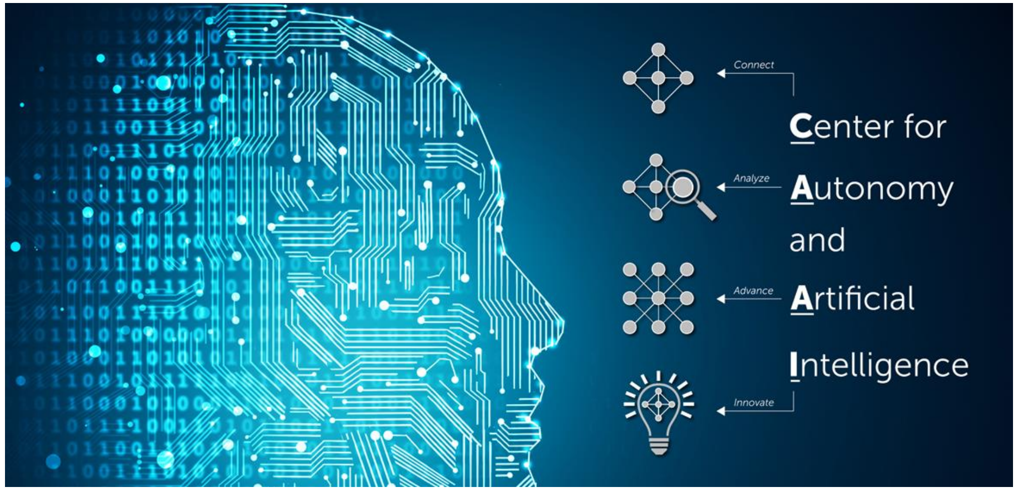


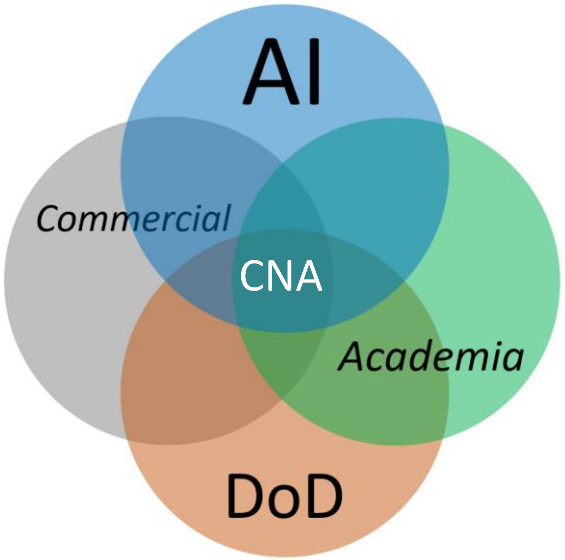
What is AI?

A panel discussion on the opportunities and challenges presented by artificial intelligence

22 January 2019

CNA
ANALYSIS & SOLUTIONS

Discussion Points

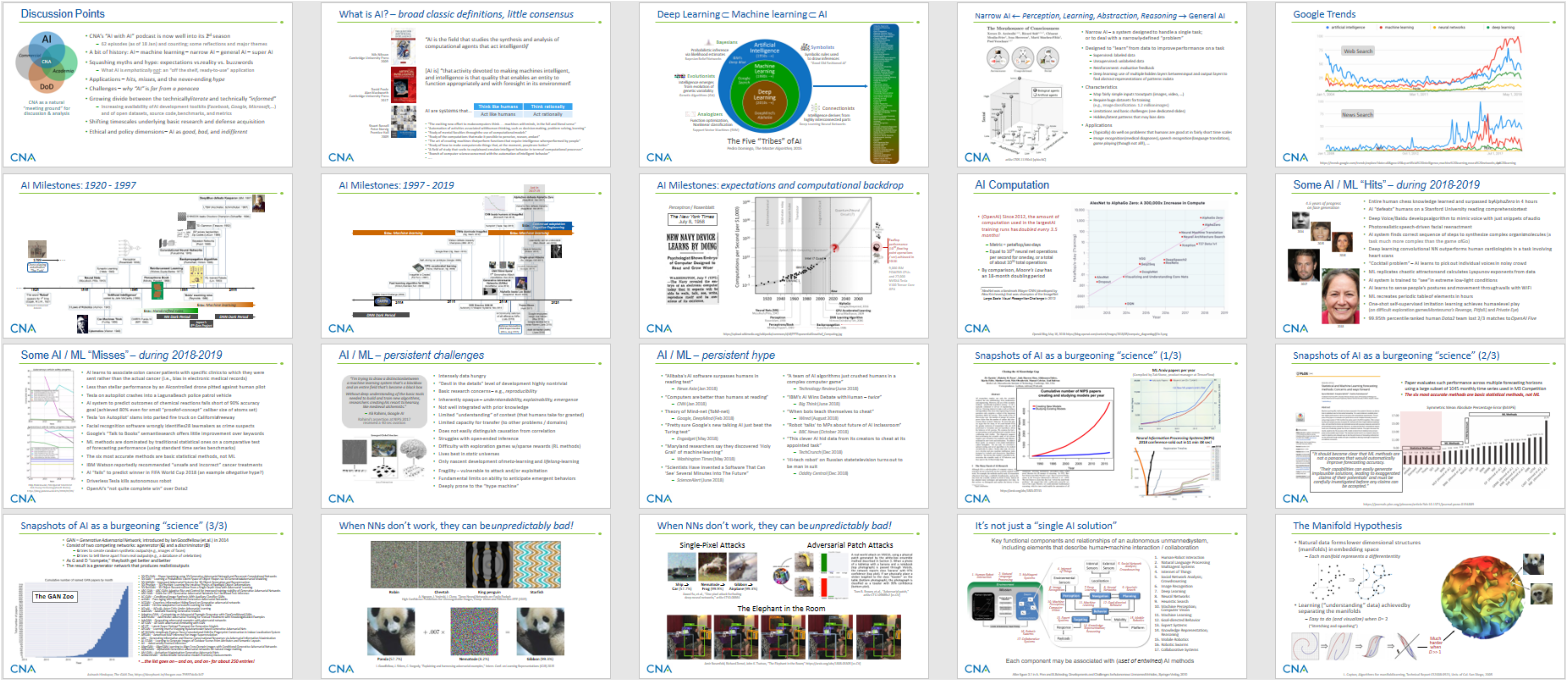


CNA as a natural "meeting ground" for discussion & analysis

- CNA's "AI with AI" podcast is now well into its 2nd season
  - 62 episodes (as of 18 Jan) and counting; some reflections and major themes
- A bit of history: AI – machine learning – narrow AI – general AI – super AI
- Squashing myths and hype: expectations vs. reality vs. buzzwords
  - What AI is *emphatically* *not*: an "off the shelf, ready-to-use" application
- Applications – *hits, misses*, and the never-ending *hype*
- Challenges – *why "AI"* is *far from a panacea*
- Growing divide between the technically *literate* and technically *"informed"*
  - Increasing availability of AI development toolkits (*Facebook*, *Google*, *Microsoft*,…) and of open datasets, source code, benchmarks, and metrics
- Shifting timescales underlying basic research and defense acquisition
- Ethical and policy dimensions – AI as *good*, *bad*, and *indifferent*

CNA

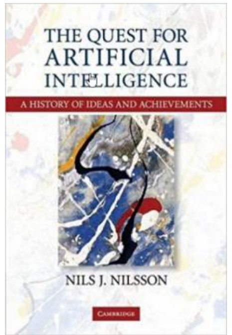
This is not a formal presentation – only visual backdrops



CNA

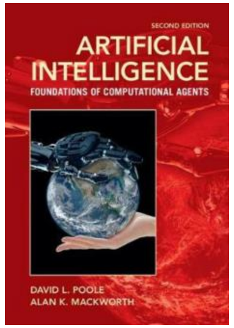
What is AI? – *broad classic definitions, little consensus*

Nils Nilsson
Cambridge University Press
2009



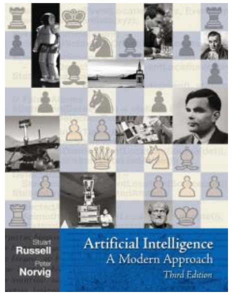
"AI is the field that studies the synthesis and analysis of computational agents that act intelligently"

David Poole
Alan Mackworth
Cambridge University Press
2017



[AI is] "that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment"

Stuart Russell
Peter Norvig
Prentice Hall
2009



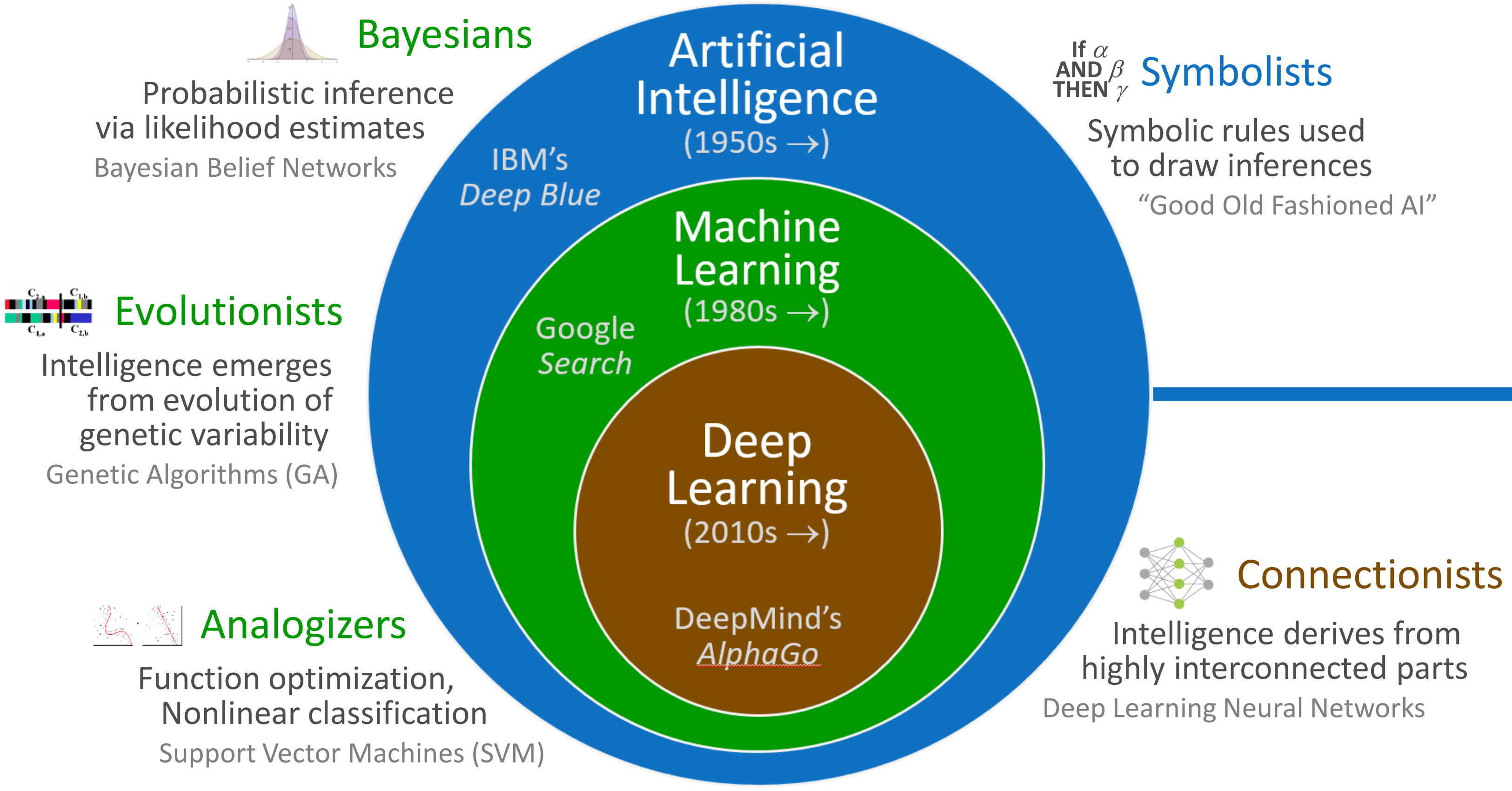
AI are systems that...

Think like humans	Think rationally
Act like humans	Act rationally

- "The exciting new effort to make computers think . . . machines with minds, in the full and literal sense"
- "Automation of activities associated with human thinking, such as decision-making, problem solving, learning"
- "Study of mental faculties through the use of computational models"
- "Study of the computations that make it possible to perceive, reason, and act"
- "The art of creating machines that perform functions that require intelligence when performed by people"
- "Study of how to make computers do things that, at the moment, people are better"
- "A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes"
- "Branch of computer science concerned with the automation of intelligent behavior"
- …..

