# The Need for a National AI Research Infrastructure Initiative

Bart Selman

Cornell University

## The Emergence of Intelligent Machines: Challenges and Opportunities

# The Emergence of Artificial Intelligence

**I** ***Emergence of (semi-)intelligent autonomous systems in society***

**--- Self-driving cars and trucks. Autonomous drones.**
**Virtual assistants. Fully autonomous trading systems.**
**Assistive robotics. Real-time translation.**

**II** ***Shift of AI research from academic to real-world***

**--- Enabled by qualitative change in the field,**
**driven in part by "Deep Learning" & Big Data.**

# Reasons for Dramatic Progress

--- series of events

--- main one: *machine perception* is starting to work (finally!)

systems are starting to *"hear"* and *"see"*

after "only" 50+ yrs of research…

--- dramatic change: lots of AI techniques (reasoning, search, reinforcement learning, planning, decision theoretic methods) were developed assuming perceptual inputs were "somehow" provided to the system. But, e.g., robots could not really see or hear anything…

(e.g. *2005* Stanley car drove around *blind*; developers were told "don't bother putting in a camera" --- Thrun, Stanford)

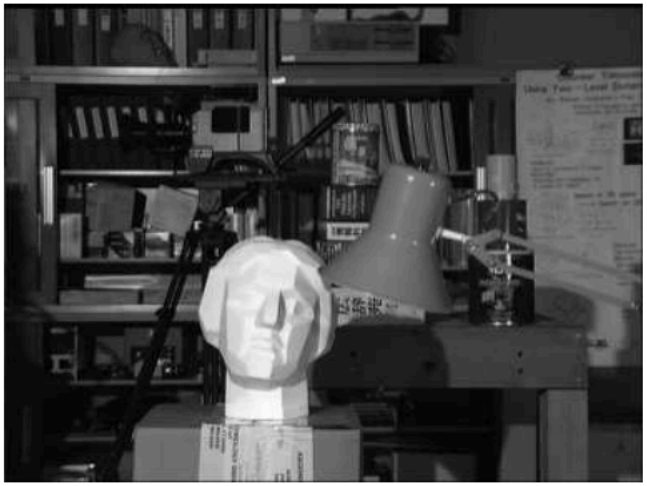
Now, we can use output from a perceptual system and leverage a broad range of existing AI techniques.

*Our systems are finally becoming "grounded in (our) world."*

Already: super-human face recognition (Facebook)

super-human traffic sign recognition (Nvidia)

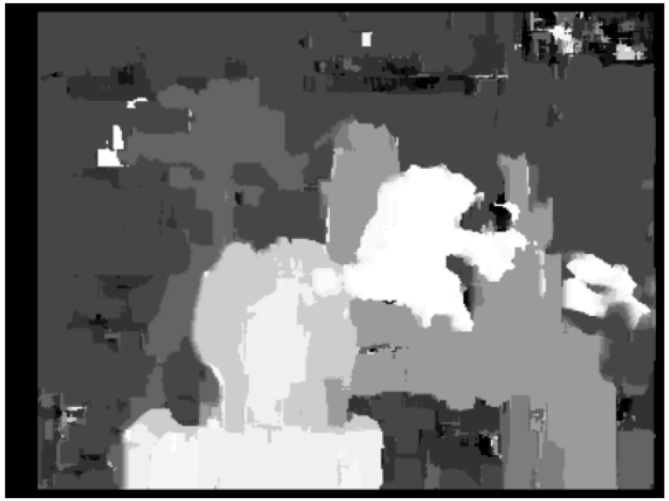
# Computer vision / Image Processing ca. 2005



