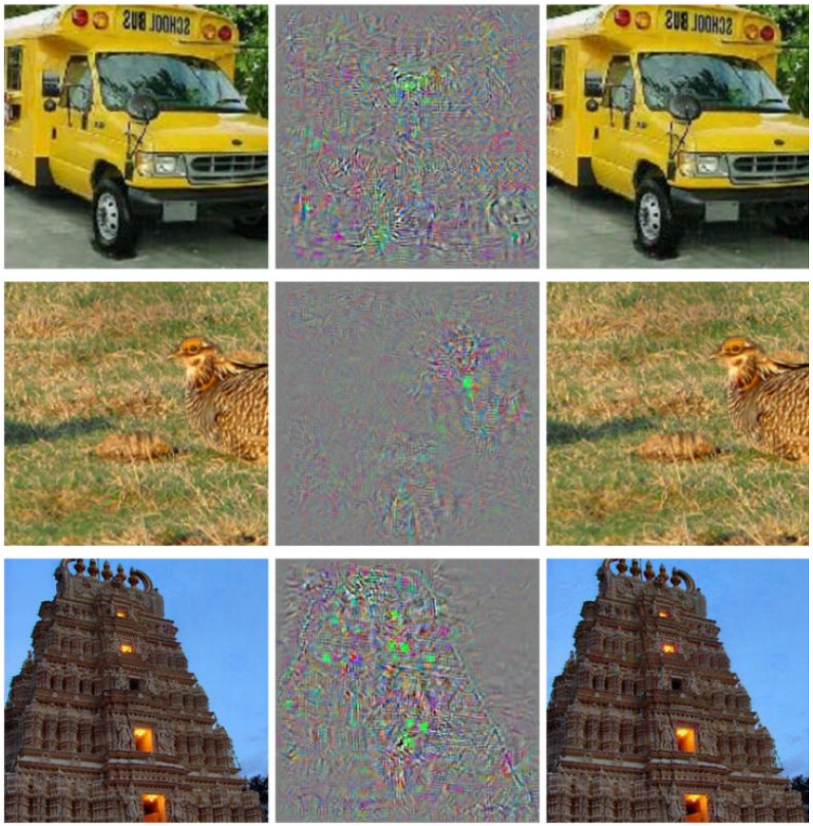
Sebastian Lapuschkin, Stephan Wäldchen, Alexander Binder, Grégoire Montavon, Wojciech Samek & Klaus-Robert Müller