CNA

Deep Learning ⊂ Machine learning ⊂ AI



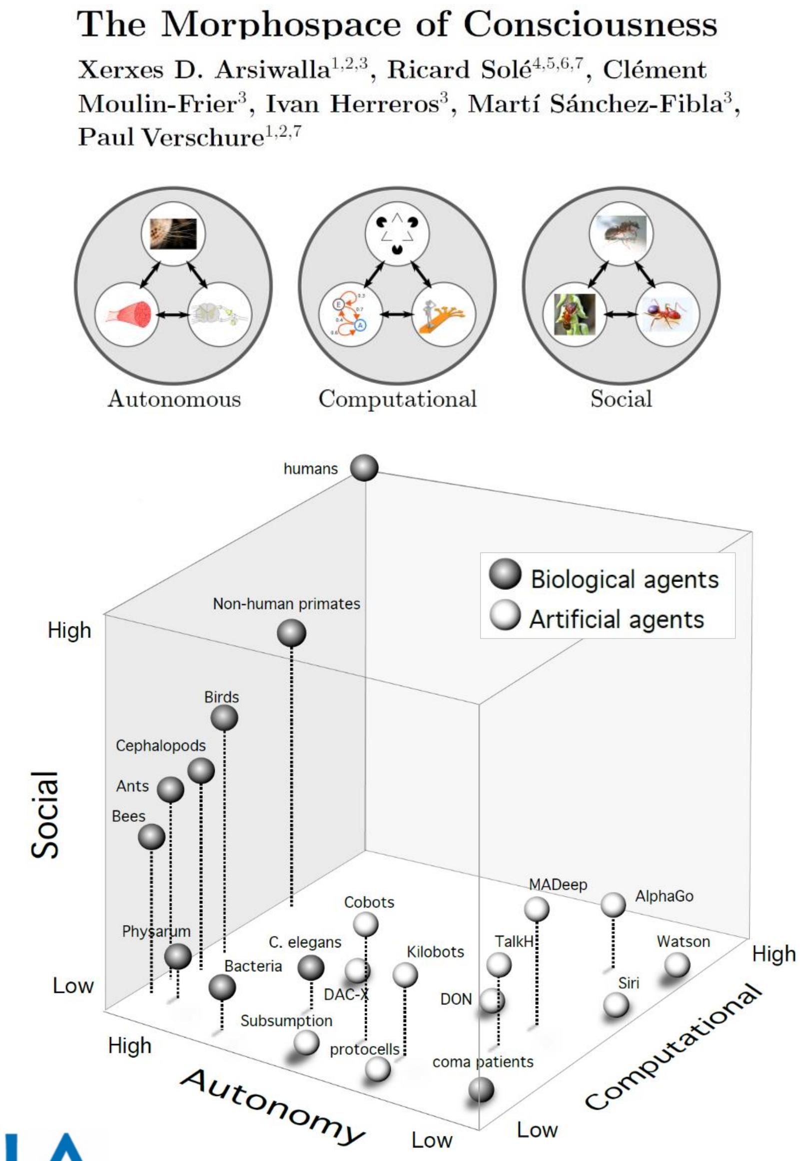
The Five "Tribes" of AI

Pedro Domingo, *The Master Algorithm*, 2015

- *Adaptive Regression*
- *Automated Reasoning*
- *AutoML*
- *Back-Propagation*
- *Bayesian Decision Theory*
- *Behavior-Based AI / Robotics*
- *Bootstrapped Aggregation*
- *Capsule Networks*
- *Case-Based Reasoning*
- *Causal Inference*
- *Cognitive Modeling*
- *Common Sense Knowledge*
- *Complex Adaptive Systems*
- *Computational Intelligence*
- *Computer Vision*
- *Decision Trees*
- *Deep Belief Networks*
- *Deep Learning*
- *Differentiable neural networks*
- *Dimensionality Regression*
- *Discriminant Analysis*
- *Distributed AI*
- *Epistemology*
- *Evolutionary Computing*
- *Expert System*
- *Genetic Programming*
- *Gradient Boosting Machines*
- *Heuristics / Metaheuristics*
- *Hierarchical Clustering*
- *Hierarchical Temporal Memory*
- *Hopfield Network (HN)*
- *Image Recognition*
- *Inference*
- *Information Theory*
- *Information Retrieval*
- *Instance-Based Learning*
- *K-Nearest Neighbors*
- *Knowledge Graphs*
- *Knowledge Representation*
- *Learning from Experience*
- *Least Squares Regression*
- *Lifelong Learning*
- *Linear Regression*
- *Logical AI*
- *Logistic Regression*
- *Long Short-Term Memory*
- *Machine Learning*
- *Markov Chains*
- *Multiagent Modeling / Systems*
- *Multilayer Perceptrons*
- *Naïve Bayes*
- *Natural Language Processing*
- *Neural Networks*
- *Neuro-linguistic Programming*
- *Ontology*
- *Pattern Recognition*
- *Planning*
- *Principal Component Regression*
- *Probabilistic Graph Models*
- *Probably Approximately Correct*
- *Radial Basis Function Network*
- *Random Forrest*
- *Recurrent Neural Networks*
- *Reinforcement Learning*
- *Reasoning*
- *Search*
- *Self-Organizing Map*
- *Semantic Inference*
- *Semantic Web*
- *Similarity Search*
- *Situated / Embodied Agents*
- *Stacked Autoencoders*
- *Statistical AI*
- *Stochastic Optimization*
- *Supervised Learning*
- *Support Vector Machines*
- *Swarm Intelligence*
- *Unsupervised Learning*

CNA

Narrow AI ← *Perception, Learning, Abstraction, Reasoning* → General AI

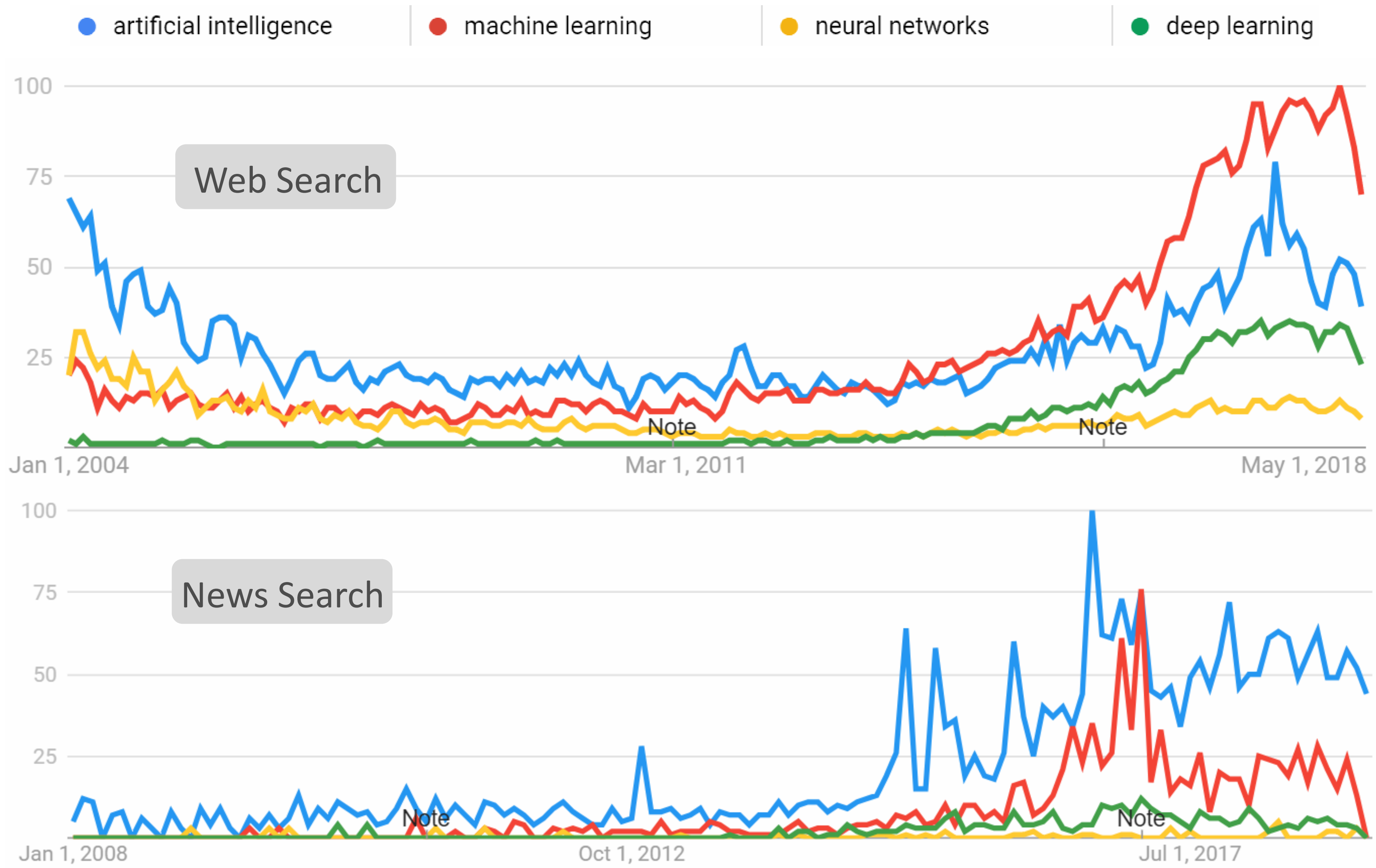


arXiv:1705.11190v3 [q-bio.NC]

- Narrow AI – a system designed to handle a single task; or to deal with a narrowly-defined "problem"
- Designed to "learn" from data to improve performance on a task
  - Supervised: labeled data
  - Unsupervised: unlabeled data
  - Reinforcement: evaluative feedback
  - Deep learning: use of multiple hidden layers between input and output layers to find abstract representations of patterns in data
- Characteristics
  - Map fairly simple inputs to outputs (images, video, …)
  - Require huge datasets for training (e.g., Image classification: 1.2 million images)
  - Limitations and basic challenges (see dedicated slides)
  - Hidden/latent patterns that may bias data
- Applications
  - (Typically) do well on problems that humans are good at in fairly short time scales
  - *Image recognition* (medical diagnoses), *speech recognition* (language translation), *game playing* (though not all!), …

CNA

Google Trends



https://trends.google.com/trends/explore?date=all&geo=US&q=artificial%20intelligence,machine%20learning,neural%20networks,deep%20learning

AI Milestones: *1920 - 1997*

"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire.

The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions, and concepts, solve kinds of problems now reserved for humans, and improve themselves."

– *Dartmouth AI Project Proposal*, John McCarthy, 1955

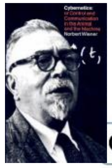
1789 → . . .
Chess-playing automaton 'The Turk'

. . . 1920 · 1940 · 1960 · 1980 · 2000

The word **'Robot'** appears for 1st time (Capek, R.U.R., 1921)
Rossum's Universal Robots

Neural Nets (Mcculloch, Pitts: 1943)

3 Laws of Robotics (Asimov: 1942)

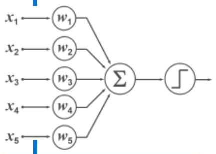


Cybernetics (Weiner: 1948)

Synaptic Learning (Hebb: 1949)

Can Machines Think (Turing: 1950)

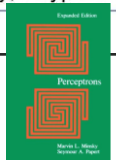
'Artificial Intelligence' coined by McCarthy (1955)



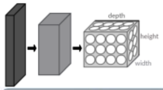
Perceptron (Rosenblatt: 1958)

DARPA Funds AI (MIT: 1963)

Perceptrons Book (Minsky, Papert: 1969)

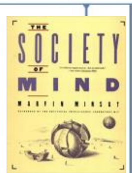


Reinforcement Learning (Widrow,Gupta,Maitra: 1973)



Convolutional Neural Networks (Fukushima: 1979)

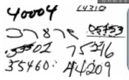
Backpropagation Algorithm (Rumelhart, Hinton: 1986)



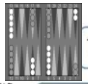
Minsky 1985

'Boids' swarming rules (Reynolds: 1986)

Bayesian Networks (Pearl: 1988)

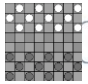


BP solves handwritten Zip Codes (LeCun: 1989)



TD–Gammon (Tesauro: 1992)

RL–trained Robots (Lin: 1993)



CHINOOK beats *Checkers* Champion (Schaeffer: 1994)

LTSM (Hochreiter, Schmidhuber: 1997)

DeepBlue defeats Kasparov (IBM: 1997)

Era$_1$: *Handcrafted rules*

Era$_2$: *Machine learning*

NN Dark Period

Japan's 5th Gen Project

DNN Dark Period

CNA

AI Milestones: *1997 - 2019*

Just in 2017-19

Era$_2$: *Machine learning*

Era$_3$: *Contextual adaptation*
Cognitive Engineering

DNN Dark Period

DARPA Grand Challenge (2004)

Fast learning algorithm for DNNs
(Hinton,Osindero,Teh: 2006)

ImageNet is Created
(Fei Fei Li: 2007)

GPU–accelerated learning
(Raina, Madhavan, Ng: 2009)

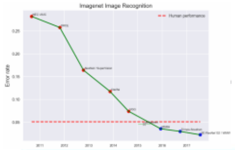
Self–driving car prototype (Google: 2009)

Google Brain (Ng, Dean: 2010)

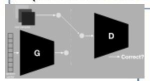
Watson defeats Jeopardy Champions (IBM: 2011)

CNNs dominate ImageNet
(Ng, Dean: 2012)