(a) Left image: 384x288, 15 labels

(b) Ground truth

**(human labeled)**

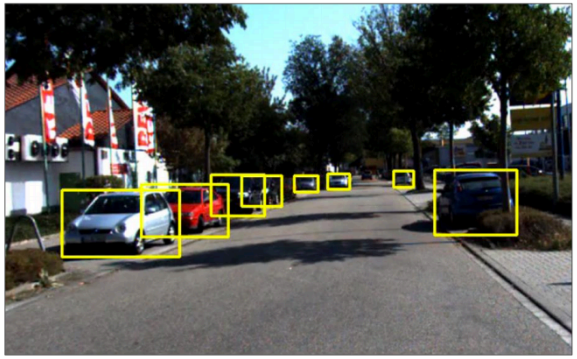


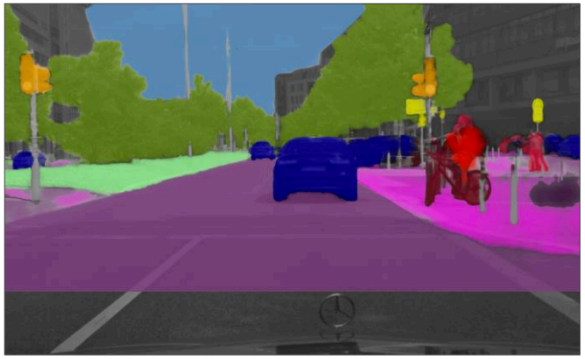
**(machine labeled)**

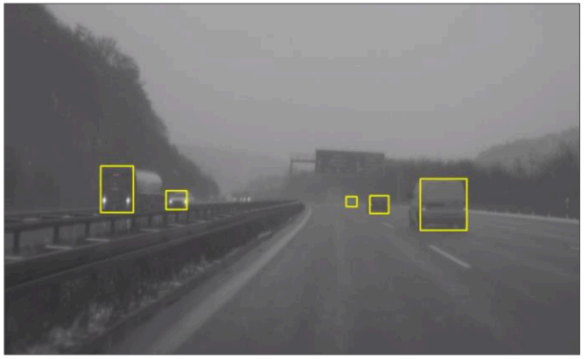
***2005 --- sigh ☹***

(c) Processed image

# DEEP LEARNING FOR SELF-DRIVING CARS







***Note labeling!***

**Statistical model (neural net) trained on >1M images;**
**Models with > 500K parameters**
**Requires GPU power**

**(Mobileye 2016; Nvidia 2016)**


Real-time tracking of environment (360 degrees/ 50+m) and decision making.