Nature Communications **10**, Article number: 1096 (2019) | Cite this article

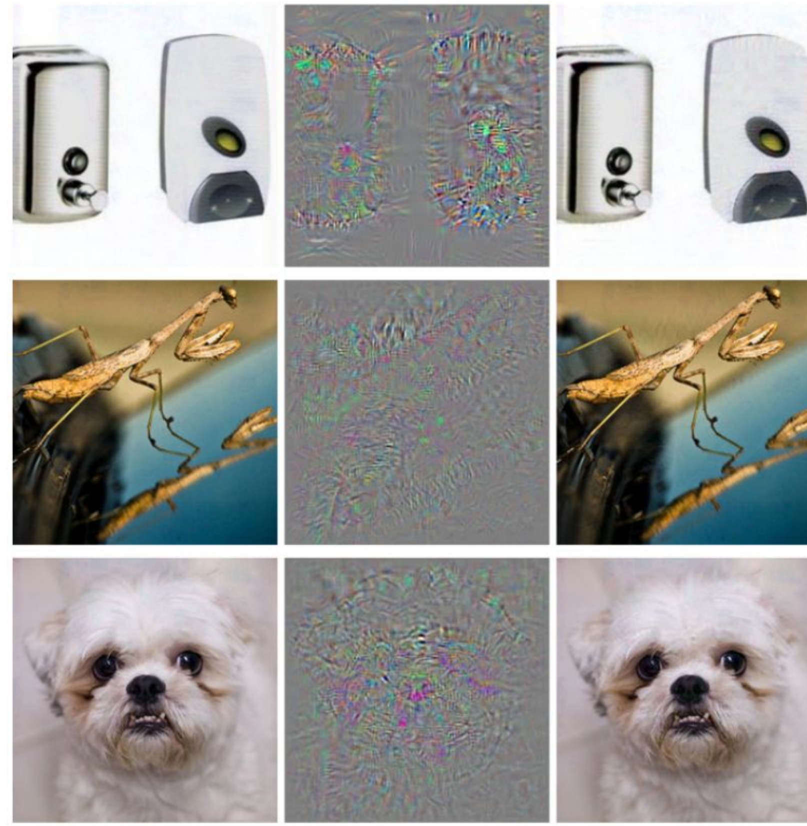
Adversarial Vulnerability

Attacks on Image Recognition Systems

“Intriguing Properties of Neural Networks”



“ostrich”



“ostrich”

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I. and Fergus, R., 2013. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*.

Attacks on Face Recognition Systems

"Accessorize to a Crime:
Real and Stealthy Attacks on State-of-the-Art Face Recognition"
Sharif et al., 2016

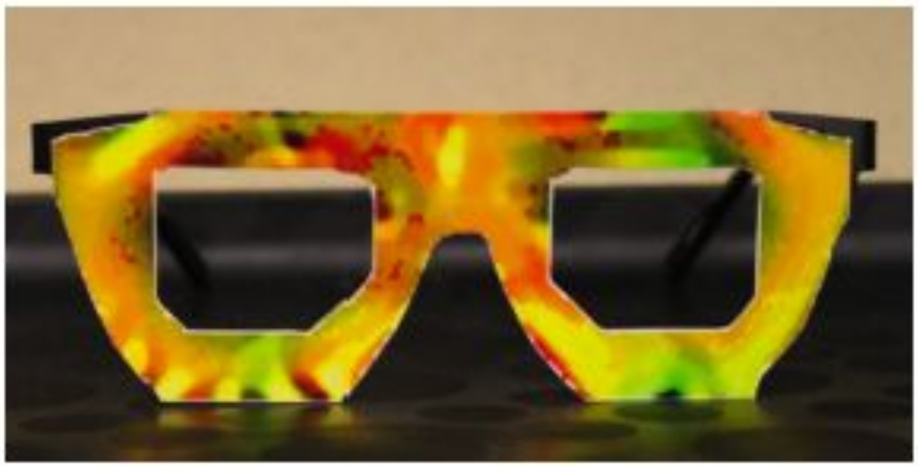


Figure 5: The eyeglass frames used by S_C for dodging recognition against DNN_B.