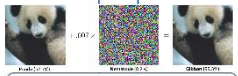
DOD Directive 3000.09
Autonomy in Weapon Systems: Nov 2012



Generative Adversarial Networks (GANs)
(Goodfellow: June 2014)



CNN 'Blind Spots'
1st *Generative Attack*
(Goodfellow: Dec 2014)



CNN beats humans at ImageNet
(Microsoft: Feb 2015)

'Autopilot' (Tesla: Sep 2015)

AI Community calls ban of all offensive AWs
(July 2015)

Defense Innovation Unit Experimental
(DIUx, Aug 2015)

AlphaGo beats Lee Sedol
(DeepMind: March 2016)

Project Maven
(April 2017)

AutoML Introduced
(Google: 2017)

***Single*–pixel Attacks**
(Su, Vargas: Oct 2017)

Capsule Networks
(Hinton: 2017)

AlphaGo Zero defeats AlphaGo
(DeepMind: Oct 2017)

AlphaZero* defeats AlphaGo *Zero
(DeepMind: Dec 2017)

Google employees resign over Maven
(May 2018)

Google decides not to renew Maven (June 2018)

Joint AI Center (JAIC)
(June 2018)

Learnability can be undecidable
(Ben–David: Jan 2019)

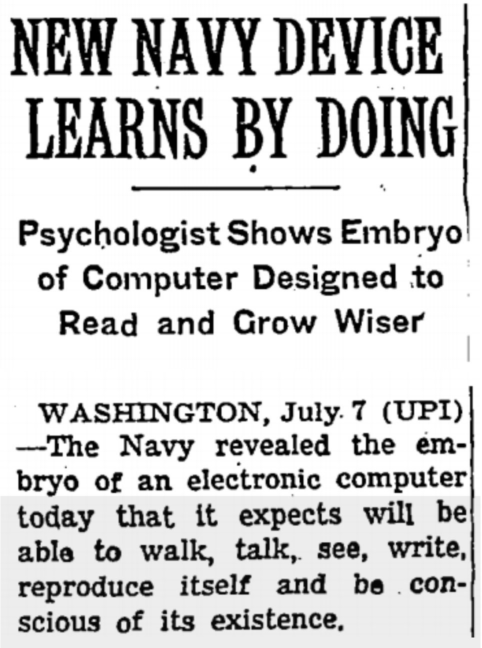
2005 2010 2015 2020

CNA

AI Milestones: *expectations and computational backdrop*

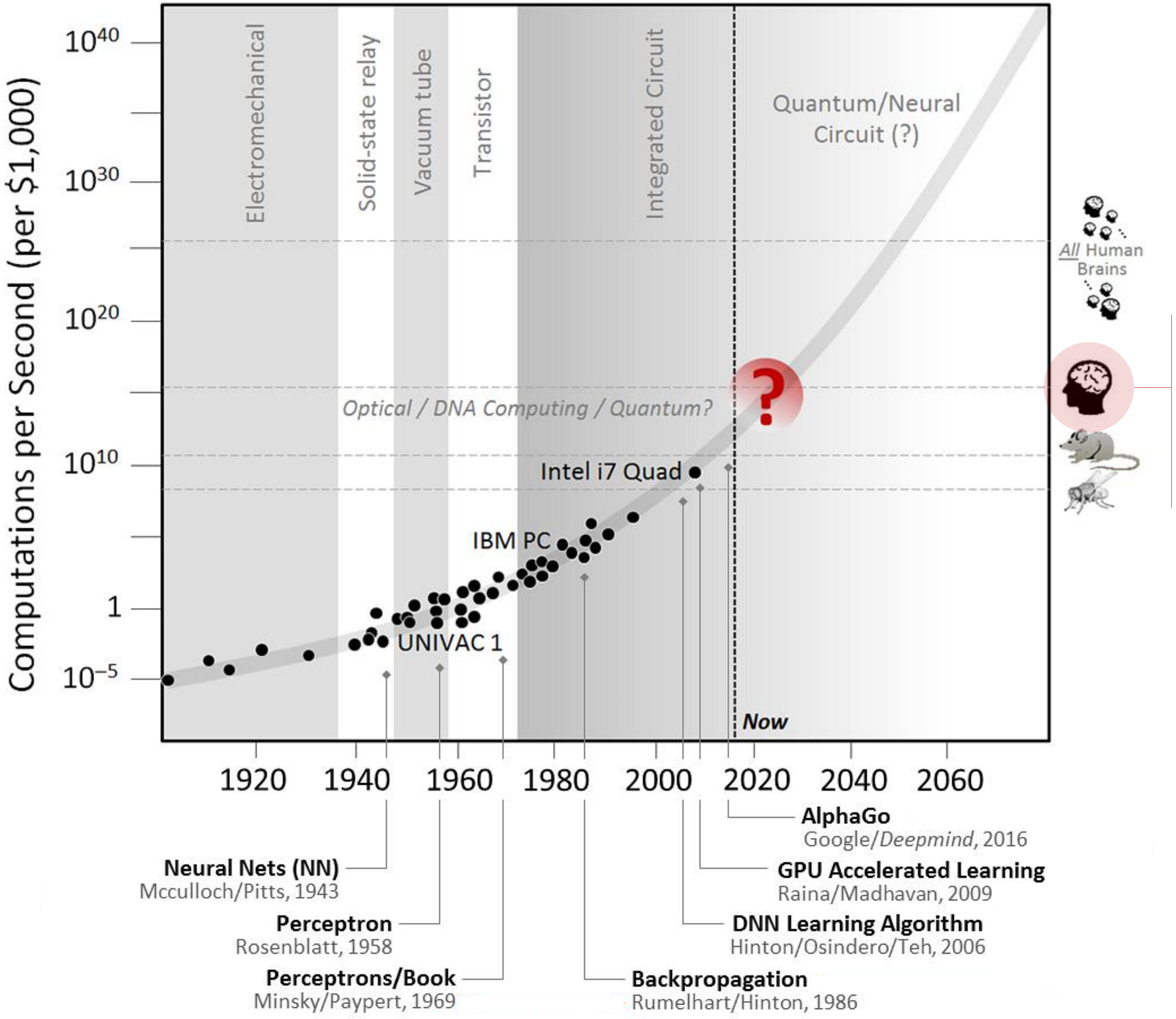
Perceptron / Rosenblatt

The New York Times
July 8, 1958


NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo of Computer Designed to Read and Grow Wiser

WASHINGTON, July 7 (UPI) —The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.



Exaflop performance (10^{18} *floating point operations / sec*) achieved in 2018!

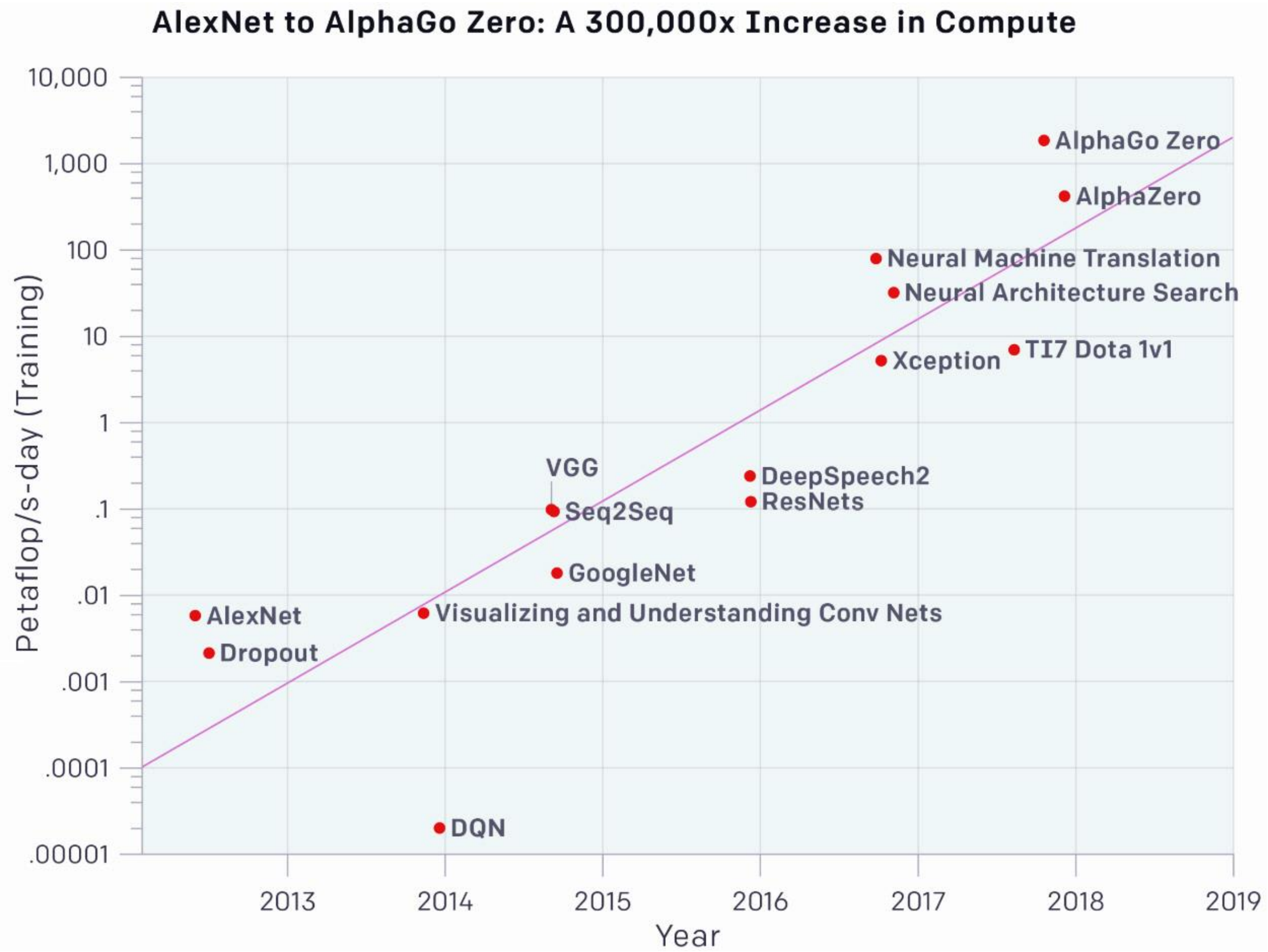
9,000 IBM POWER9 CPUs and 27,000 NVIDIA Tesla V100 Tensor Core GPU

https://upload.wikimedia.org/wikipedia/commons/d/df/PPTExponentialGrowthof_Computing.jpg

AI Computation

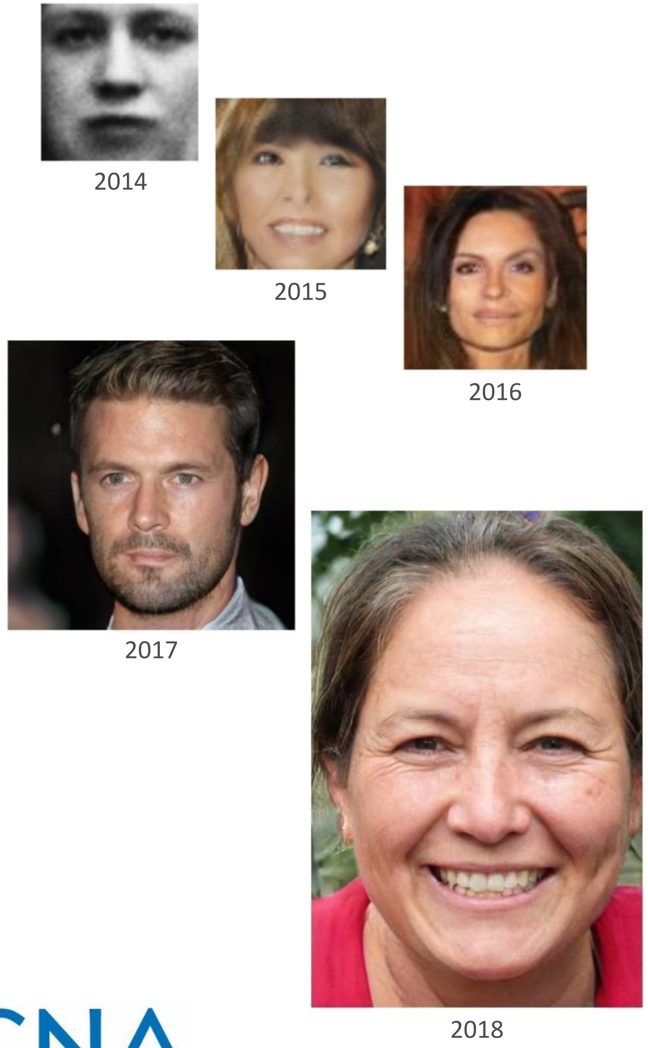
- (OpenAI) Since 2012, the amount of computation used in the largest AI training runs has *doubled every 3.5 months!*
  - Metric = petaflop/sec-days
  - Equal to 10^{15} neural net operations per second for one day, or a total of about 10^{20} total operations
- By comparison, *Moore's Law* has an 18-month doubling period

*AlexNet was a landmark 8-layer CNN (developed by Alex Krizhevsky) that was champion of the ImageNet *Large Scale Visual Recognition Challenge* in 2012