# Factors in accelerated progress, cont.

**--- deep learning / deep neural nets**

**success is evidence in support of the "hardware hypothesis"**

**(need to get near brain compute power; Moravec)**

***core neural net ideas from mid 1980s***

**needed: several orders of magnitude increase in computational power and data**

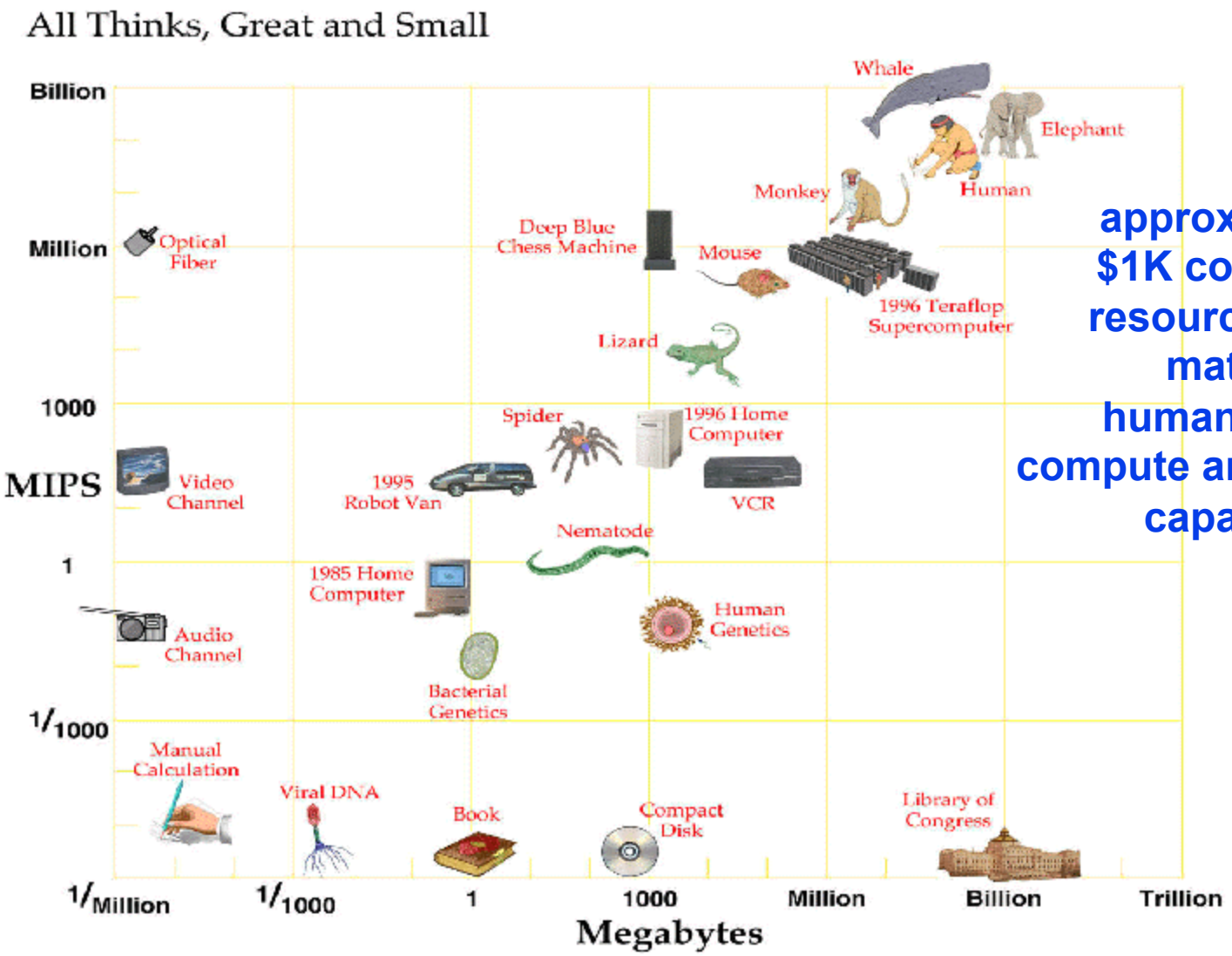
**Aside:**

**(1)** ***This advance was not anticipated/predicted at all.***
**by 2000, almost all AI/ML researchers had moved away from neural nets… changed around 2011/12.**

**(2) Algorithmic advances still provided larger part of speedups than hardware. Core algorithmic concept from 1980s but** ***key additional advances since*****.**

**+ BIG DATA!**

# Computer vs. Brain




**approx. 2030**
**$1K compute resources will match human brain compute and storage capacity**

# The unreasonable effectiveness of deep learning in artificial intelligence

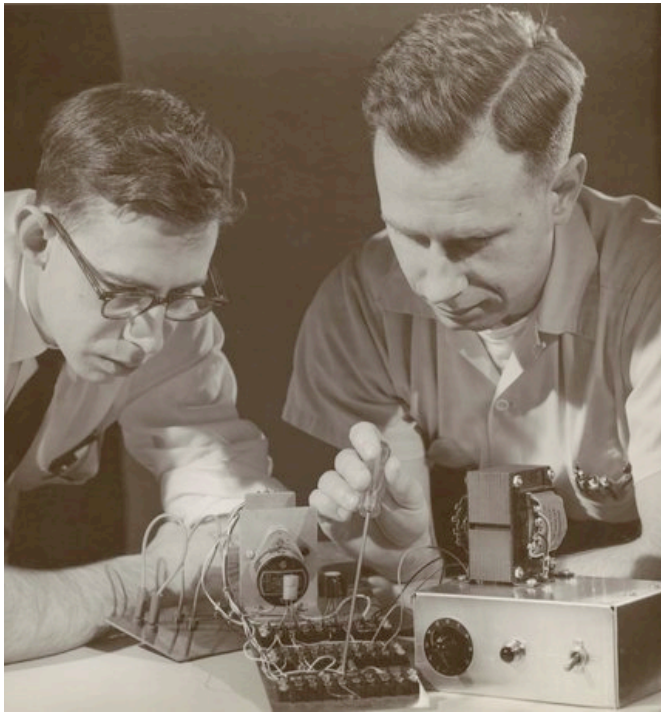
**Terrence J. Sejnowski**[a,b,1]

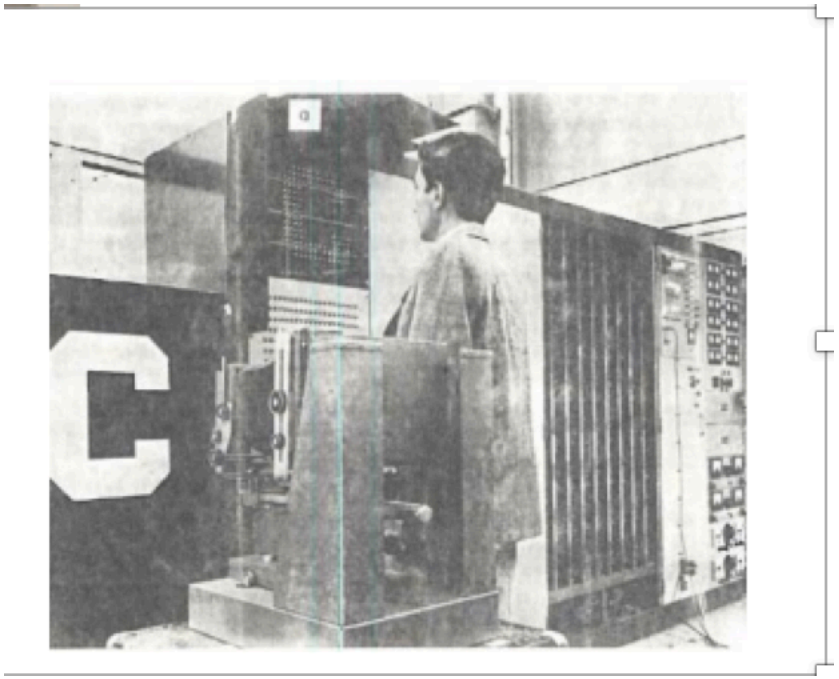
[a]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037; and [b]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved November 22, 2019 (received for review September 17, 2019)