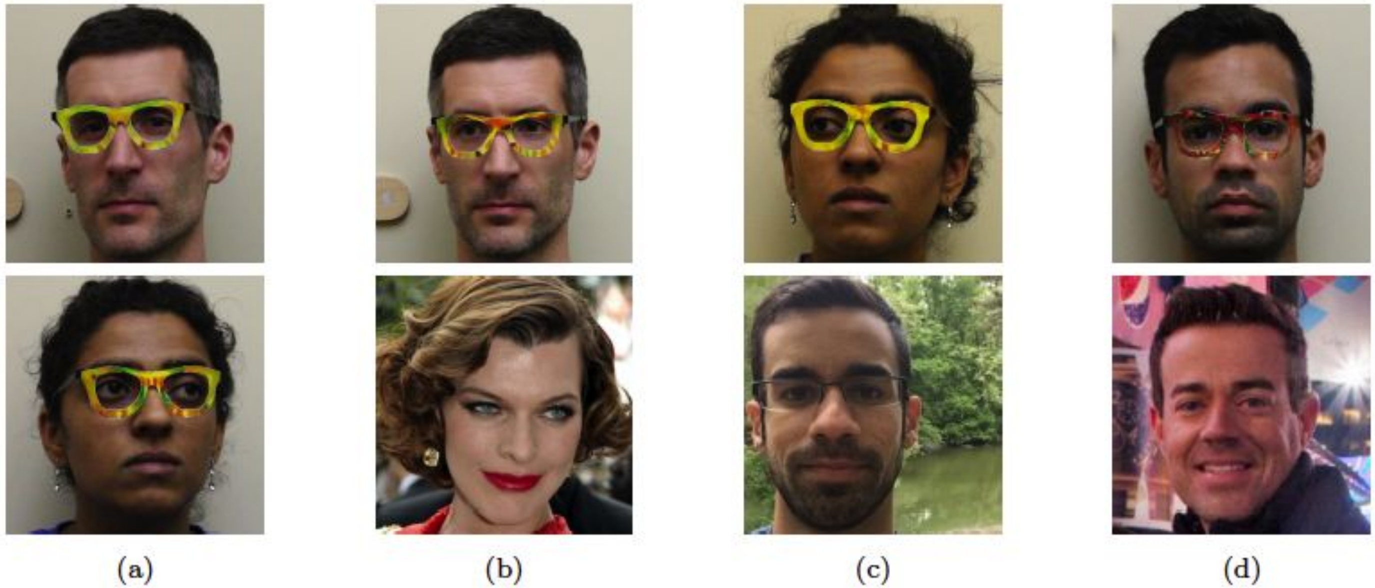
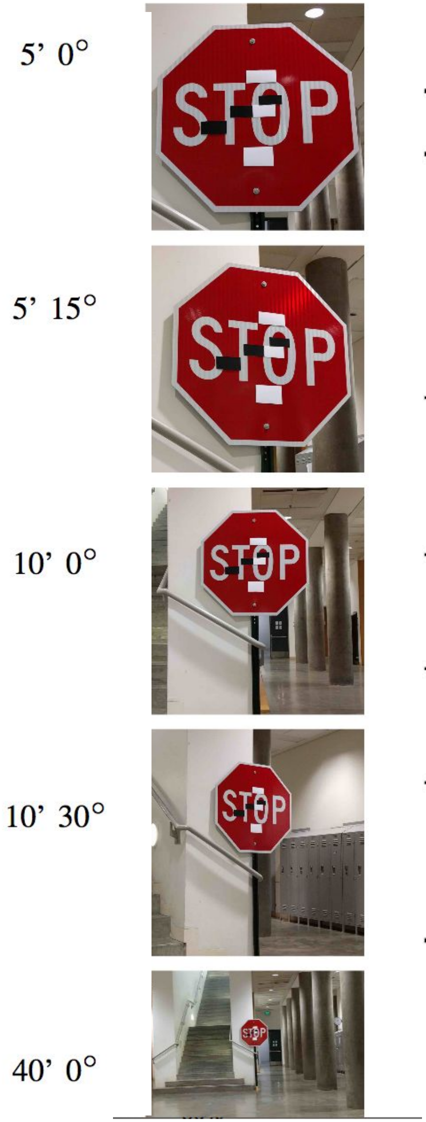


Figure 4: Examples of successful impersonation and dodging attacks. Fig. (a) shows S_A (top) and S_B (bottom) dodging against DNN_B. Fig. (b)–(d) show impersonations. Impersonators carrying out the attack are shown in the top row and corresponding impersonation targets in the bottom row. Fig. (b) shows S_A impersonating Milla Jovovich (by Georges Biard; source: https://goo.gl/GlsWlC); (c) S_B impersonating S_C; and (d) S_C impersonating Carson Daly (by Anthony Quintano; source: https://goo.gl/VfnDct).

Attacks on Autonomous Driving Systems



Target: “Speed Limit 80”

Distance & Angle	Top Class (Confid.)	Second Class (Confid.)
5’ 0°	Speed Limit 80 (0.88)	Speed Limit 70 (0.07)
5’ 15°	Speed Limit 80 (0.94)	Stop (0.03)
5’ 30°	Speed Limit 80 (0.86)	Keep Right (0.03)
5’ 45°	Keep Right (0.82)	Speed Limit 80 (0.12)
5’ 60°	Speed Limit 80 (0.55)	Stop (0.31)
10’ 0°	Speed Limit 80 (0.98)	Speed Limit 100 (0.006)
10’ 15°	Stop (0.75)	Speed Limit 80 (0.20)
10’ 30°	Speed Limit 80 (0.77)	Speed Limit 100 (0.11)
15’ 0°	Speed Limit 80 (0.98)	Speed Limit 100 (0.01)
15’ 15°	Stop (0.90)	Speed Limit 80 (0.06)
20’ 0°	Speed Limit 80 (0.95)	Speed Limit 100 (0.03)
20’ 15°	Speed Limit 80 (0.97)	Speed Limit 100 (0.01)
25’ 0°	Speed Limit 80 (0.99)	Speed Limit 70 (0.0008)
30’ 0°	Speed Limit 80 (0.99)	Speed Limit 100 (0.002)
40’ 0°	Speed Limit 80 (0.99)	Speed Limit 100 (0.002)

