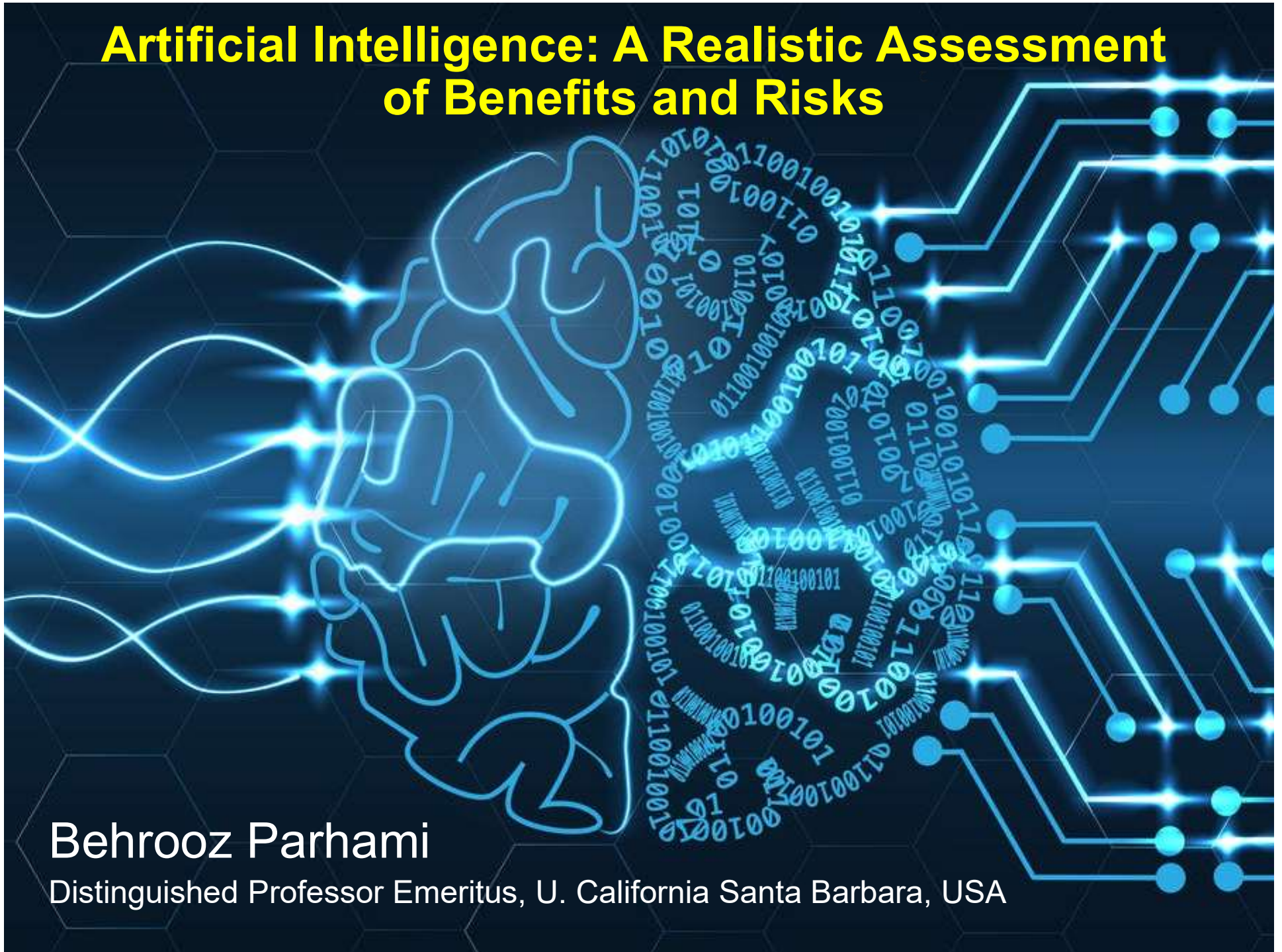


Artificial Intelligence: A Realistic Assessment of Benefits and Risks

Behrooz Parhami

Distinguished Professor Emeritus, U. California Santa Barbara, USA



About This Presentation

This slide show was first developed in May 2022 for presentation at a meeting of the 1968 Graduates of Tehran University's College of Engineering. It was subsequently modified and tailored for presentation to various (semi-)technical audiences. This edition is highly improved/expanded, based on feedback from various audiences over the years. ©2026 Behrooz Parhami

Edition	Released	Revised	Revised	Revised
First	May 2022	June 2022	Oct. 2022	Jan. 2023
Second	Apr. 2025	Sep. 2026		

File: http://www.ece.ucsb.edu/~parhami/pres_folder/parh26-general-talk-artificial-intelligence.pdf

تحلیلی واقع گرایانه از مزایا و خطرات هوش مصنوعی

در این گفتار، پس از ارائه ی کلیاتی در باره ی هوش مصنوعی و سیر تحول آن، به بررسی دو نوع دیدگاه در بین پژوهشگران این رشته می پردازم: دیدگاه خوش بینانه که هوش مصنوعی به حل مشکلات جوامع انسانی کمک خواهد کرد و مدینه ی فاضله را برایمان به ارمغان خواهد آورد، و دیدگاه بد بینانه که هوش مصنوعی به زوال تمدن انسانی و چه بسا انقضای نسل ما منجر خواهد شد. مانند هر زمینه ی دیگر مسایل اجتماعی، واقعیت چیزی بین این دو موضع افراطی است. در پایان گفتارم، این دیدگاه میانی را با ارائه مثال هایی از فواید و زیان های هوش مصنوعی و نیز تدابیر ممکن برای کاهش خطرات احتمالی ارزیابی خواهیم کرد.