Brains have 11 orders of magnitude of spatially structured computing components (Fig. 5). At the level of synapses, each cubic millimeter of the cerebral cortex, about the size of a rice grain, contains a billion synapses. The largest deep learning networks today are reaching a billion weights. The cortex has the equivalent power of hundreds of thousands of deep learning networks, each specialized for solving specific problems. How are all these expert networks organized? The levels of investigation

## Historical Aside: The first learning Artificial Neural Net was developed at Cornell.





**Rosenblatt (left), 1958**.

(unfortunately,

patent long expired…)

# Progress, cont.

**--- crowd-sourced human data --- *machines need to understand our conceptualization of the world.* E.g. vision for self driving cars trained on 100,000+ images of labeled road data.**

**--- engineering teams (e.g. IBM's Watson)**
**strong commercial interests**
**at a scale never seen before in our field**

***An AI arms race***

**--- Investments in AI systems are being scaled-up by an order of magnitude (to billions).**

**Google, Facebook, Baidu, IBM, Microsoft, Tesla etc. ($2B+)**
**+ military ($19B proposed) + China, Canada, France, et al.**

# AI milestones starting in the late 90s

**1997 IBM's Deep Blue defeats Kasparov**

**2005 Stanley --- self-driving car (controlled environment)**

**2011 IBM's Watson wins Jeopardy! (question answering)**

**2012 Speech recognition via "deep learning" (Geoff Hinton)**

**2014 Computer vision is starting to work (deep learning)**

**2015 Microsoft demos real-time translation (speech to speech)**

**2016 Google's AlphaGo defeats Lee Sedol**

**Google's WaveNet --- human level speech synthesis**

**2017 Watson technology automates 30 mid-level office insurance claim workers, Japan (IBM).**

**Automated dermatologists, human expert accuracy (Stanford)**

**Poker, Heads-up, No-Limit Texas Hold'em, CMU program beats top human players**

## Historical aside:

## World's first collision between fully autonomous cars (2007)


CORNELL MIT

# Next Phase

**Further integration of techniques --- perception, (deep) learning, inference, planning --- *will be a game changer for AI systems.***

**Example: AlphaGo:**

**Deep Learning**

**+**

**Reasoning (MCTS/UCT)**

**(Google/Deepmind 2016, 17)**

**Synthetic Chemistry (‘18)**

# *What We Can't Do Yet*

**Need deeper semantics of natural language**

***Requires commonsense knowledge and reasoning***

**Aside:**
***Google translation is really done without any understanding of the text! (very unexpected)***

**Example:**

**"The large ball crashed through the table because *it* was made of Styrofoam."**

**What was made of Styrofoam? The large ball or the table?**

**"The large ball crashed through the table because *it* was made of steel."**

**Hmm… Can't Google figure this out? No! (Carla Gomes)**

***English to French (from Carla Gomes)***

The **vase** crashed through the table because **it** was made of *steel*.
***Le* vase** s'est écrasé à travers ***la*** table parce **qu'il** était en *acier*.

The vase crashed through **the table** because **it** was made of *Styrofoam*.
***Le* vase** s'est écrasé à travers ***la*** table parce **qu'il** était fait de *polystyrène*.

**Commonsense is needed to deal with unforeseen cases.**
**("corner cases," i.e., cases not in training data)**


**China Tesla crash --- consider how human driver handles this!**
**You Tube:** Tesla crashes into an orange streetsweeper on Autopilot –Chinese Media

# Artificial Non-Human Intelligence

**AI focus: Human intelligence because that's the intelligence we know…**

**Cognition: Perception, learning, reasoning, planning, and knowledge.**

**Deep learning is changing what we thought we could do, at least in perception and learning (with enough data).**

**Separate development --- "non-human": Reasoning and planning. Similar qualitative and quantitative advances but "under the radar."**

> **Part of the world of software verification, program synthesis, and *automating science and mathematical discovery.***

**Developments proceed without attempts to mimic human intelligence or even human intelligence capabilities.**

**Truly machine-focused (digital): e.g., "verify this software procedure" or "synthesize procedure" --- can use billions of inference steps --- or "synthesize an optimal plan with 1,000 steps." (Near-optimal: 10,000+ steps.)**

**Next: Mathematical Discovery**

**Example** Consider a sequence of 1s and -1s, e.g.:

-1, 1, 1, -1, 1, 1, -1, 1, -1 …

1 2 3 4 5 6 7 8 9 …

2 4 6 8 …

3 6 9 …

**Erdos Discrepancy Conjecture**

and look at the sum of sequences and subsequences

-1 + 1 = 0

-1 + 1 + 1 = 1

-1 + 1 + 1 + -1 = 0

-1 + 1 + 1 + -1 + 1 = 1

-1 + 1 + 1 + -1 + 1 + 1 = 2

-1 + 1 + 1 + -1 + 1 + 1 + -1 = 1

-1 + 1 + 1 + -1 + 1 + 1 + -1 + 1 = 2

-1 + 1 + 1 + -1 + 1 + 1 + -1 + 1 + - 1 = 1

etc.

and "skip by 1"

1 + -1 = 0

1 + -1 + 1 = 1

1 + -1 + 1 + 1 = 2

etc.

and "skip by 2"

1 + 1 = 2

1 + 1 + -1 = 1

etc.