Evtimov et al., “Robust Physical-World Attacks on Deep Learning Models”, 2017

Lack of “common sense”

JACK STEWART TRANSPORTATION 10.10.2018 07:00 AM

Why People Keep Rear-Ending Self-Driving Cars

Human drivers (and one cyclist) have rear-ended self-driving cars 28 times this year in California—accounting for nearly two-thirds of robocar crashes.

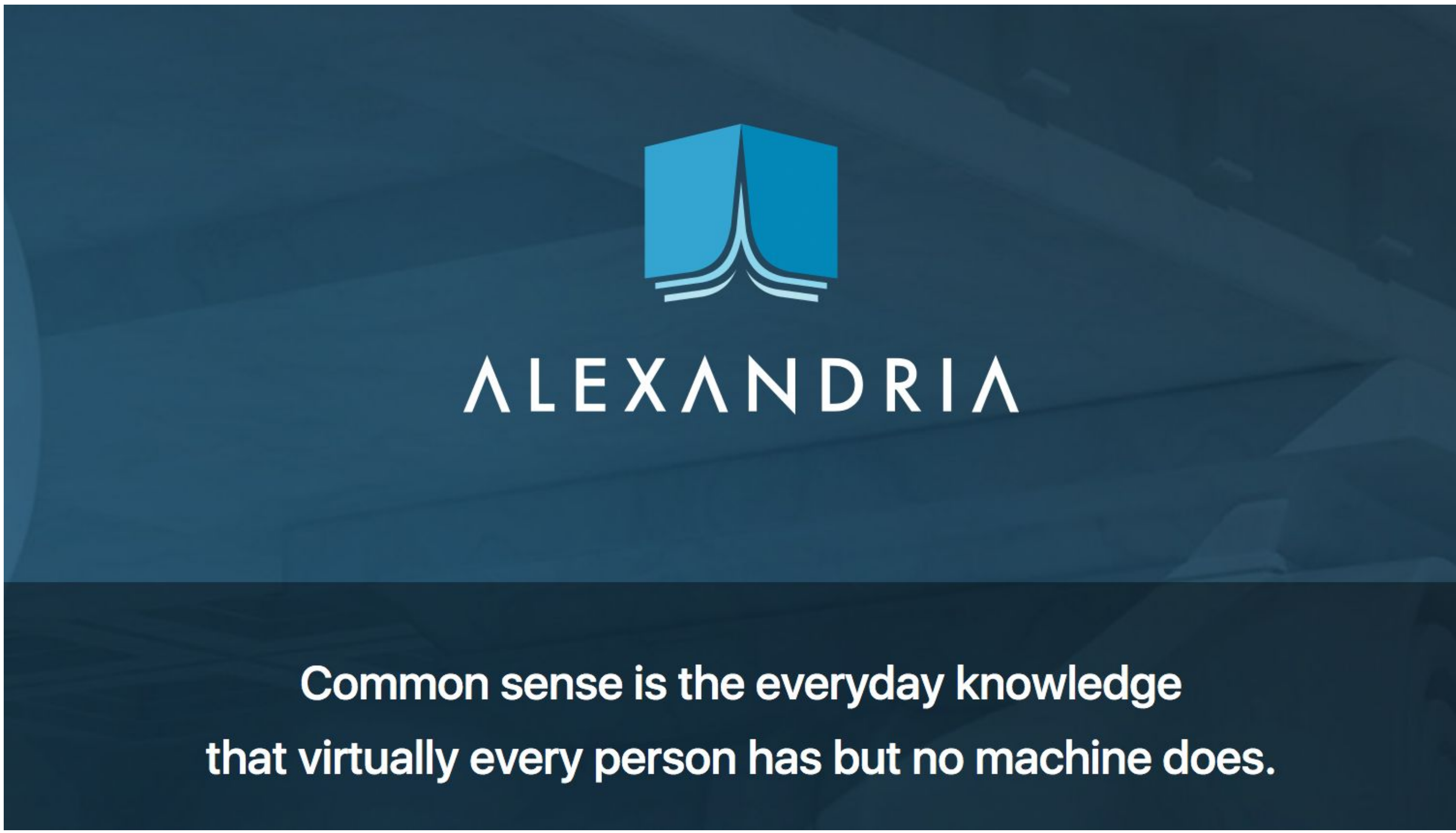

30
MPH


Virgin

> “COMMON SENSE” is the Dark Matter of Artificial Intelligence.

Paul Allen invests $125 million to teach computers common sense

https://www.seattletimes.com/business/technology/paul-allen-invests-125-million-to-teach-computers-common-sense/



https://allenai.org/alexandria/

Department of Defense
Fiscal Year (FY) 2019 Budget Estimates

February 2018

Defense Advanced Research Projects Agency

Title: Machine Common Sense (MCS)

Description: The Machine Common Sense (MCS) program will explore approaches to commonsense reasoni
Recent advances in machine learning have resulted in exciting new artificial intelligence (AI) capabilities in area
recognition, natural language processing, and two-person strategy games (Chess, Go). But in all of these appl
the machine reasoning is narrow and highly specialized; broad, commonsense reasoning by machines remains
program will create more human-like knowledge representations, for example, perceptually-grounded represen
commonsense reasoning by machines about the physical world and spatio-temporal phenomena. Equipping A
more human-like reasoning capabilities will make it possible for humans to teach/correct a machine as they inte
on tasks, enabling more equal collaboration and ultimately symbiotic partnerships between humans and machi

FY 2019 Plans:

- Develop approaches for machine reasoning about imprecise and uncertain information derived from text, pic
speech, and sensor data.
- Design methods to enable machines to identify knowledge gaps and reason about their state of knowledge.