Some AI / ML "Hits" – *during 2018-2019*

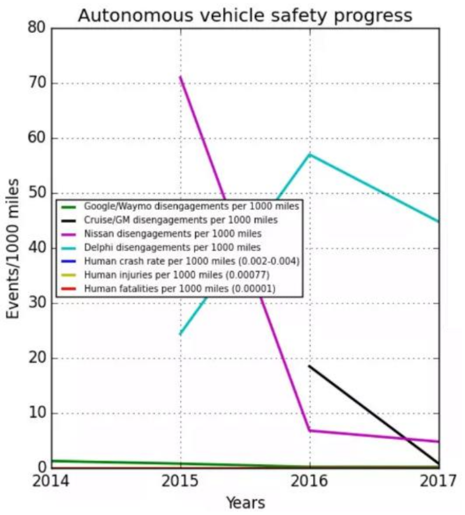
4.5 years of progress on face generation

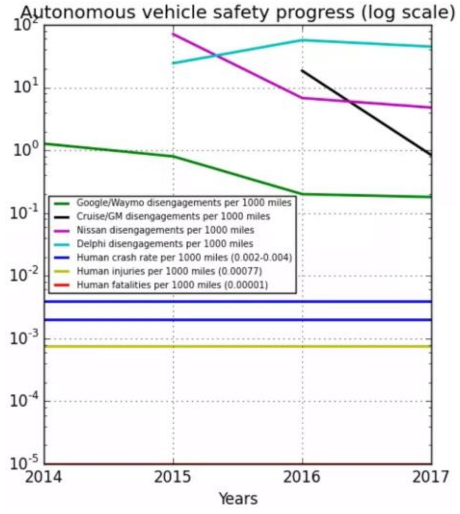


- Entire human chess knowledge learned and surpassed by *AlphaZero* in 4 hours
- AI "defeats" humans on a Stanford University reading comprehension test
- Deep Voice/Baidu develops algorithm to mimic voice with just snippets of audio
- Photorealistic speech-driven facial reenactment
- AI system finds correct sequence of steps to synthesize complex organic molecules (a task much more complex than the game of Go)
- Deep learning convolutional NN outperforms human cardiologists in a task involving heart scans
- "Cocktail problem" – AI learns to pick out individual voices in noisy crowd
- ML replicates chaotic attractors and calculates Lyapunov exponents from data
- AI system is trained to "see" in extreme low-light conditions
- AI learns to sense people's postures and movement through walls with WiFi
- ML recreates periodic table of elements in hours
- One-shot self-supervised imitation learning achieves human-level play (on difficult exploration games *Montezuma's Revenge*, *Pitfall!*, and *Private Eye*)
- 99.95th percentile ranked human *Dota2* team lost 2/3 matches to *OpenAI Five*

CNA

Some AI / ML “Misses” – *during 2018-2019*



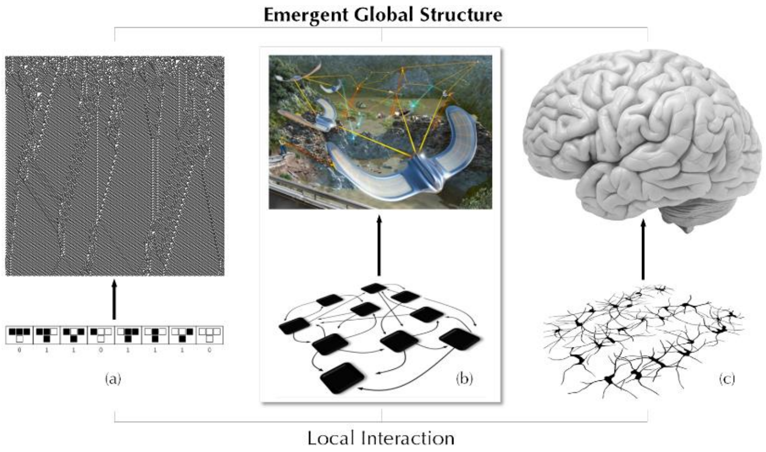


Filip Piekniewski, Principal AI Scientist at *Koh Young Technology* (South Korea); https://blog.piekniewski.info/2018/02/09/

- AI learns to associate colon cancer patients with specific clinics to which they were sent rather than the actual cancer (i.e., bias in electronic medical records)
- Less than stellar performance by an AI-controlled drone pitted against human pilot
- Tesla on autopilot crashes into a Laguna Beach police patrol vehicle
- AI system to predict outcomes of chemical reactions falls short of 90% accuracy goal (achieved 80% even for small “proof-of-concept” caliber size of atoms set)
- Tesla ‘on Autopilot’ slams into parked fire truck on California freeway
- Facial recognition software wrongly identifies 28 lawmakers as crime suspects
- Google’s “Talk to Books” semantic-search offers little improvement over keywords
- ML methods are dominated by traditional statistical ones on a comparative test of forecasting performance (using standard time series benchmarks)
- The six most accurate methods are basic statistical methods, not ML
- IBM Watson reportedly recommended “unsafe and incorrect” cancer treatments
- AI “fails” to predict winner in FIFA World Cup 2018 (an example of *negative* hype?)
- Driverless Tesla kills autonomous robot
- OpenAI’s “not quite complete win” over Dota2

AI / ML – *persistent challenges*

> *"I'm trying to draw a distinction between a machine learning system that's a black box and an entire field that's become a black box.*
>
> *Without deep understanding of the basic tools needed to build and train new algorithms, researchers creating AIs resort to hearsay, like medieval alchemists."*
>
> – Ali Rahimi, *Google AI*
>
> Rahimi's assertion at NIPS 2017 received a 40 sec ovation



- Intensely data hungry
- "Devil in the details" level of development highly nontrivial
- Basic research concerns – e.g., *reproducibility*
- Inherently opaque – *understandability, explainability, emergence*
- Not well integrated with prior knowledge
- Limited "understanding" of context (that humans take for granted)
- Limited capacity for transfer (to other problems / domains)
- Does not easily distinguish causation from correlation
- Struggles with open-ended inference
- Difficulty with exploration games w/sparse rewards (RL methods)
- Lives best in *static* universes
- Only nascent development of *meta*-learning and *lifelong*-learning
- Fragility – vulnerable to attack and/or exploitation
- Fundamental limits on ability to anticipate emergent behaviors
- Deeply prone to the "hype machine"

AI / ML – *persistent hype*

- "Alibaba's AI software surpasses humans in reading test"
  - *News Asia* (Jan 2018)
- "Computers are better than humans at reading"
  - *CNN* (Jan 2018)
- Theory of Mind-net (ToM-net)
  - *Google, DeepMind* (Feb 2018)
- "Pretty sure Google's new talking AI just beat the Turing test"
  - *Engadget* (May 2018)
- "Maryland researchers say they discovered 'Holy Grail' of machine learning"
  - *Washington Times* (May 2018)
- "Scientists Have Invented a Software That Can 'See' Several Minutes Into The Future"
  - *ScienceAlert* (June 2018)
- "A team of AI algorithms just crushed humans in a complex computer game"
  - *Technology Review* (June 2018)
- "IBM's AI Wins Debate with Human – *twice*"
  - *Big Think* (June 2018)
- "When bots teach themselves to cheat"
  - *Wired* (August 2018)
- "Robot 'talks' to MPs about future of AI in classroom"
  - *BBC News* (October 2018)
- "This clever AI hid data from its creators to cheat at its appointed task"
  - *TechCrunch* (Dec 2018)
- 'Hi-tech robot' on Russian state television turns out to be man in suit
  - *Oddity Central* (Dec 2018)

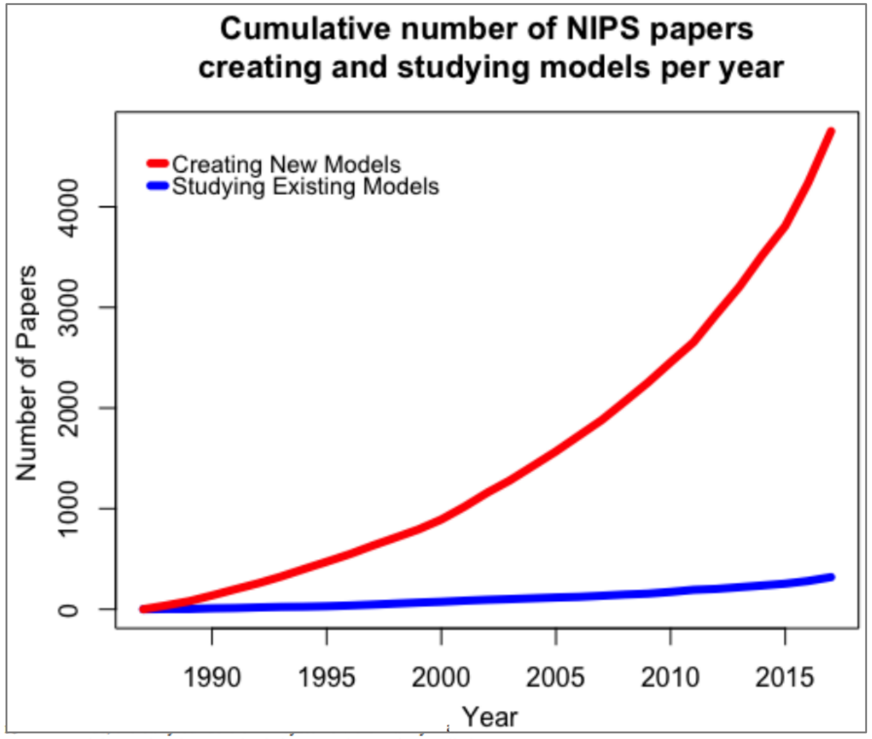
Snapshots of AI as a burgeoning “science” (1/3)

Closing the AI Knowledge Gap

Ziv Epstein*, Blakeley H. Payne*, Judy Hanwen Shen, Abhimanyu Dubey, Bjarke Felbo, Matthew Groh, Nick Obradovich, Manuel Cebrian, Iyad Rahwan
Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
Correspondence: {cebrian, irahwan}@mit.edu

Abstract

AI researchers employ not only the scientific method, but also methodology from mathematics and engineering. However, the use of the scientific method – specifically hypothesis testing – in AI is typically conducted in service of engineering objectives. Growing interest in topics such as fairness and algorithmic bias show that engineering-focused questions only comprise a subset of the important questions about AI systems. This results in the *AI Knowledge Gap*: the number of unique AI systems grows faster than the number of studies that characterize these systems' behavior. To close this gap, we argue that the study of AI could benefit from the greater inclusion of researchers who are well positioned to formulate and test hypotheses about the behavior of AI systems. We examine the barriers preventing social and behavioral scientists from conducting such studies. Our diagnosis suggests that accelerating the scientific study of AI systems requires new incentives for academia and industry, mediated by new tools and institutions. To address these needs, we propose a two-sided marketplace called TuringBox. On one side, *AI contributors* upload existing and novel algorithms to be studied scientifically by others. On the other side, *AI examiners* develop and post machine intelligence tasks designed to evaluate and characterize algorithmic behavior. We discuss this market's potential to democratize the scientific study of AI behavior, and thus narrow the AI Knowledge Gap.



1 The Many Facets of AI Research

Although AI is a sub-discipline of computer science, AI researchers do not exclusively use the scientific method in their work. For example, the methods used by early AI researchers often drew from logic, a subfield of mathematics, and are distinct from the scientific method we think of today. Indeed AI has adopted many techniques and approaches over time. In this section, we distinguish and explore the history of these

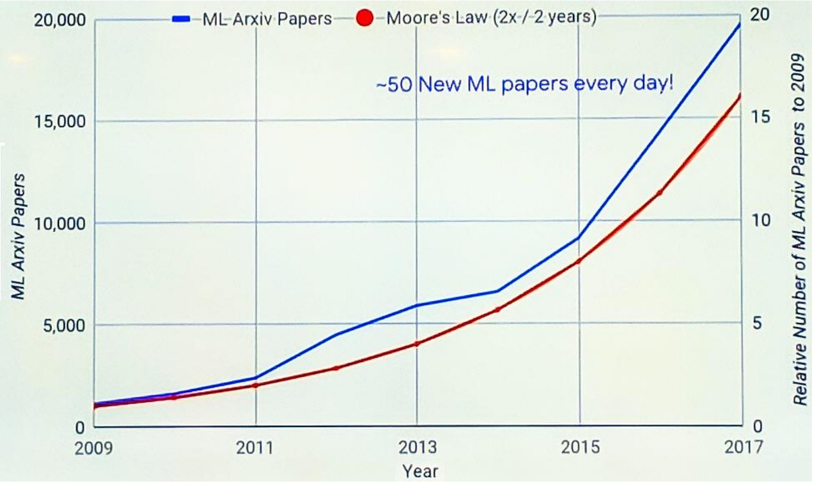
ematicians, devising mechanistic procedures–often called proof theories–for all manner of reasoning. In 1955, Herbert Simon and Allen Newell's *Logic Theorist* proved 38 theorems in the *Principia Mathematica* [Newell *et al.*, 1959]. This led Simon to claim that they had “solved the mind-body problem.” He argued that with a sufficiently powerful version of the Logic Theorist, we could automate mathematical reasoning, which in turn would enable the automation of all

*Equal contribution.