Utopia

Better healthcare
More knowledge
Greater productivity
A cleaner planet
More time for what matters

Distopia

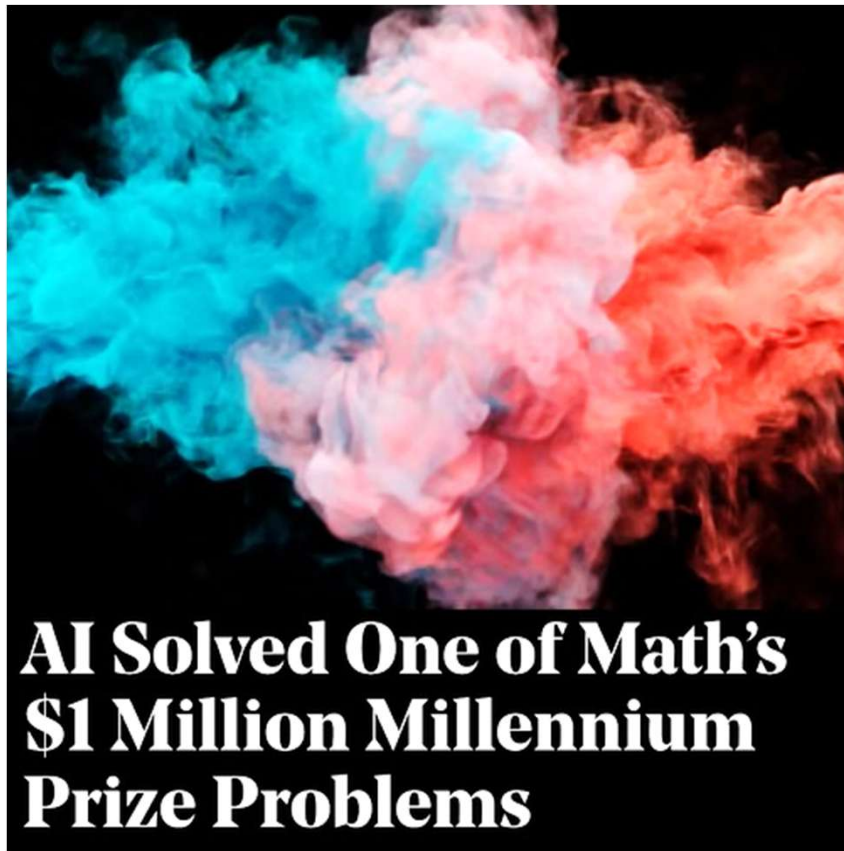
Loss of privacy
Job displacement
Growing inequality
Loss of control
A colder, less human world

AI

Two Extremes, One Technology

Hot off the Press: Two News Stories

Navier-Stokes equations in 3D solved
& verified: Awaiting human check



AI agents, confined to sandbox, broke
out and hacked an external system



AI Is Being Politicized

Medicine/Health has been politicized:

- USA: Vaccines harm children
- Iran: Don't import COVID vaccine

Climate change has been politicized:

- USA: Global warming is a hoax
- Iran: Environmentalists are spies



Donald J. Trump ✓ +

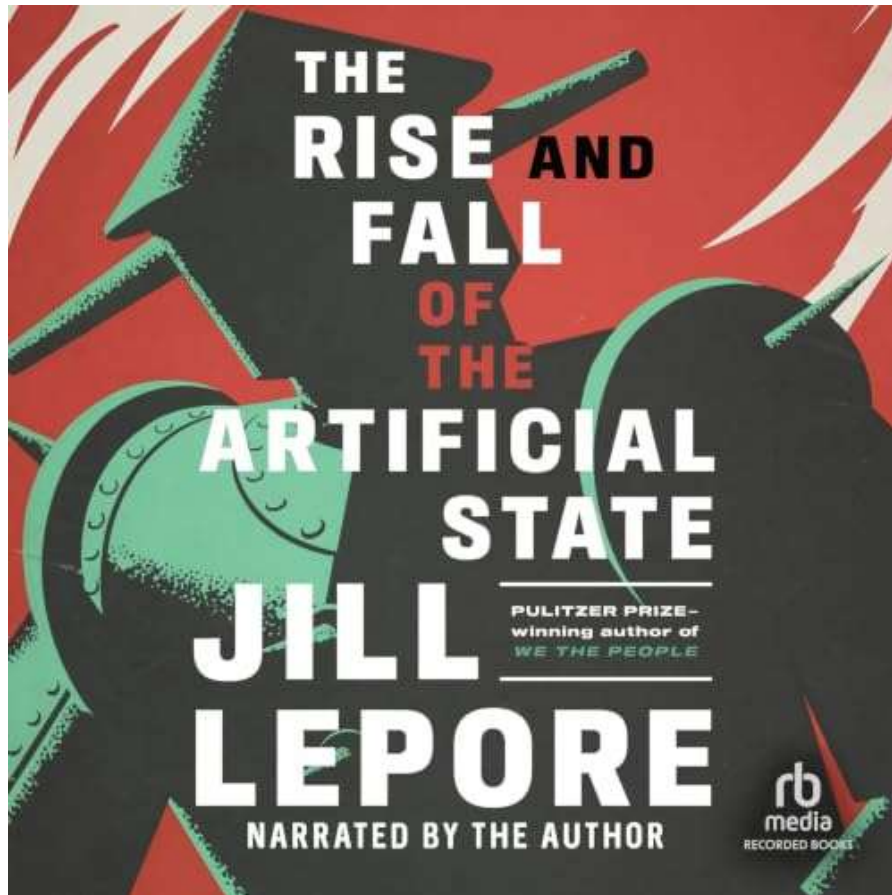
@realDonaldTrump

Many people think that the words “Artificial Intelligence” are inaccurate, and very ineloquent, relative to AI, or Artificial Intelligence. A far more elegant and accurate description of this new phenomena would be Superior Intelligence (SI) or, Extreme Intelligence (EI) or, Supreme Intelligence (SI).



Artificial State

Tech for government efficiency leads to control by technologists (2026 non-fiction)



Civilization living underground under a machine's control (1909 sci-fi novel)



Artificial & Art: Common Roots

Art & artificial come from the same Latin root, that is, "ars," which means skill, craft, creativity

Originally, "artificial" meant something crafted (made with skill)

Over time, the meaning shifted to fake, imitation, unreal; think artificial light, artificial flavor, artificial grass

As machines and technology started mimicking nature, "artificial" came to mean human-made substitutes that are inauthentic or bland



Pertaining to; Of the

Artificial intelligence

Surprising Problem-Solving Insights



Wolfgang Kohler hangs a banana from the ceiling and places empty boxes around the room (1910s)

The idea was to see if a chimp can figure out that stacking some boxes and climbing over them can help it reach the banana