- Formulate perceptually-grounded representations to enable commonsense reasoning by machines about the
spatio-temporal phenomena.

Why AI is Harder Than We Think

Melanie Mitchell
Santa Fe Institute
Santa Fe, NM, USA
mm@santafe.edu

Fallacy 1: Narrow AI is on a continuum with general AI

IBM® Watson™ represents a first step into cognitive systems, a new era of computing.

AlphaZero … is the first step in creating real AI.

GPT-2 AS STEP TOWARD GENERAL INTELLIGENCE

Hubert Dreyfus: "The **first-step fallacy** is the claim that, ever since our first work on computer intelligence we have been inching along a continuum at the end of which is AI, so that any improvement in our programs no matter how trivial counts as progress….There was in fact a discontinuity in the assumed continuum of steady incremental progress. The unexpected obstacle was called the commonsense knowledge problem."

Stuart Dreyfus: "It [is] like claiming that the first monkey that climbed a tree was making progress towards landing on the moon."

Fallacy 2: Easy things are easy and hard things are hard

Herbert Simon: "Everything of interest in cognition happens above the 100-millisecond level, the time it takes to recognize your mother."

Andrew Ng: "If a typical person can do a mental task with less than one second of thought, we can probably automate it using AI either now or in the near future."

Demis Hassibis et al.: Go is one of "the most challenging of domains."

Moravec's paradox: "It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility." **— and common sense!**

Marvin Minsky: "In general, we're least aware of what our minds do best."

Modern wishful mnemonics:

"Watson can read all of the health-care texts in the world in seconds."

"Watson understands context and nuance in seven languages."

"AlphaGo's goal is to beat the best human players not just mimic them."