We now know (2015): there exists a sequence of *1160* +1s and -1s such that sums of all subsequences *never* < -2 or > +2.

**Example** Consider a sequence of 1s and -1s, e.g.:

-1, 1, 1, -1, 1, 1, -1, 1, -1, -1 …

1 -1 1 1 -1 …

1 1 -1 …

-1 1 …

**Erdos Discrepancy Conjecture**

and look at the sum of the sequence and its subsequences

-1 + 1 = 0

-1 + 1 + 1 = 1

-1 + 1 + 1 + -1 = 0

-1 + 1 + 1 + -1 + 1 = 1

-1 + 1 + 1 + -1 + 1 + 1 = 2

-1 + 1 + 1 + -1 + 1 + 1 + -1 = 1

-1 + 1 + 1 + -1 + 1 + 1 + -1 + 1 = 2

-1 + 1 + 1 + -1 + 1 + 1 + -1 + 1 + - 1 = 1

etc.

and "skip by 1"

1 + -1 = 0

1 + -1 + 1 = 1

1 + -1 + 1 + 1 = 2

etc.

and "skip by 2"

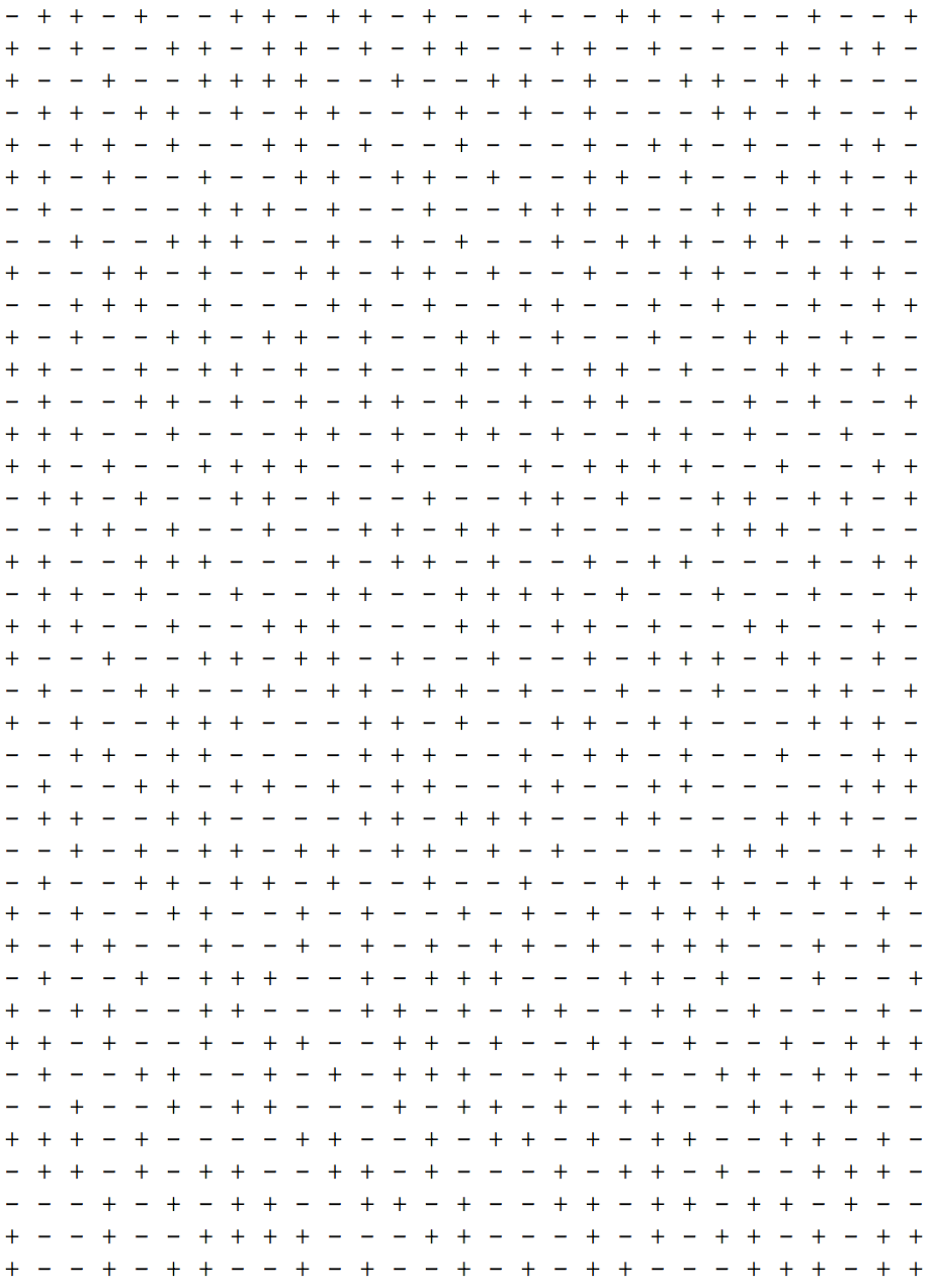
1 + 1 = 2

1 + 1 + -1 = 1

etc.