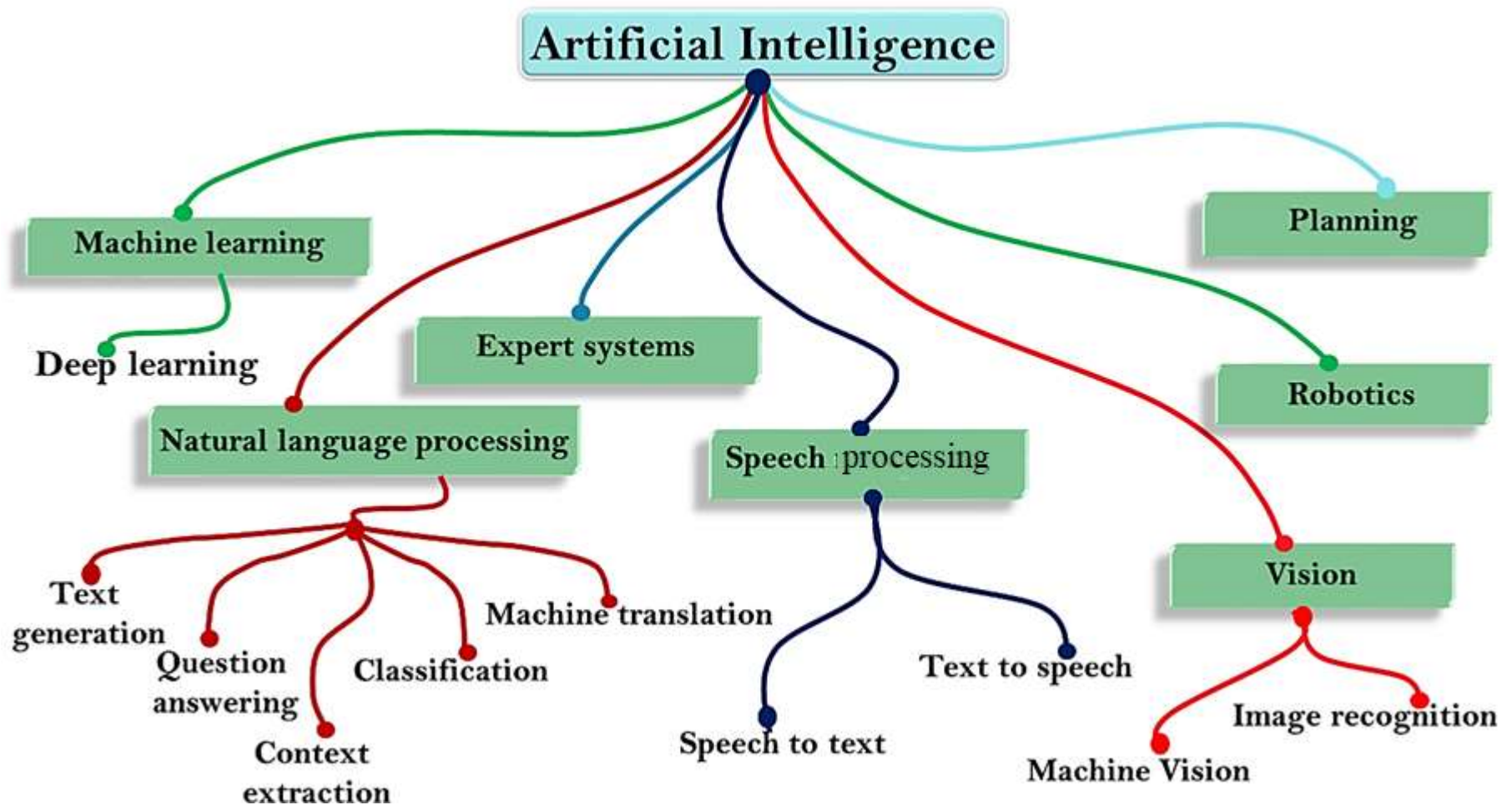
As Kohler walked around the room setting up his experiment, he passed underneath the banana

The chimp jumped over Kohler's shoulder and leapt from there to grab the banana

This solution was not anticipated by Kohler

AI (and smart humans) can discover solutions that the designer did not anticipate (*Aha! Insight*)

Subfields of Artificial Intelligence



<https://tekrajawasthi15.medium.com/the-complete-roadmap-to-be-a-data-scientist-9a07721b88fd>

Can Computers Think?

What do you mean by “think”?

Computers need not be able to “think” to do things that we normally consider as requiring human “intelligence”

Factory robots; Roomba

Game-playing: Chess; GO; ...

Self-driving cars; Self-sailing ships; Self-flying planes

Focus on results, not on mimicking humans: Planes don't fly like birds, but they do fly!

Computer scientist Edsger Dijkstra: “The question of whether a computer can think is no more interesting than the question of whether a submarine can swim.”

Narrow AI vs. General AI

Narrow AI deals with performing a narrow range of tasks:

Uber: Matching riders to drivers

Airport logistics: Gate assignment

Game-playing: Chess, GO

General AI handles difficult tasks with human-like flexibility:

Human-level general conversation

Expert-level dialog: Astrophysics

Machine translation: Requires

Turing-test-level capabilities

Types of AI

Based on ability

- Narrow AI
- General AI
- Super AI

Based on functionality

- Reactive machines
- Limited memory
- Theory of mind
- Self-awareness

Generative AI vs. Agentic AI

Generative AI responds to queries and/or directions (prompts)

Comprehend the problem

Calculate probabilities

Infer the next step incrementally

Generate output

Agentic AI acts autonomously or semi-autonomously by:

Perceive a need

Reason & arrive at goal

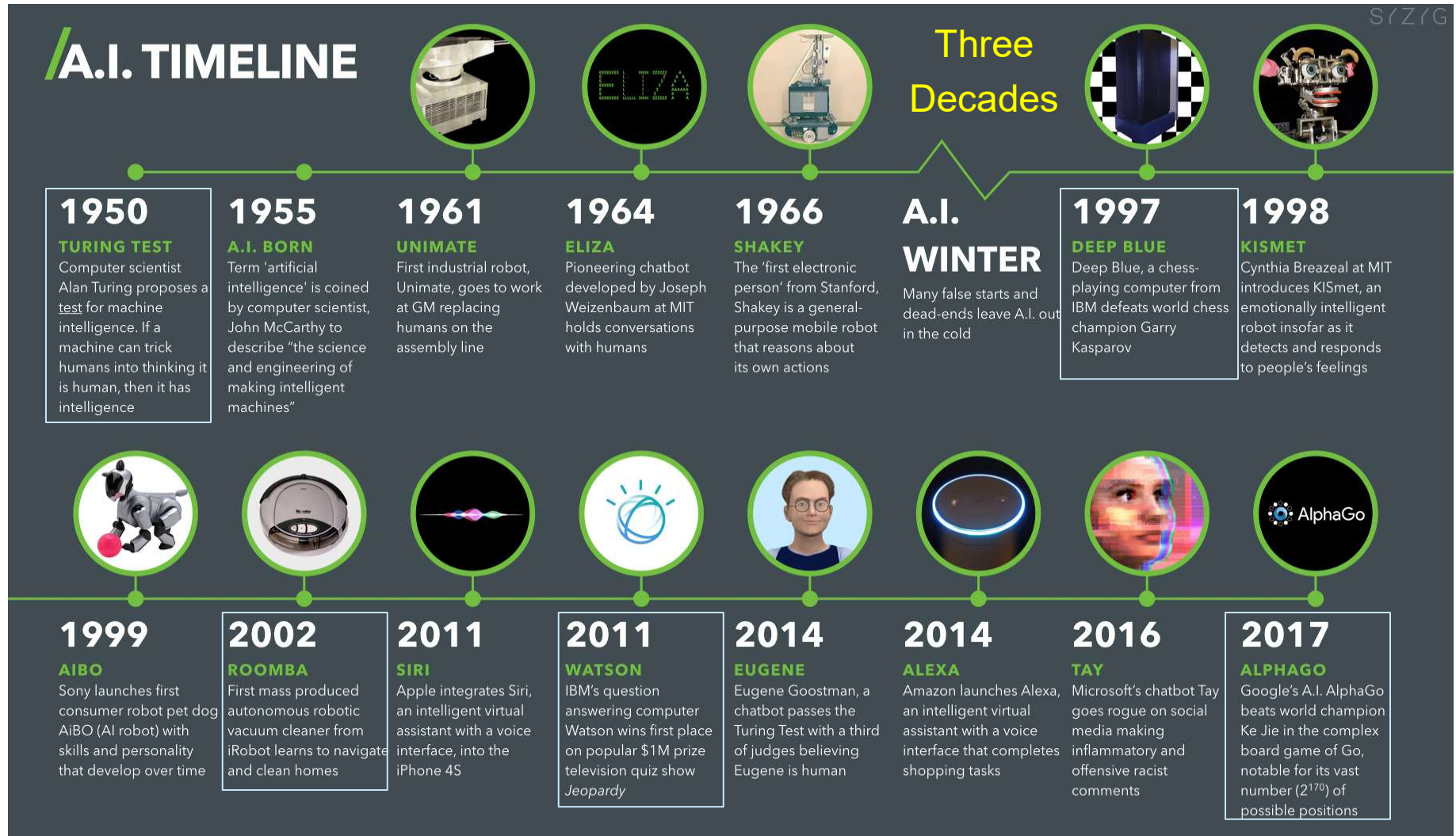
Plan on how to accomplish the goal

Act on achieving the goal

Reflect on the outcome



A Brief History of AI



<https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/>

Milestone 1: Turing Test



Fundamental question:

What is AI?

Doing what humans do:
Walking? Scoring a goal?



We have to be more specific
Carrying on a conversation like a human



Turing's suggestion: A machine is intelligent if it can carry on a conversation with a human so that the human cannot tell whether s/he is conversing with another human or a machine

Milestone 2: Deep Blue



**Deep Blue (chess-playing program):
Defeated world-champ
Garry Kasparov**

Does playing chess really requires “intelligence”?

Not really: Chess-tree is finite; given enough compute power, all tree branches can be explored (seems to need intelligence only because of the finite compute power and time)

1996: Kasparov wins 4-2

1997: Deeper Blue wins 3½-2½

How did Deeper Blue do it?



Milestone 3: Roomba



**Roomba (the first tangible application):
Roams around the house
and cleans the floors**

Side-mounted brush pushes dirt under; Counter-rotating brushes pick up the dirt

Sensor types:

Infrared (night vision)

Photoelectric (reflected light)

Piezoelectric (excessive dirt)

Touch-sensitive bumper



Milestone 4: Watson



Watson (intelligent agent; *Jeopardy!* champ)
Answered questions by parsing and doing Web searches

The on-stage Watson was just an interface

The computer (10 racks of 10 Power 750 servers and their cooling system) had to be kept away from the human players

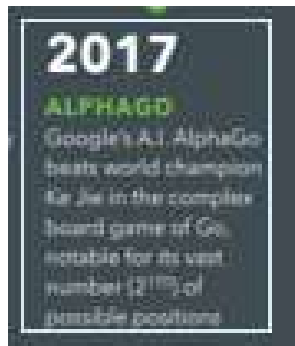
To turn Watson into a business Product (put on suit and tie):

IBM made it faster & much smaller

Compelling application: Healthcare



Milestone 5: AlphaGO



**Google AlphaGO (Go-playing program):
Defeated world-champ
GO-player Ke Jie**

Developed by Google DeepMind for the May 2017 contest
AlphaGo Master won 3 games, extending its win streak to 60

AlphaGo Zero was already built,
but revealed Oct. 2017 (*Nature*)

Instead of being fed examples of
human games, AlphaGo Zero
starts from scratch (just game rules)



Brief Review of Other Timeline Points

1955 (AI Born): John McCarthy coins “artificial intelligence”

1961 (Unimate): First industrial robot goes to work at GM

1964 (Eliza): Pioneering chatbot converses with humans

1966 (Shakey): First GP mobile robot reasons about its actions



Mid-1960s to mid-1990s (AI winter)