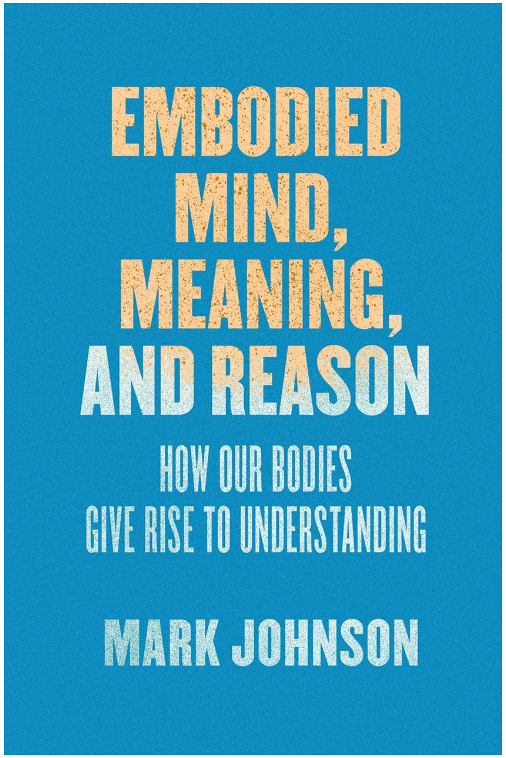
Benchmark datasets called "reading comprehension" , "common sense understanding", "general language understanding evaluation"

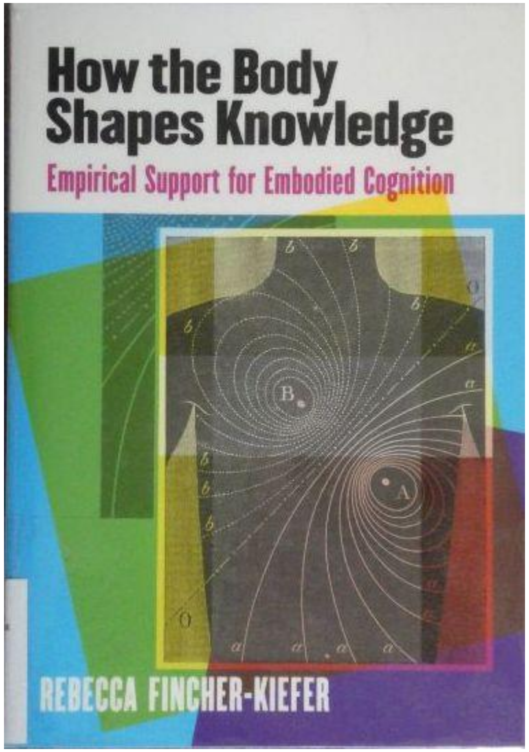
Methods called "deep *learning*", "*neural* networks"

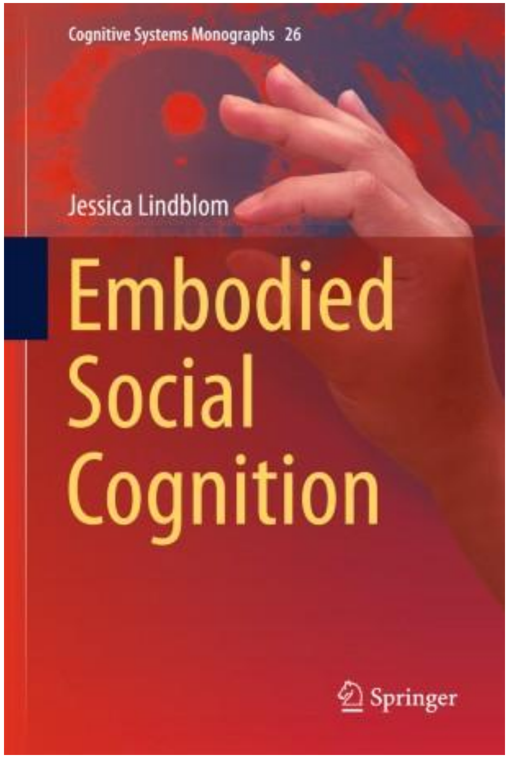
Fallacy 4: Intelligence is all in the brain

Joseph Carlsmith: "I think it more likely than not that 10^{15} FLOP/s is enough to perform tasks as well as the human brain (given the right software, which may be very hard to create)."

Geoffrey Hinton: "To understand [documents] at a human level, we're probably going to need human-level resources and we have trillions of connections [in our brains]. ...But the biggest networks we have built so far only have billions of connections. So we're a few orders of magnitude off, but I'm sure the hardware people will fix that."