We now know (2015): there exists a sequence of *1160* +1s and -1s such that sums of all subsequences *never* < -2 or > +2.

**1160 elements all sub-sums stay between -2 and +2**



**So, we now know (2015): there exists a sequence of *1160* +1s and -1s such that sums of all subsequences *never* < -2 or > +2.**

**Result was obtained with a *general* reasoning program (a Boolean Satisfiability or SAT solver). *Surprisingly*, the approach far outperformed specialized search methods written for the problem, including ones based on other known types of sequences. (A PolyMath project started in January 2010.)**

**But, remarkably, no such sequence of *1161* or longer exists!**

**(> 10^300 such sequences; each has a subsequence adding to a +3 (or -3) somewhere)**

**Encoding: 37,462 Boolean variables and 161,644 clauses / constraints.**

***Proof of non-existence of discrepancy 2 sequence found in about 10 hours (SAT Solver, MacBook Air).***

***Proof: 13 gigabytes and independently verified (50 line proof checking program). Proof is around a billion small inference steps.***

**Longest known math proof (2015).**

**Machine "understands" and can verify result easily (milliseconds).**

**Humans: probably never. ☹ Still, we can be certain of the result because of the verifier.**

***So, future human math can be augmented with machine discovered math.***

**(Similarly, in game play, AlphaGo augments human Go play.)**

# Non-Human Intelligence
# Comp. Complexity / Intelligence Hierarchy

EXP-complete:
games like Go, …

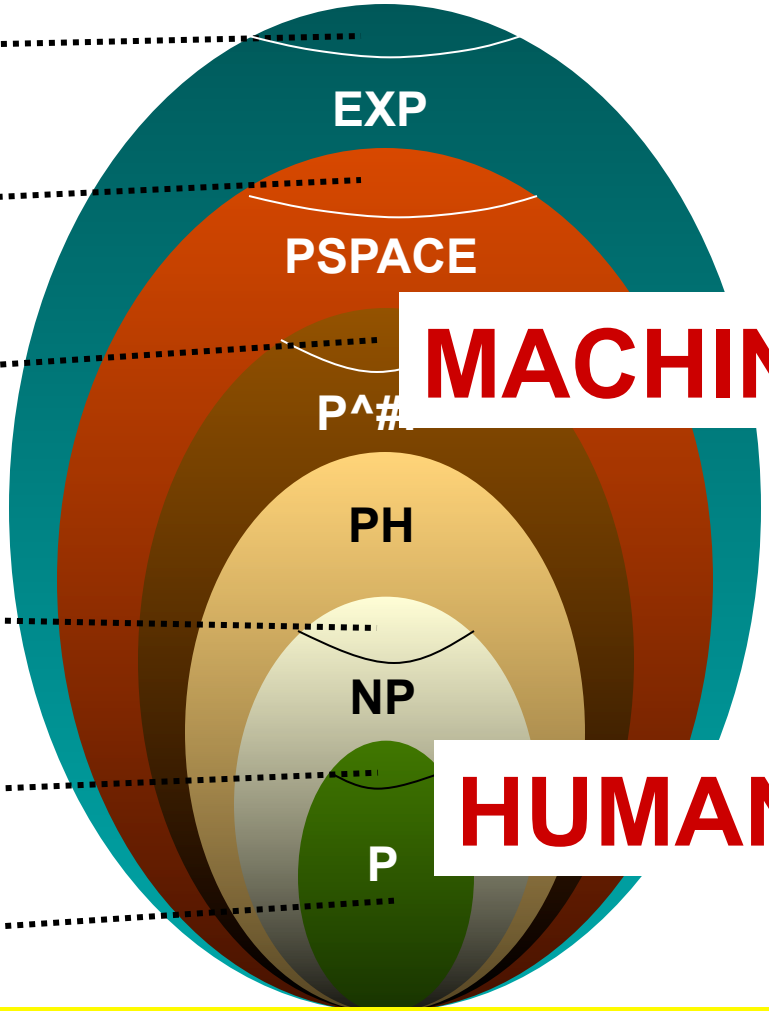
**PSPACE-complete**:
QBF, ***planning***, chess (bounded), …