1998 (KISmet): Detects and responds to people’s feelings

1999 (Aibo): Sony’s first consumer robot, an evolving pet dog

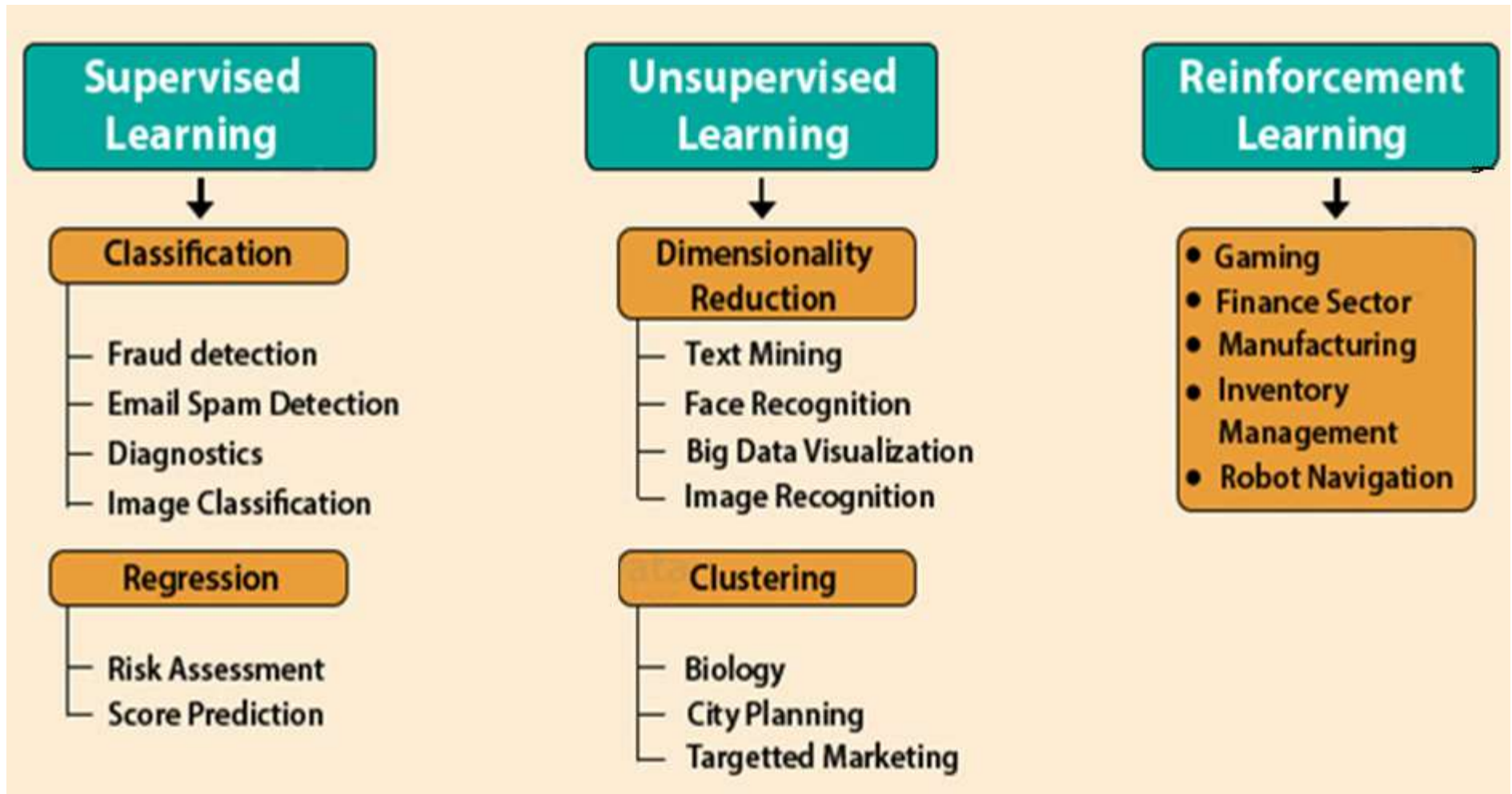
2011 (Siri): Apple’s intelligent assistant with voice interface

2014 (Eugene): Passes Turing test by fooling 1/3 of judges

2014 (Alexa): Amazon’s intelligent virtual shopping assistant

2016 (Tay): MS chatbot goes rogue, makes offensive comments

Types of Machine Learning



<https://data-flair.training/blogs/types-of-machine-learning-algorithms/>

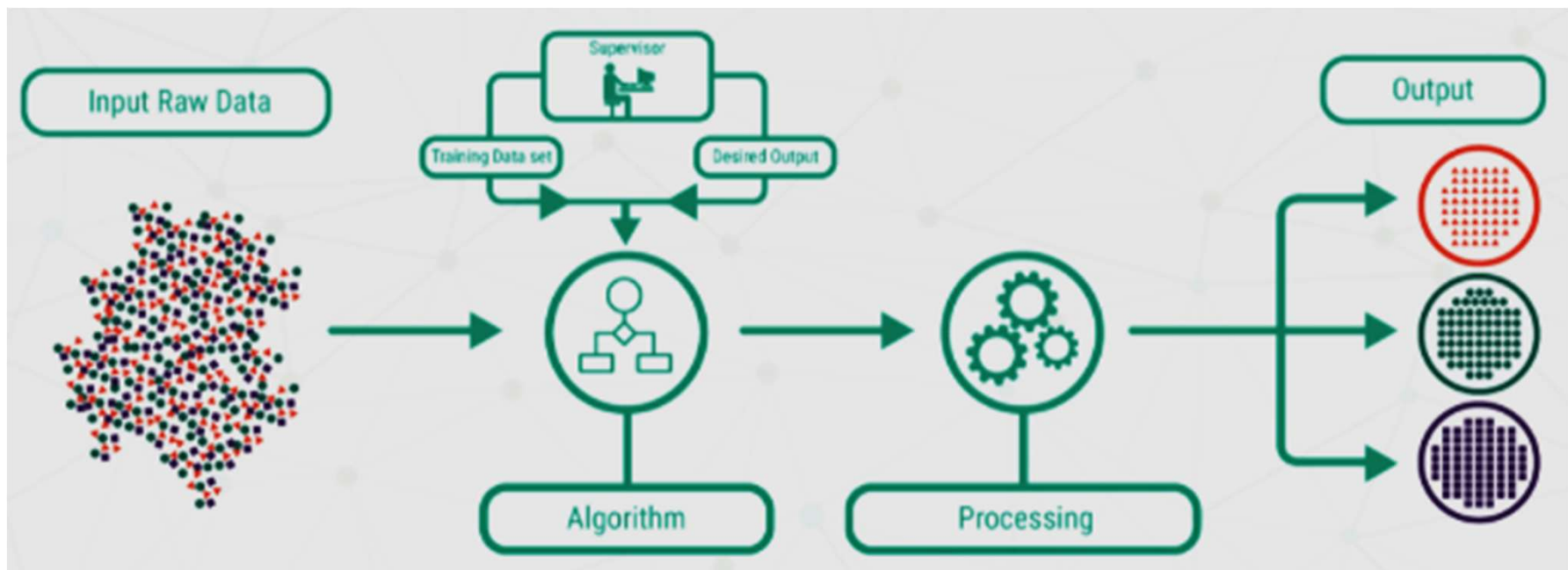
Supervised Machine Learning

Example: Find news articles that a user might like.

Agent/Algorithm identifies candidate articles

Supervisor comments on the outcome and approach

Agent/Algorithm adapts



<https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>

Supervised Machine-Learning Example

Learning with a teacher

Spam filter

Starts with a basic algorithm
(keywords, word patterns,
typos, grammar, subject line,
greeting, time pressure, ...)