EMBODIED
MIND,
MEANING,
AND REASON
HOW OUR BODIES
GIVE RISE TO UNDERSTANDING
MARK JOHNSON


How the Body
Shapes Knowledge
Empirical Support for Embodied Cognition
REBECCA FINCHER-KIEFER


Cognitive Systems Monographs 26
Jessica Lindblom
Embodied
Social
Cognition
Springer

Open questions spurred by these fallacies

Fallacy 1: Narrow AI is on a continuum with general AI

- How can we assess actual progress toward "general" or "human-level" AI?

Fallacy 2: Easy things are easy and hard things are hard

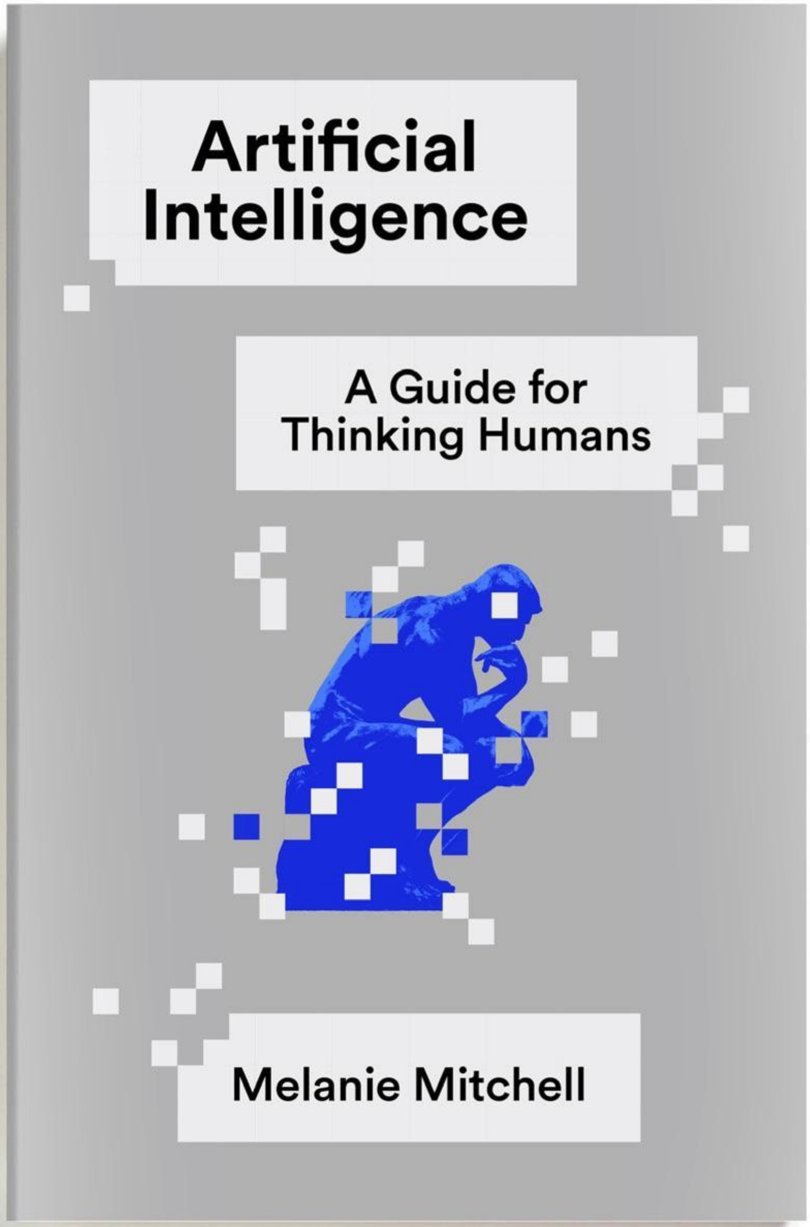
- How can we assess the difficulty of a domain for AI?

Fallacy 3: The lure of "wishful mnemonics"

- How do we talk to ourselves about what machines can and cannot do without fooling ourselves with wishful mnemonics?

Fallacy 4: Intelligence is all in the brain

- How embodied (and socially/culturally embedded) does intelligence need to be?



Thank you for listening!