**#P-complete/hard**:
**#SAT, sampling**,
***probabilistic inference***, …

**NP-complete**:
**SAT**, ***propositional reasoning***, scheduling, graph coloring, puzzles, …

P-complete:
circuit-value, …

In P:
sorting, shortest path




**What are the consequences for human understanding of machine intelligence?**

**The emergence of intelligent autonomous machines among us is expected to have a major impact on society.**

***"Preparing for the Future of Artificial Intelligence"***

**White House Report,**

**Executive Office of the President, Oct. 2016**

**Societal issues:**

**1) Economics (wealth inequality) & Employment**

**2) AI Safety & Ethics**

**3) Military Impact (Smart autonomous weapon systems)**

**4) The Future: Super-Intelligence? Living with smart machines.**

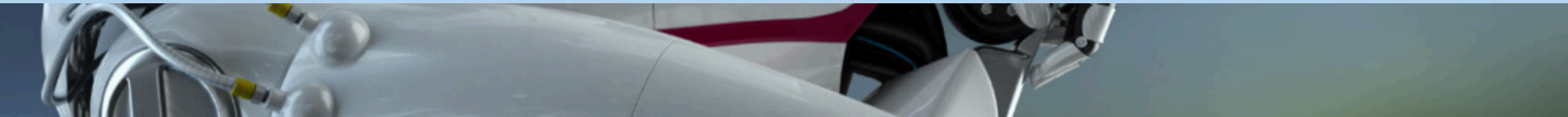
**Elon Musk: Future of Life Institute (Max Tegmark, MIT)**

**AI Safety research program**

# Lecture Program

**Mondays, 7:30-8:30pm, Olin 155**

**02/27** – Bart Selman, Cornell: *The Future of AI: Benefits vs. Risks*
**03/06** – Jon Kleinberg, Cornell: *Trade-Offs in Algorithmic Fairness*
**03/13** – Kilian Weinberger, Cornell: *Interpretable Machine Learning*
**03/20** – Dan Weld, Univ. of Washington: *Computational Ethics for AI*
**03/27** – Moshe Vardi, Rice Univ.: *Humans, Machines, and Work*
04/03 – spring break
04/10 – no lecture
**04/17** – Karen Levy, Cornell: *Working with and Against AI*
**04/24** – Ross Knepper, Cornell: *The Ethics of Robotics*
**05/01** – Joe Halpern, Cornell: *Morality: In Search of Formal Definitions*



# 1) Economic Impact: Technological Unemployment

**Example 1: self-driving vehicles (5 - 10 yrs).**
**90+% accident reduction BUT**

***Transportation covers about 1 in 10 US jobs!***
***Not so easy to replace… Also hospital emergency room reduction…***

***Retrain? But for what?***
***Knowledge worker? (see next)***
***STEM field? (too small)***

**Example 2: IBM Watson style automation of 30 insurance admin jobs (2017, Japan).**

**Expensive to create system but easy to duplicate…**
**Places *mid-level knowledge-based* jobs at risk.**

*Most jobs with a significant routine component will be affected.*
*Significant economic incentive for companies to pursue automation.*
*40+% of jobs at risk.*

**It appears inevitable that advanced AI *(systems that can hear, see, reason, plan, and learn)* will have a significant impact on employment and our society in general.**

**Human society will need to prepare itself.**
**Universal basic income?**
**Without work, how do we feel useful?**
**Amplification of wealth inequality?**

# 2) AI Safety & Ethics

## Area 1: Issues with Machine Learning (ML) Data-Driven Approaches

***Data-driven ML approaches are starting to provide decision support at all levels of society.***

**Examples:**

***a) Financial loan approvals***
***b) Hiring / interview decisions***
***c) Google search order rankings***
***d) College applicant selection***
***e) Medical diagnosis***
***f) What's in your news feed…***
***g) Your year-end raise***

***Etc.***

**What about hidden biases in these decisions? & Are data-driven decisions fair?**

**ML approaches include hidden biases from data (e.g. past hiring / performance data) and from algorithms (e.g., what types of unfair bias cannot be eliminated?)**

**EU on the forefront: Working on laws to require explainable machine learning results. Also, statistical models need to be shown to adhere to non-discrimination laws.**

**Problem: not so easy to do!**

**But, at least, Google can no longer just say "Results are fair because they are decided by an algorithm and data. And, algorithms and data are always fair." That worked great for a while… :-)**