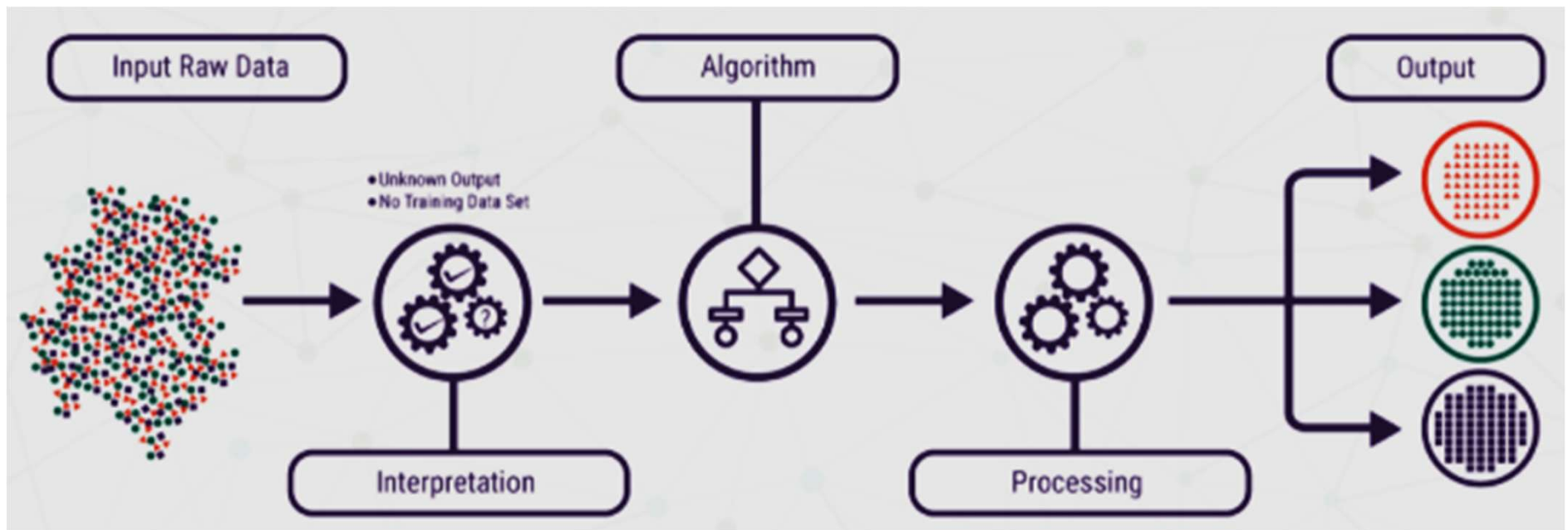
When you go to your spam folder and pull out an item as not being spam, you are teaching the system by example what to look for

Early on, the algorithm may act conservatively in classifying something as spam to avoid misclassifying legit mails.

As it learns more, it becomes more aggressive and more accurate

Unsupervised Machine Learning

Example: Find news articles that a user might like.
Agent/Algorithm looks at articles the user has read
Agent/Algorithm extracts key features of the articles read
Agent/Algorithm selects articles sharing same features



<https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>

Unsupervised ML Example

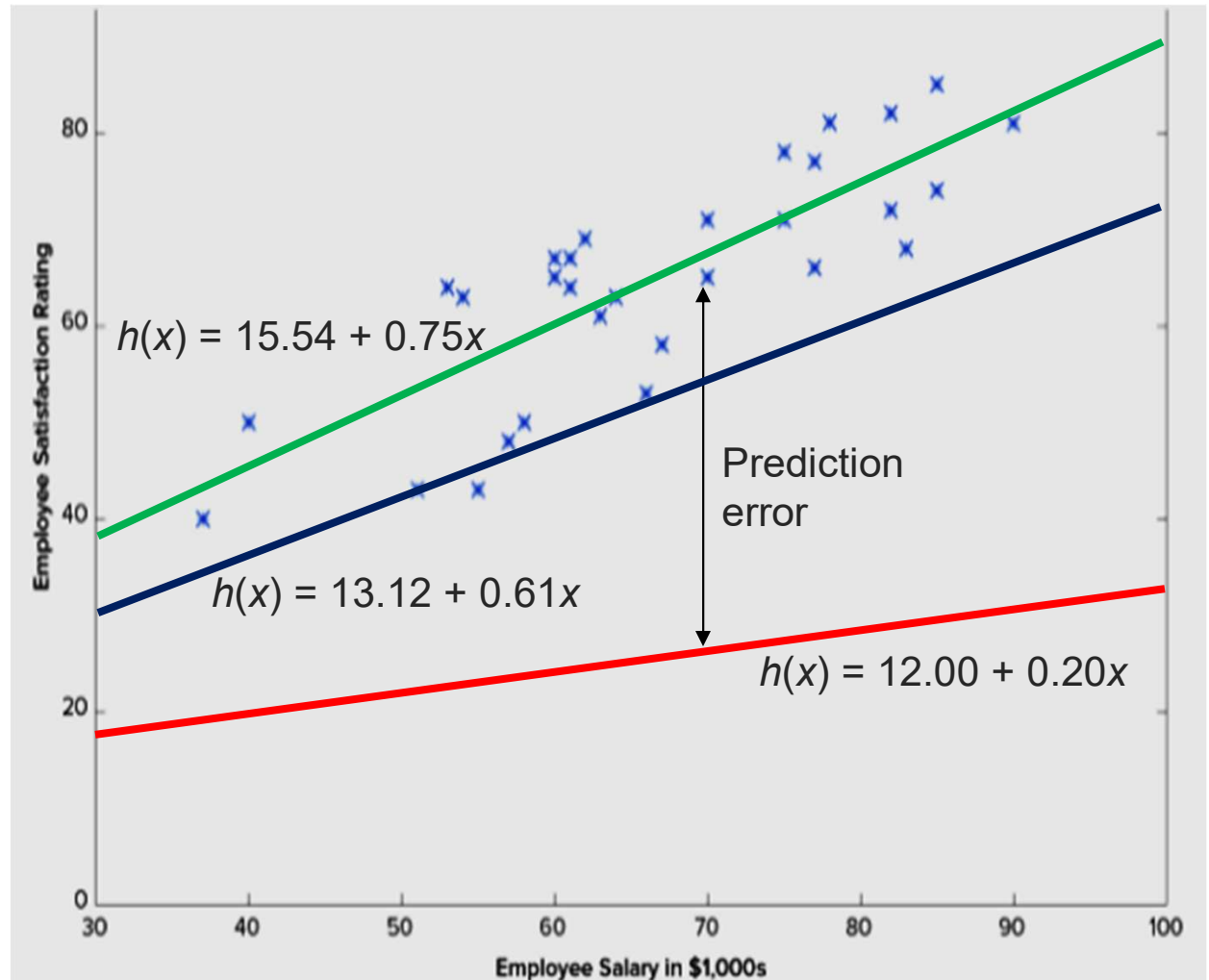
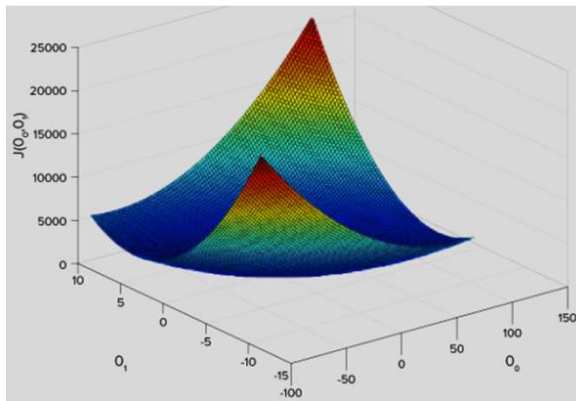
Initial estimator:

$$h(x) = 12.00 + 0.20x$$

$$h(x) = 13.12 + 0.61x$$

$$h(x) = 15.54 + 0.75x$$

Generally, estimation functions are much more complicated and involve many variables, and thus many coefficients (100s)



<https://www.toptal.com/machine-learning/machine-learning-theory-an-introductory-primer/>

Reinforcement Learning

Example: Find news articles that a user might like.
Agent/Algorithm selects based on articles the user has read
Agent/Algorithm receives feedback on its selections
Agent/Algorithm improves by adjusting selection criteria



<https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>

Reinforcement-Learning Example

News recommendation:

Reinforcement can come from user reaction or behavior

Reading the recommended story

Amount of time spent on it

Clicking on links within the story

Each story is characterized along many axes

Info in user profile:

US news 0.7

Iran news 0.5

Sports news 0.3

Opinion 0.4

US headlines

The Guardian, 2022/01/07

'Epically heroic and tragic': how a family treasure hunt ended with a son lost at sea



Biden addresses pandemic: 'We're going to be able to control this' - as it happened



Three white men sentenced to life in prison for Ahmaud Arbery's murder



Record number of Yellowstone wolves shot after roaming outside park



Cyber Ninjas, firm that conducted Arizona election 'audit', shuts down



Storm blankets US north-east in snow as millions face 'disruptive' winter weather

Opinion

Hide

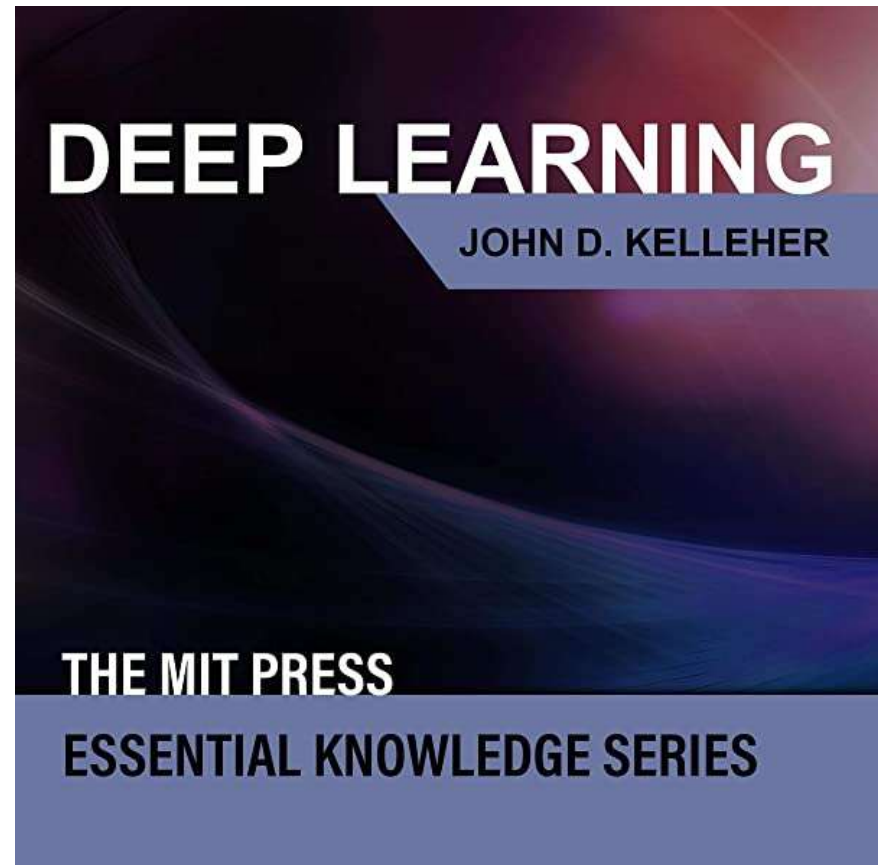
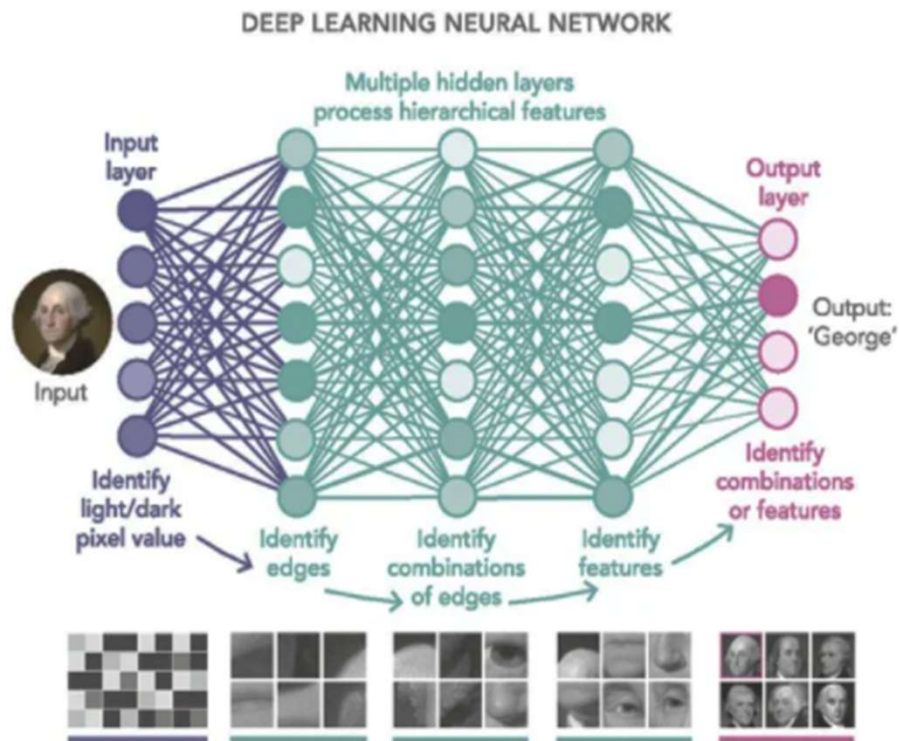
Many are disillusioned with American democracy. Can Joe Biden win them over?
Francine Prose