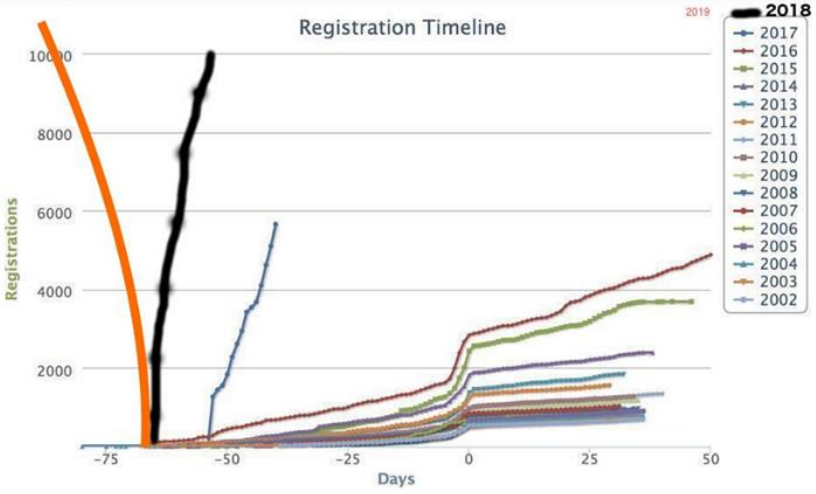
https://arxiv.org/abs/1803.07233

ML Arxiv papers per year

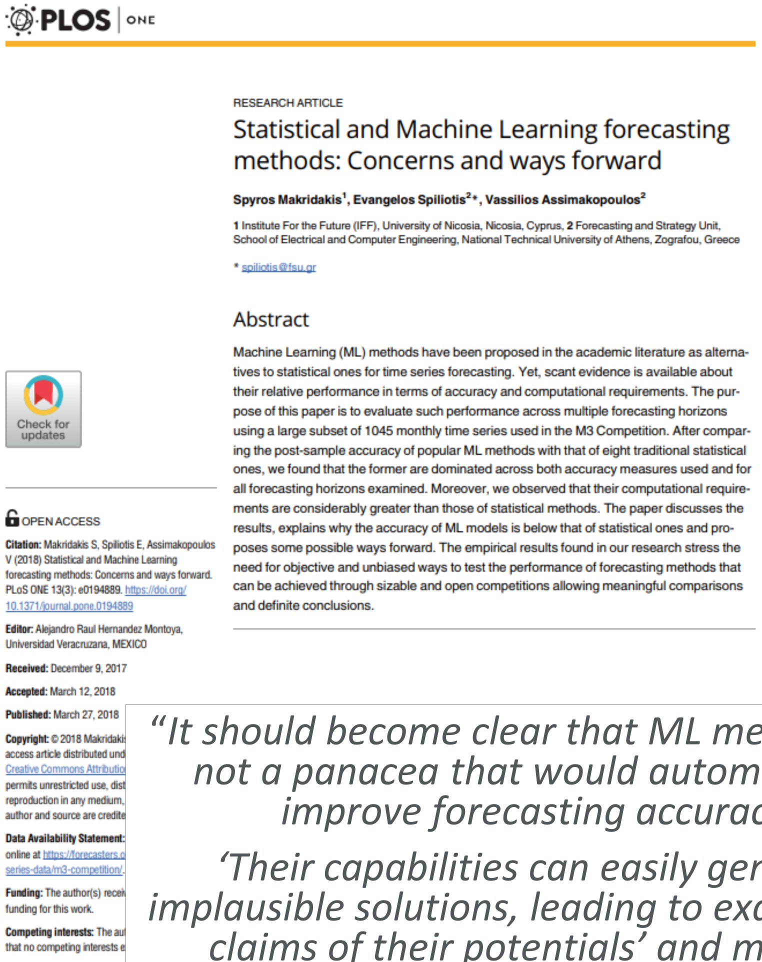
(Compiled by Zak Stone, product manager at *TensorFlow*)



Neural Information Processing Systems **(NIPS) 2018 conference sold out in 11 min 38 sec!**



Snapshots of AI as a burgeoning "science" (2/3)



PLOS | ONE

RESEARCH ARTICLE

Statistical and Machine Learning forecasting methods: Concerns and ways forward

Spyros Makridakis[1], Evangelos Spiliotis[2]*, Vassilios Assimakopoulos[2]

1 Institute For the Future (IFF), University of Nicosia, Nicosia, Cyprus, 2 Forecasting and Strategy Unit, School of Electrical and Computer Engineering, National Technical University of Athens, Zografou, Greece

* spiliotis@fsu.gr

Abstract

Machine Learning (ML) methods have been proposed in the academic literature as alternatives to statistical ones for time series forecasting. Yet, scant evidence is available about their relative performance in terms of accuracy and computational requirements. The purpose of this paper is to evaluate such performance across multiple forecasting horizons using a large subset of 1045 monthly time series used in the M3 Competition. After comparing the post-sample accuracy of popular ML methods with that of eight traditional statistical ones, we found that the former are dominated across both accuracy measures used and for all forecasting horizons examined. Moreover, we observed that their computational requirements are considerably greater than those of statistical methods. The paper discusses the results, explains why the accuracy of ML models is below that of statistical ones and proposes some possible ways forward. The empirical results found in our research stress the need for objective and unbiased ways to test the performance of forecasting methods that can be achieved through sizable and open competitions allowing meaningful comparisons and definite conclusions.

Check for updates

OPEN ACCESS

Citation: Makridakis S, Spiliotis E, Assimakopoulos V (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 13(3): e0194889. https://doi.org/10.1371/journal.pone.0194889

Editor: Alejandro Raul Hernandez Montoya, Universidad Veracruzana, MEXICO

Received: December 9, 2017

Accepted: March 12, 2018

Published: March 27, 2018

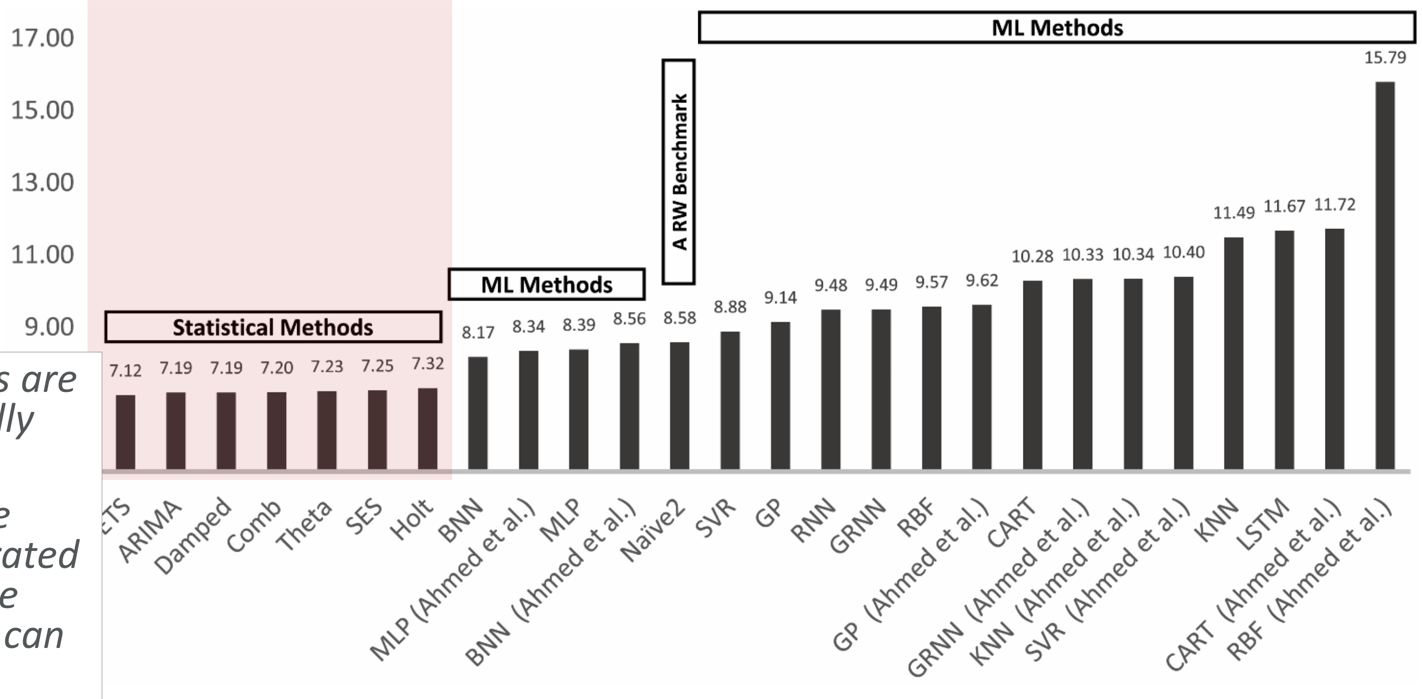
Data Availability Statement:
online at https://forecasters.o
series-data/m3-competition/.

Funding: The author(s) recei
funding for this work.

Competing interests: The au
that no competing interests e

- Paper evaluates such performance across multiple forecasting horizons using a large subset of 1045 monthly time series used in M3 Competition
- ***The six most accurate methods are basic statistical methods, not ML***

Symmetric Mean Absolute Percentage Error (sMAPE)

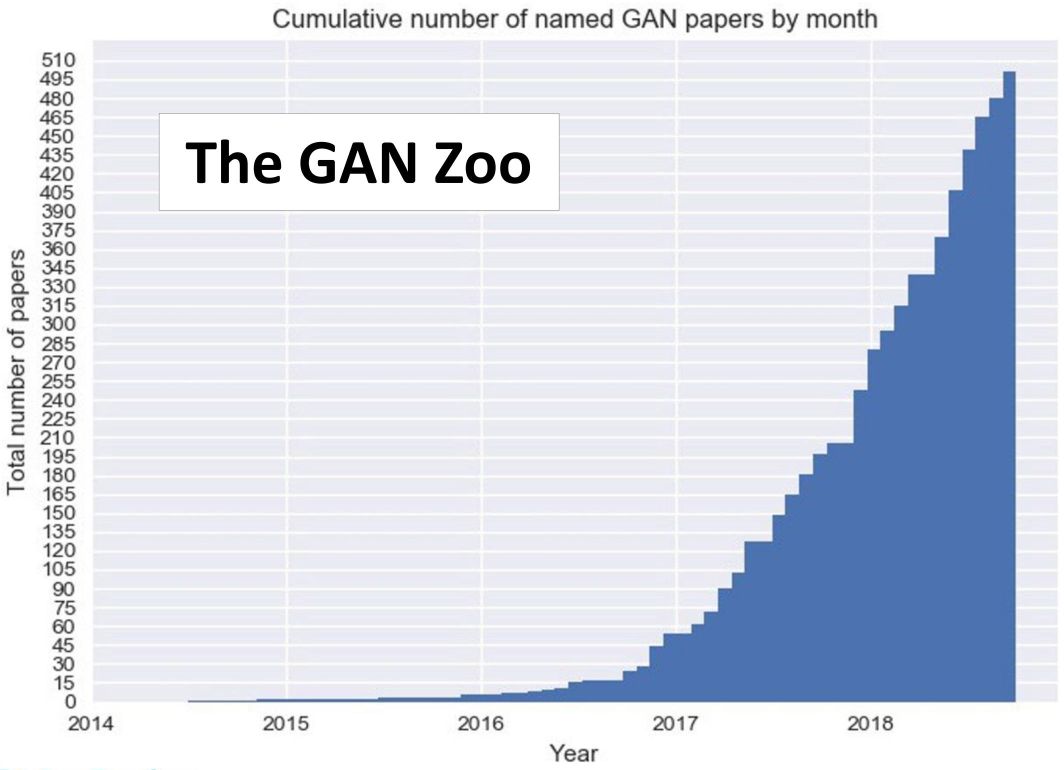


"It should become clear that ML methods are not a panacea that would automatically improve forecasting accuracy.

'Their capabilities can easily generate implausible solutions, leading to exaggerated claims of their potentials' and must be carefully investigated before any claims can be accepted."