# 2) AI Safety & Ethics, cont.

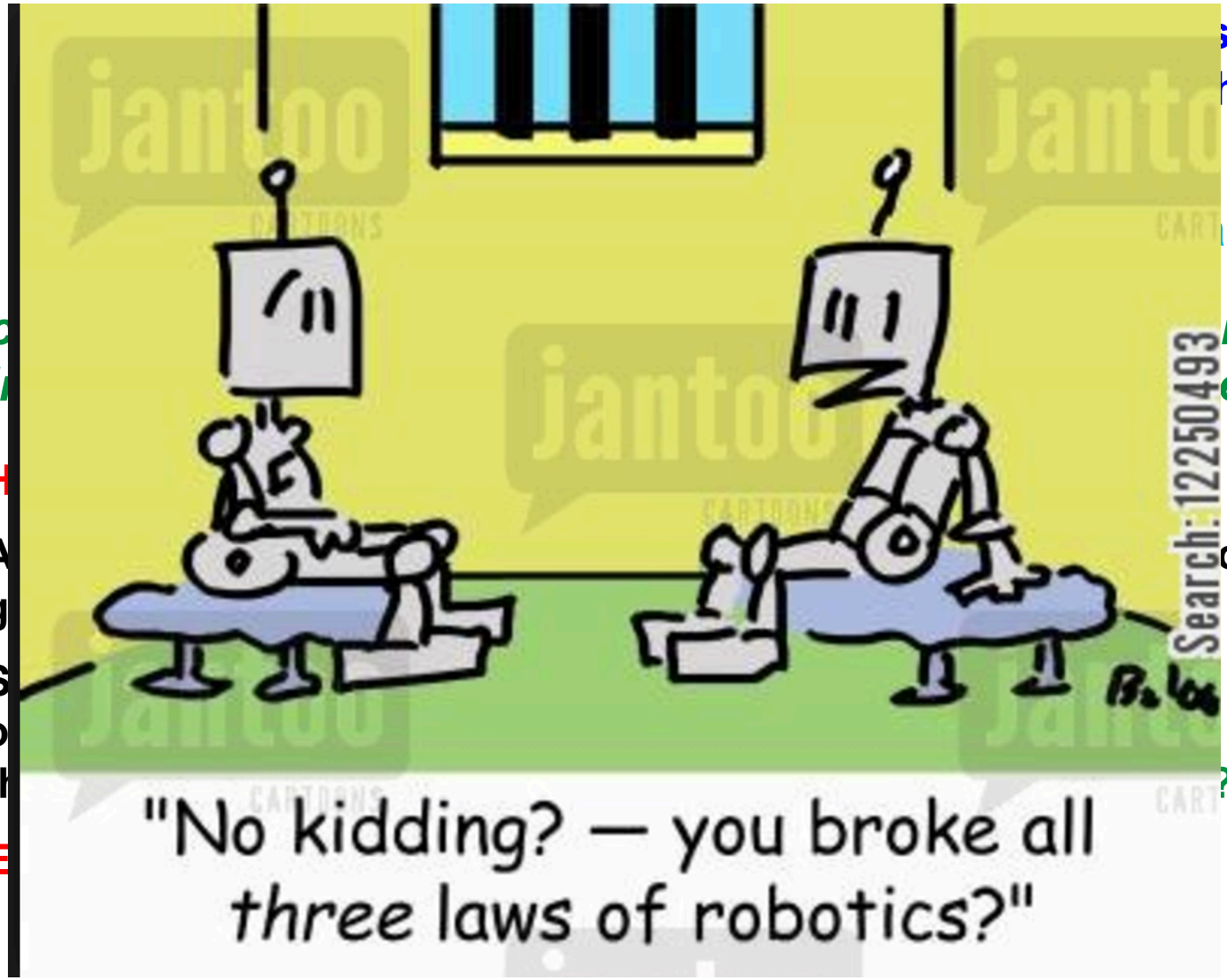
## Area 2: Autonomous Goal-Driven Systems that Plan and Reason

***Autonomous AI systems (eg robots or virtual assistants) no longer follow the traditional programming paradigm with detailed hand-coded sequence of instructions.***

**Instead:** ***only high-level goals or instructions are given, and the system synthesize sequences of actions to perform.***

**How do we ensure that these decision making systems do what we want them to do and do so in a responsible matter benefiting humans?**

**"The Value Alignment Problem." Stuart Russell, UC Berkeley.**


"No kidding? — you broke all three laws of robotics?"

# 3) War & Peace

**AI scientists and others have recently raised significant concerns about the risks of an smart, AI-based Autonomous Weapons race.**

**Lots of press**  **pon systems because of th .**

**Also part of c s, with countries wo**

**Issue far from tional and international are being considered.**

**AI researchers discussing the risk of an AI Arms Race at the White House, 2016.**

# 4) Future: Super-Human Intelligence?

Stephen Hawking: AI will be 'either best or worst thing' for humanity

**Super-human AI often gets the most press.**
**Will we be "superseded" by smart machines?**

**May work out much better than some have argued. Push for AI Safety Research (funded by Elon Musk and others) will quite likely ensure a tight coupling between human and machine interests.**

**Also, even if machines outperform us on a range of intellectual tasks, that does necessarily mean we won't be able to understand the systems. *Humans can understand complex solutions even if we do not discover them ourselves!***

***We're on an exciting intellectual journey in the history of humanity!***