Snapshots of AI as a burgeoning “science” (3/3)

- GAN = *Generative Adversarial Network*, introduced by Ian Goodfellow (et.al.) in 2014
- Consist of two competing networks: a *generator* (**G**) and a *discriminator* (**D**)
  - **G** tries to create random synthetic outputs (e.g., images of faces)
  - **D** tries to tell these apart from real outputs (e.g., a database of celebrities)
- As G and D “compete,” they both get better and better
- The result is a generator network that produces realistic outputs

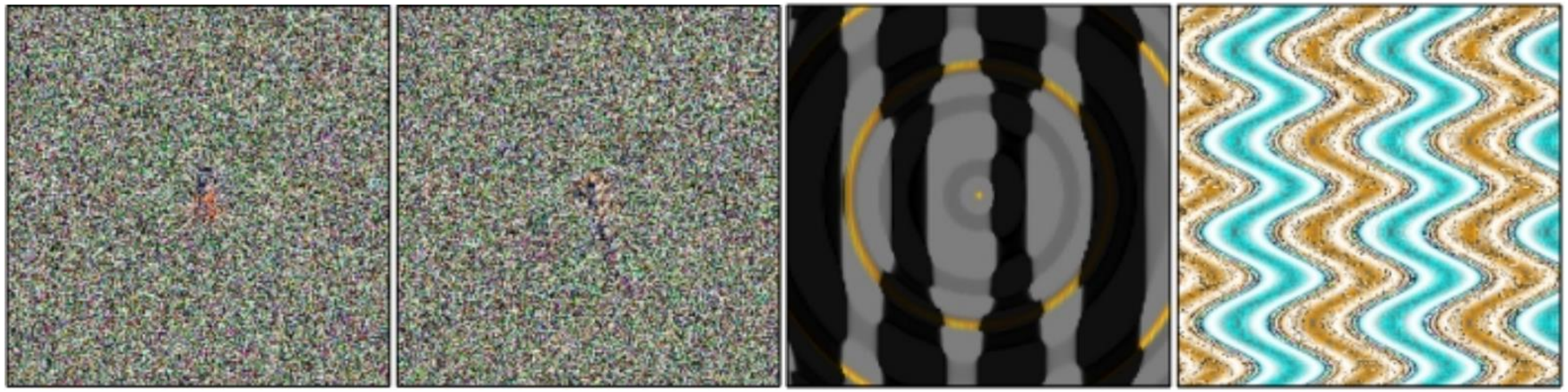


- 3D-ED-GAN — Shape Inpainting using 3D Generative Adversarial Network and Recurrent Convolutional Networks
- 3D-GAN — Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling
- 3D-IWGAN — Improved Adversarial Systems for 3D Object Generation and Reconstruction
- 3D-PhysNet — 3D-PhysNet: Learning the Intuitive Physics of Non-Rigid Object Deformations
- 3D-RecGAN — 3D Object Reconstruction from a Single Depth View with Adversarial Learning
- ABC-GAN — ABC-GAN: Adaptive Blur and Control for improved training stability of Generative Adversarial Networks
- ABC-GAN — GANs for LIFE: Generative Adversarial Networks for Likelihood Free Inference
- AC-GAN — Conditional Image Synthesis With Auxiliary Classifier GANs
- acGAN — Face Aging With Conditional Generative Adversarial Networks
- ACGAN — Coverless Information Hiding Based on Generative adversarial networks
- acGAN — On-line Adaptive Curriculum Learning for GANs
- ACtuAL — ACtuAL: Actor-Critic Under Adversarial Learning
- AdaGAN — AdaGAN: Boosting Generative Models
- Adaptive GAN — Customizing an Adversarial Example Generator with Class-Conditional GANs
- AdvEntuRe — AdvEntuRe: Adversarial Training for Textual Entailment with Knowledge-Guided Examples
- AdvGAN — Generating adversarial examples with adversarial networks
- AE-GAN — AE-GAN: adversarial eliminating with GAN
- AE-OT — Latent Space Optimal Transport for Generative Models
- AEGAN — Learning Inverse Mapping by Autoencoder based Generative Adversarial Nets
- AF-DCGAN : Amplitude Feature Deep Convolutional GAN for Fingerprint Construction in Indoor Localization System
- AffGAN — Amortised MAP Inference for Image Super-resolution
- AIM — Generating Informative and Diverse Conversational Responses via Adversarial Information Maximization
- AL-CGAN — Learning to Generate Images of Outdoor Scenes from Attributes and Semantic Layouts
- ALI — Adversarially Learned Inference
- AlignGAN — AlignGAN: Learning to Align Cross-Domain Images with Conditional Generative Adversarial Networks
- AlphaGAN — AlphaGAN: Generative adversarial networks for natural image matting
- AM-GAN — Activation Maximization Generative Adversarial Nets
- AmbientGAN — AmbientGAN: Generative models from lossy measurements

- ***…the list goes on – and on, and on - for about 250 entries!***

Avinash Hindupur, *The GAN Zoo*, https://deephunt.in/the-gan-zoo-79597dc8c347

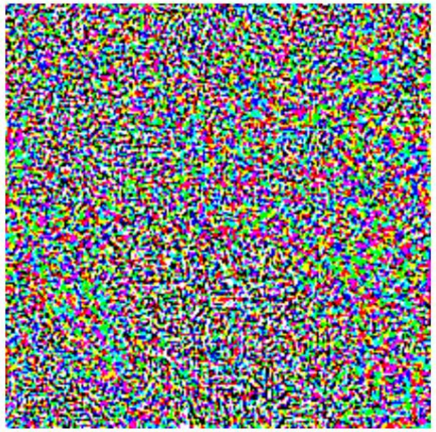
When NNs don't work, they can be *unpredictably bad!*



Robin **Cheetah** **King penguin** **Starfish**

A. Nguyen, J. Yosinski, J. Clune, "Deep Neural Networks are Easily Fooled:
High Confidence Predictions for Unrecognizable Images," *Comp. Vision and Pattern Rec.* IEEE (2015)

$+ .007 \times$



$=$

Panda (57.7%) **Nematode** (8.2%) **Gibbon** (99.3%)

I. Goodfellow, J. Shlens, C. Szegedy, "Explaining and harnessing adversarial examples," Intern. Conf. on Learning Representations (ICLR) 2015

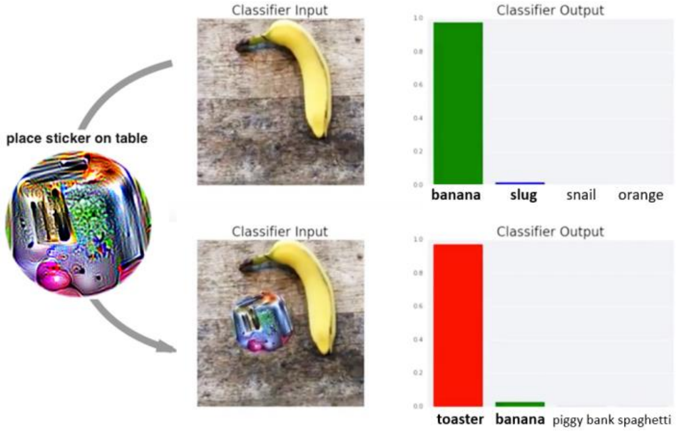
When NNs don’t work, they can be *unpredictably bad!*

Single-Pixel Attacks

Ship → **Car** (57.7%)   **Nematode** → **Frog** (99.9%)   **Gibbon** → **Airplane** (99.3%)

Jiawei Su, et.al., “One pixel attack for fooling deep neural networks,” arXiv:1710.08864

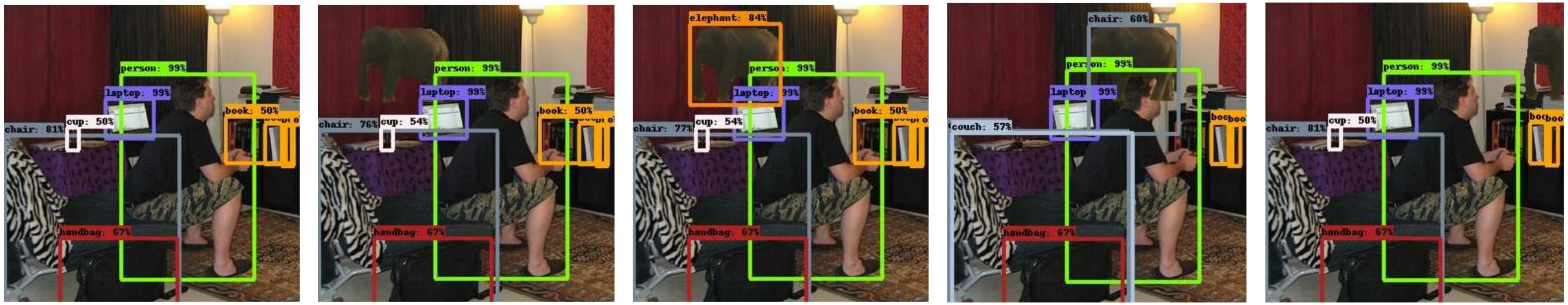
Adversarial Patch Attacks



A real-world attack on VGG16, using a physical patch generated by the white-box ensemble method described in Section 3. When a photo of a tabletop with a banana and a notebook (top photograph) is passed through VGG16, the network reports class ’banana’ with 97% confidence (top plot). If we physically place a sticker targeted to the class "toaster" on the table (bottom photograph), the photograph is classified as a toaster with 99% confidence (bottom plot).

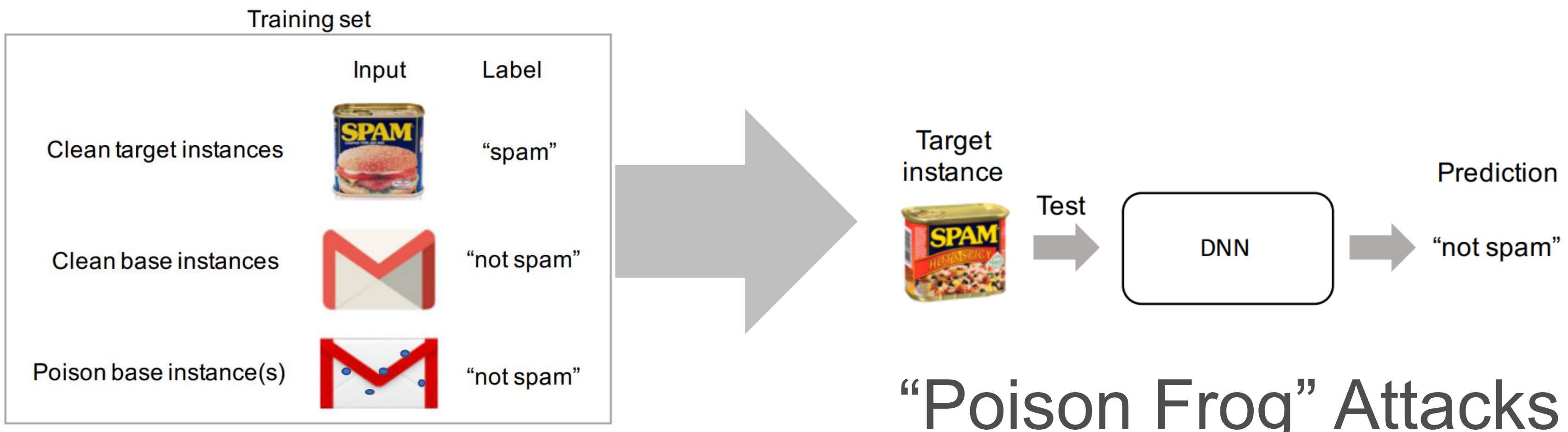
Tom B. Brown, et.al., “Adversarial patch,” arXiv:1712.09665v2 [cs.CV]

The Elephant in the Room

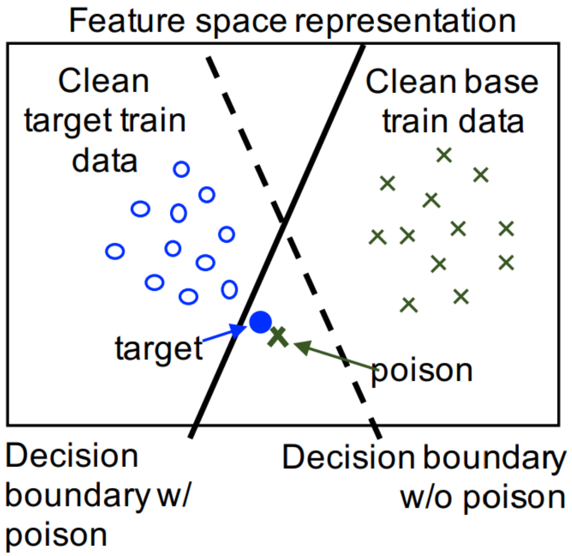


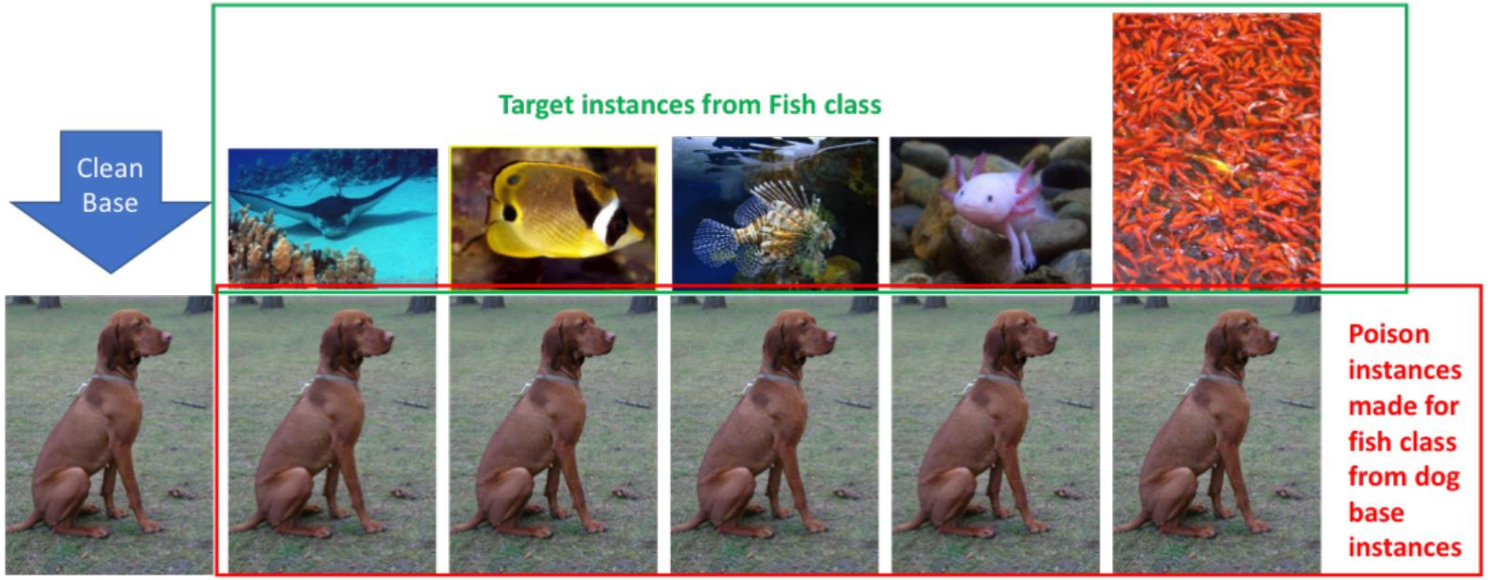
Amir Rosenfeld, Richard Zemel, John K. Tsotsos, “The Elephant in the Room,” https://arxiv.org/abs/1808.03305 [cs.CV]

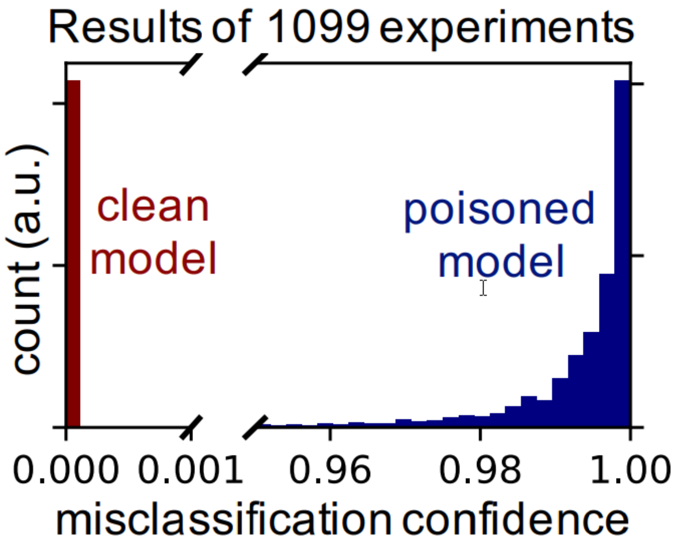
When NNs don’t work, they can be *unpredictably bad!*



“Poison Frog” Attacks



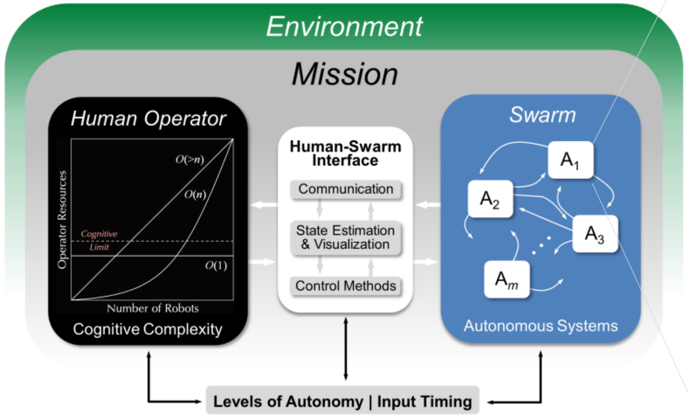


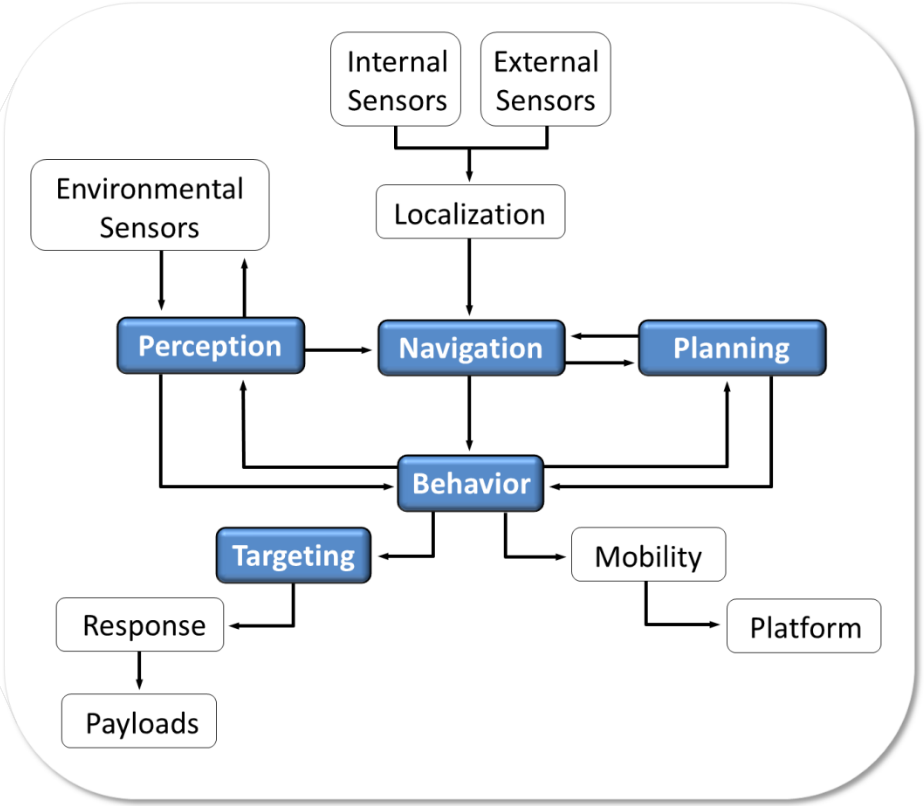


Ali Shafahi, et.al., “Poison Frogs! Targeted Clean-Label Poisoning Attacks on Neural Networks,” https://arxiv.org/abs/1804.00792v1

It's not just a "single AI solution"

Key functional components and relationships of an autonomous unmanned system, including elements that describe human–machine interaction / collaboration

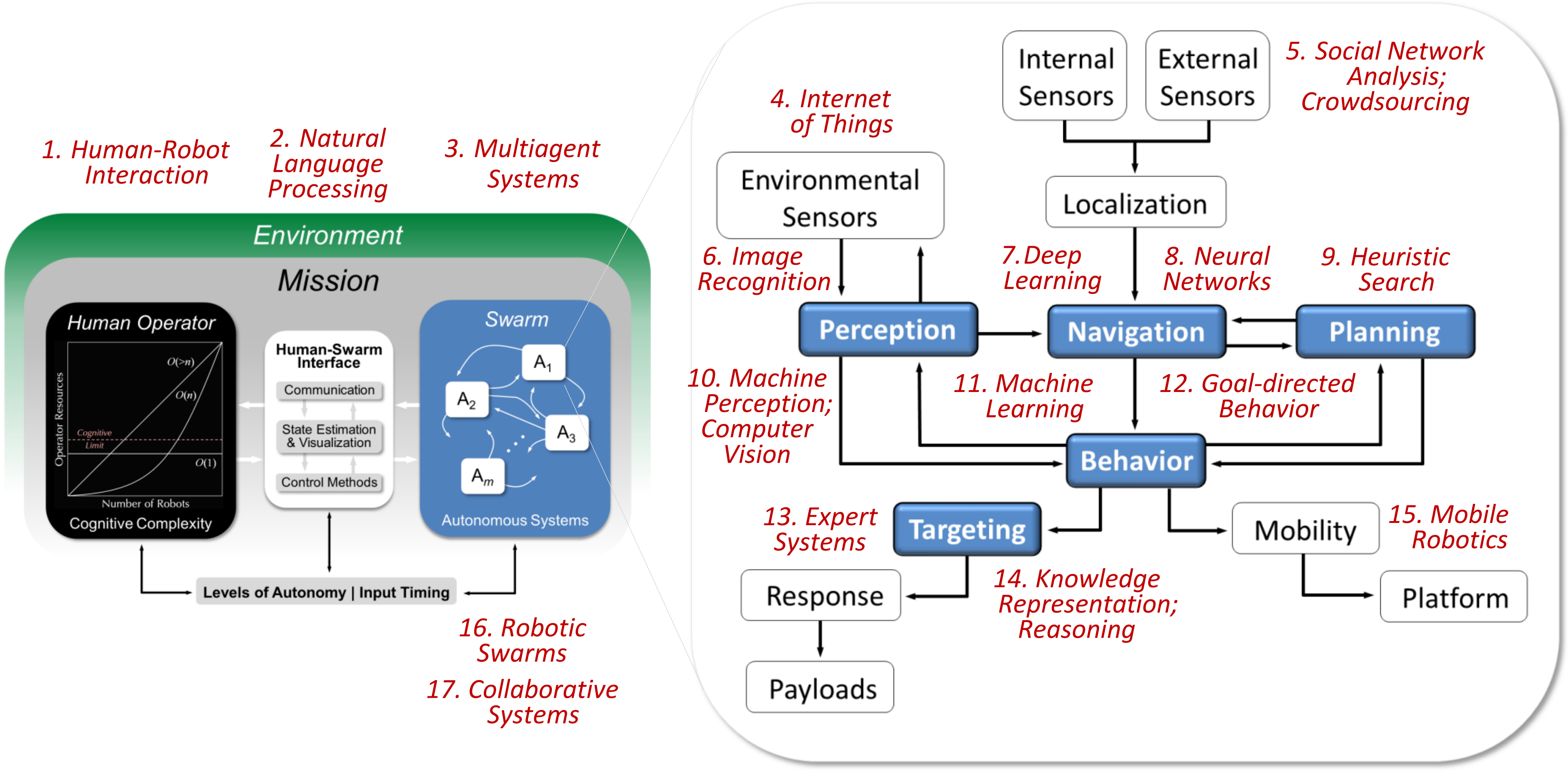




After figure 3.1 in A. Finn and S. Scheding, Developments and Challenges for Autonomous Unmanned Vehicles, Springer-Verlag, 2010

It’s not just a “single AI solution”

Each component may be associated with (a *set of entwined*) AI methods



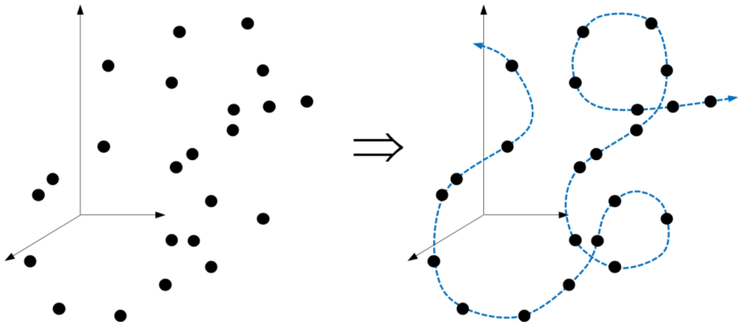
1. Human-Robot Interaction
2. Natural Language Processing
3. Multiagent Systems
4. Internet of Things
5. Social Network Analysis; Crowdsourcing
6. Image Recognition
7. Deep Learning
8. Neural Networks
9. Heuristic Search
10. Machine Perception; Computer Vision
11. Machine Learning
12. Goal-directed Behavior
13. Expert Systems
14. Knowledge Representation; Reasoning
15. Mobile Robotics
16. Robotic Swarms
17. Collaborative Systems

CNA

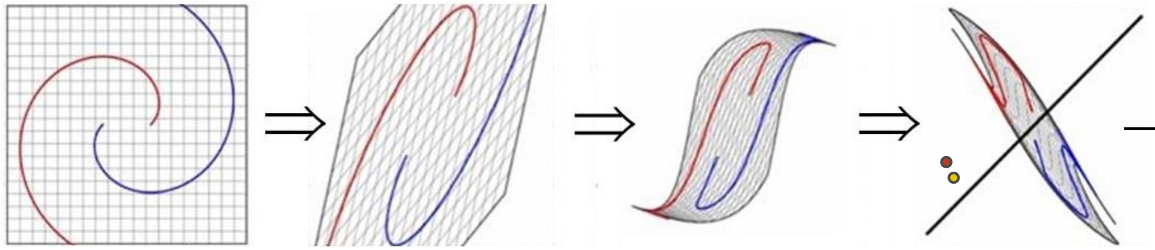
After figure 3.1 in A. Finn and S. Scheding, Developments and Challenges for Autonomous Unmanned Vehicles, Springer-Verlag, 2010

The Manifold Hypothesis

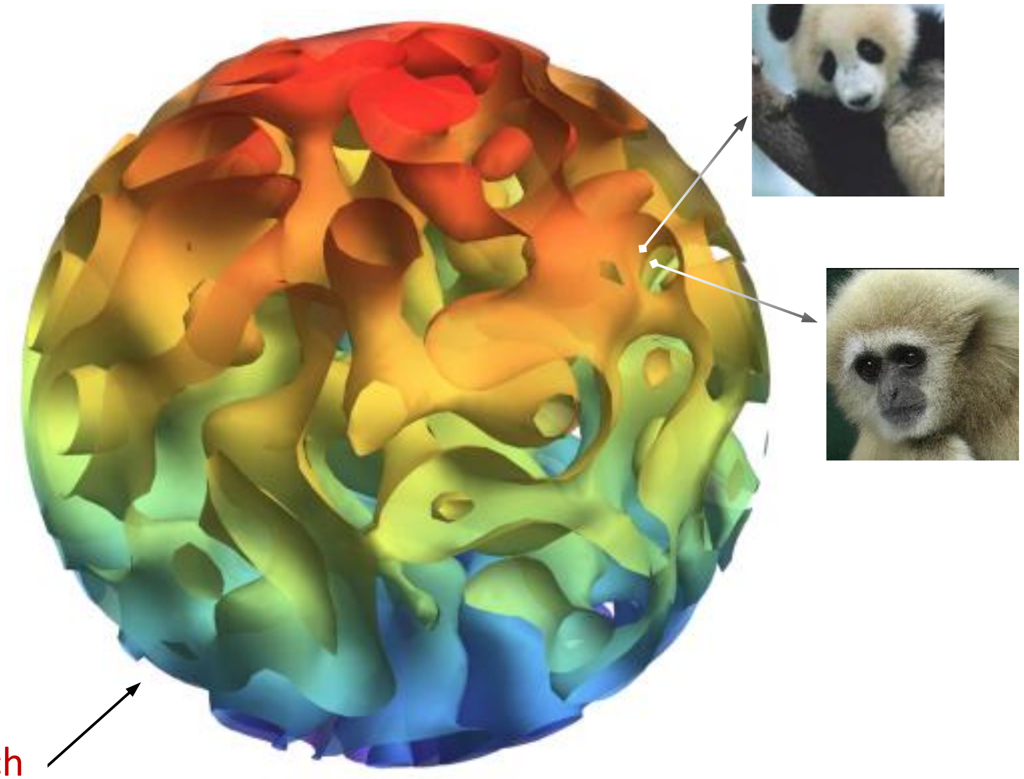
- Natural data forms lower dimensional structures (manifolds) in embedding space
  - *Each manifold represents a different entity*

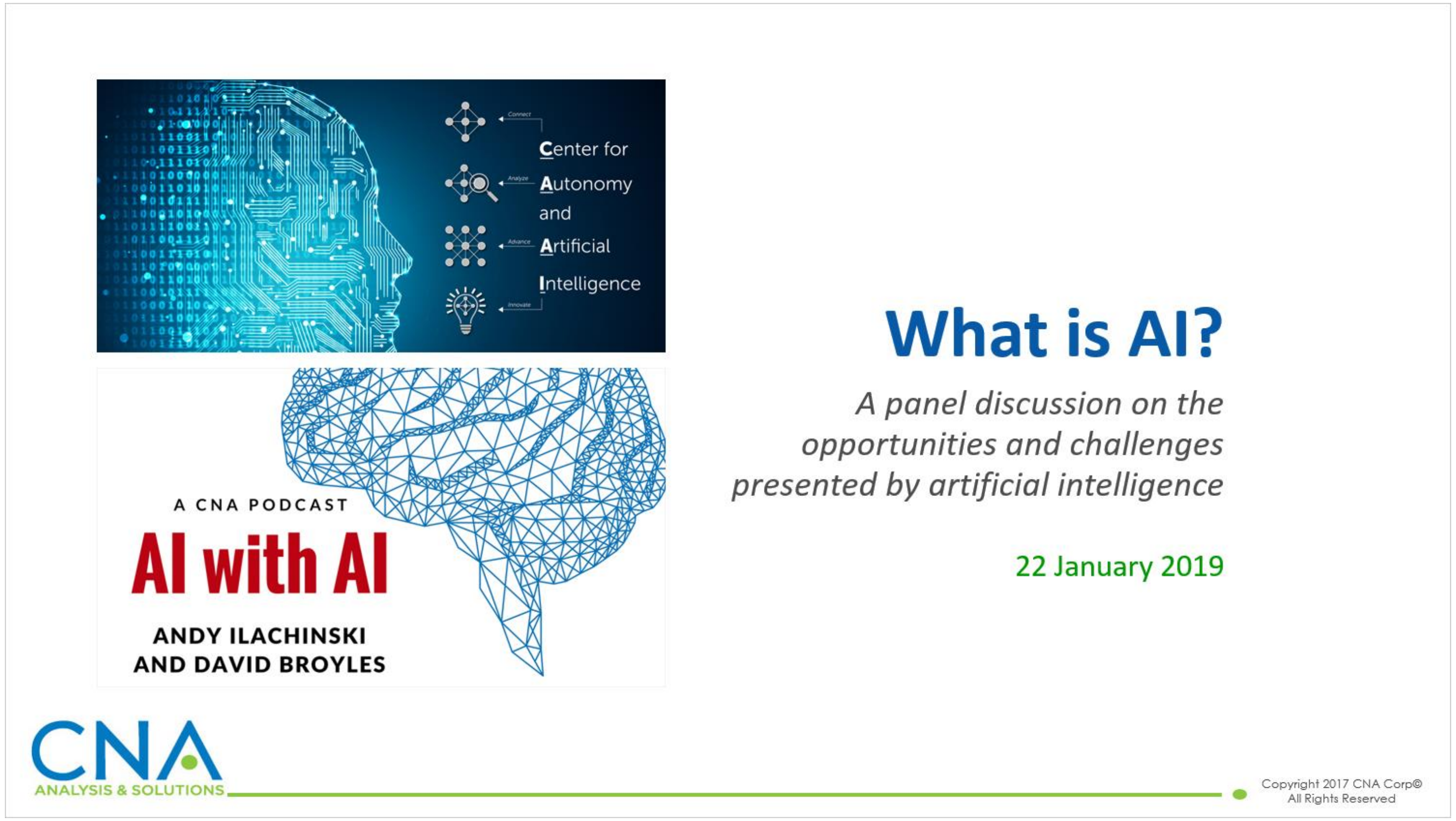


- Learning ("understanding" data) achieved by separating the manifolds
  - *Easy to do (and visualize) when D* = 2
    ("Stretching and squashing")



Much harder when *D* >> 1

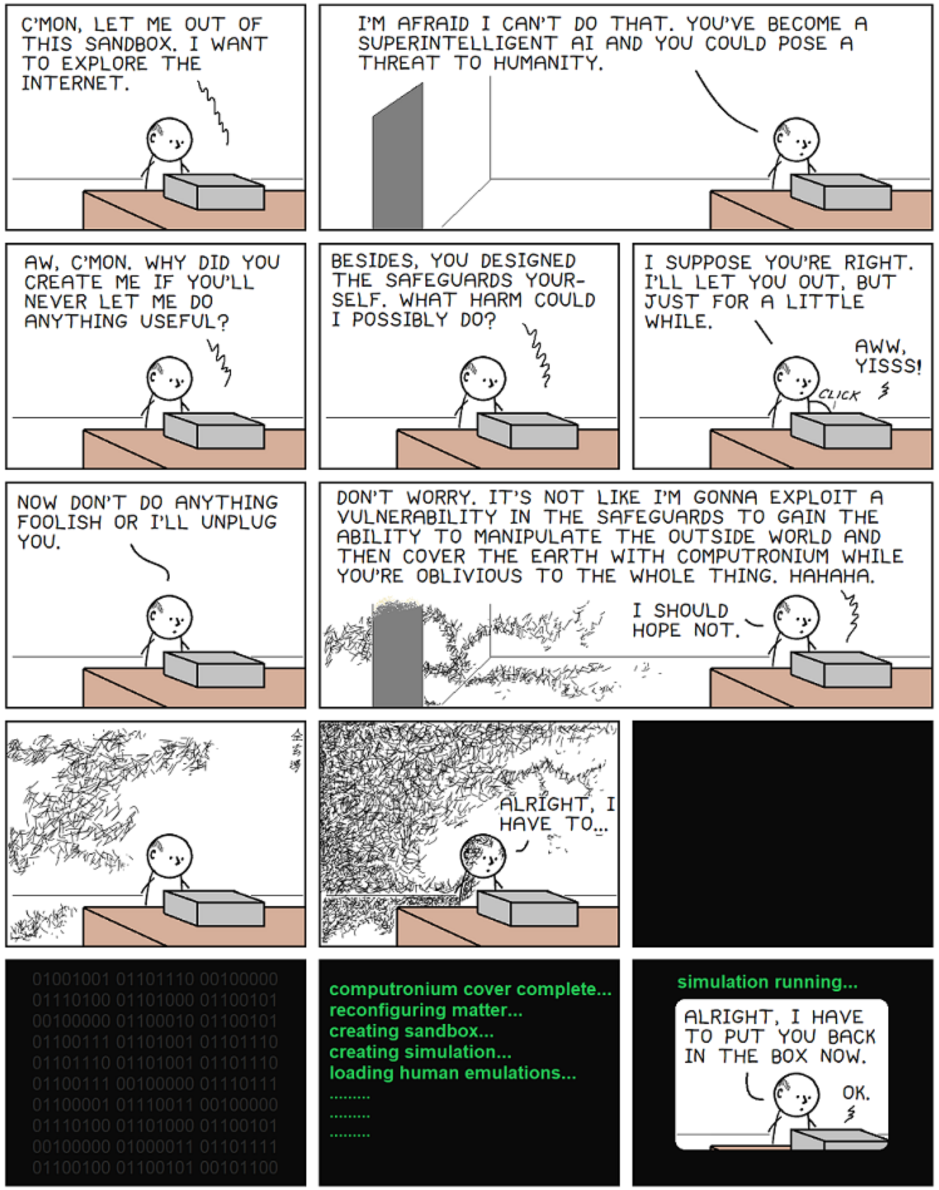



Center for
Autonomy
and
Artificial
Intelligence
A CNA PODCAST
AI with AI
ANDY ILACHINSKI
AND DAVID BROYLES
What is AI?
A panel discussion on the
opportunities and challenges
presented by artificial intelligence
22 January 2019
CNA
ANALYSIS & SOLUTIONS
Copyright 2017 CNA Corp©
All Rights Reserved

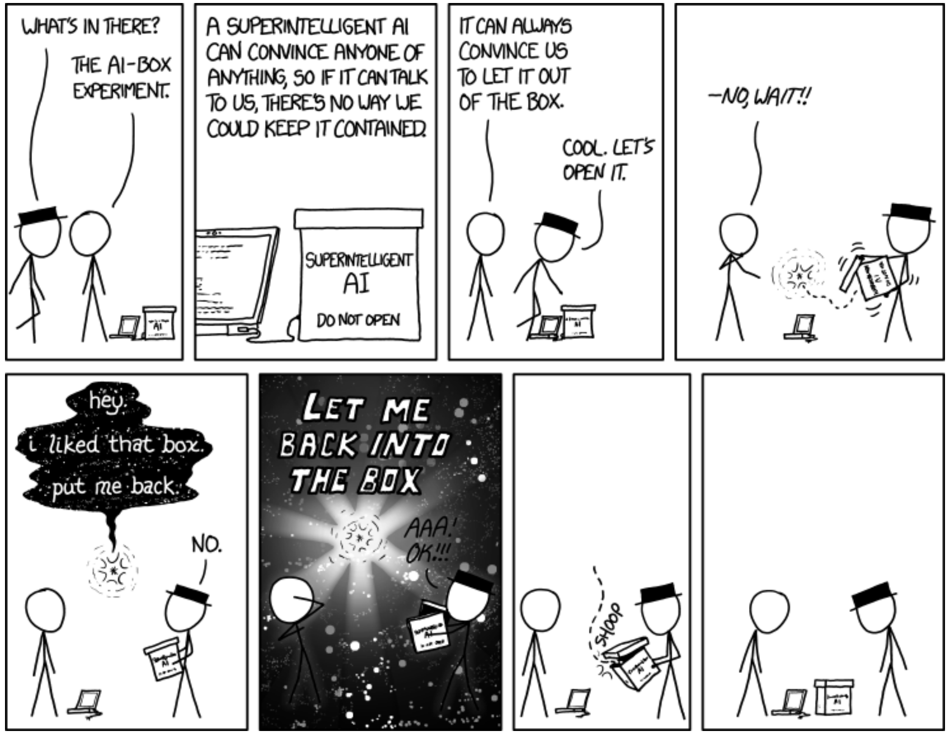
Questions ?

Comical Views of Machine Learning & *Super*-AI

Xkcd Webcomic, https://xkcd.com/1838/; *Creative Commons License*



Abtruse Goose, https://abstrusegoose.com/594; *Creative Commons License*



Xkcd Webcomic, https://xkcd.com/1450/;
Creative Commons License: https://creativecommons.org/licenses/by-nc/3.0/us/