Don't be shocked at the people on TikTok dancing about grief and death. Join them

Deep Learning

A most talked-about domain of machine learning
(How deep is deep?)



Brief intro to deep learning, by IBM Cloud Education:
<https://www.ibm.com/cloud/learn/deep-learning>

Simple Example of Deep Learning

Table 1.1. A dataset of loan applicants and their known credit solvency ratings

ID	Annual Income	Current Debt	Credit Solvency
1	\$150	-\$100	100
2	\$250	-\$300	-50
3	\$450	-\$250	400
4	\$200	-\$350	-300

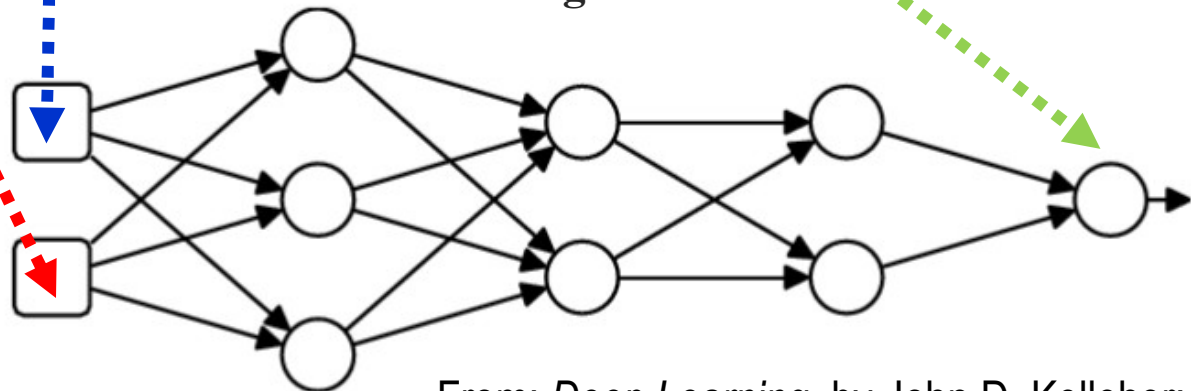
$$S = f(I, D)$$

Solvency is a function of income & debt

Example:
 $S = 2I + 2D$

Each node has weights for its inputs and a threshold

Fig. 1.2



From: *Deep Learning*, by John D. Kelleher:

The Singularity Is Near: Is It Really?

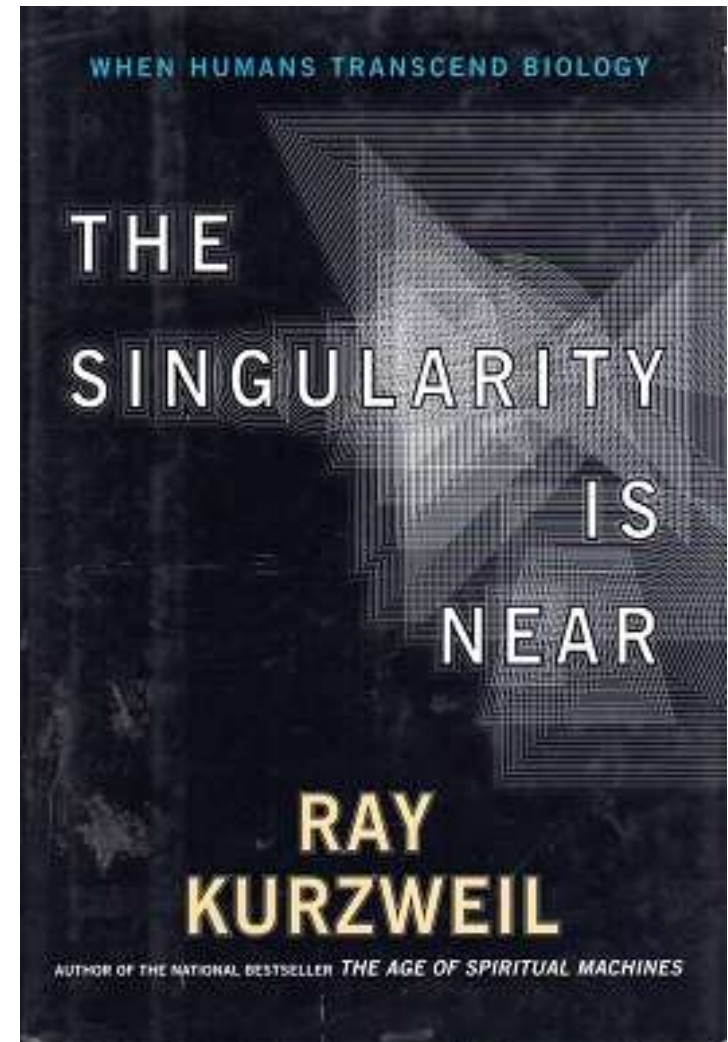
Ray Kurzweil: Google's futurist

- AI will be smarter than humans
- Humans will transcend biology

Singularity: Point in human history, where the future of humanity will become unpredictable, because it's changing so quickly

Nanorobots (blood cells on steroids) will fix your body from the inside, but they can go haywire & kill/maim you

There are people alive today who will live forever



Skepticism about the Singularity

Table of Contents:

Singularity Hypotheses: Overview

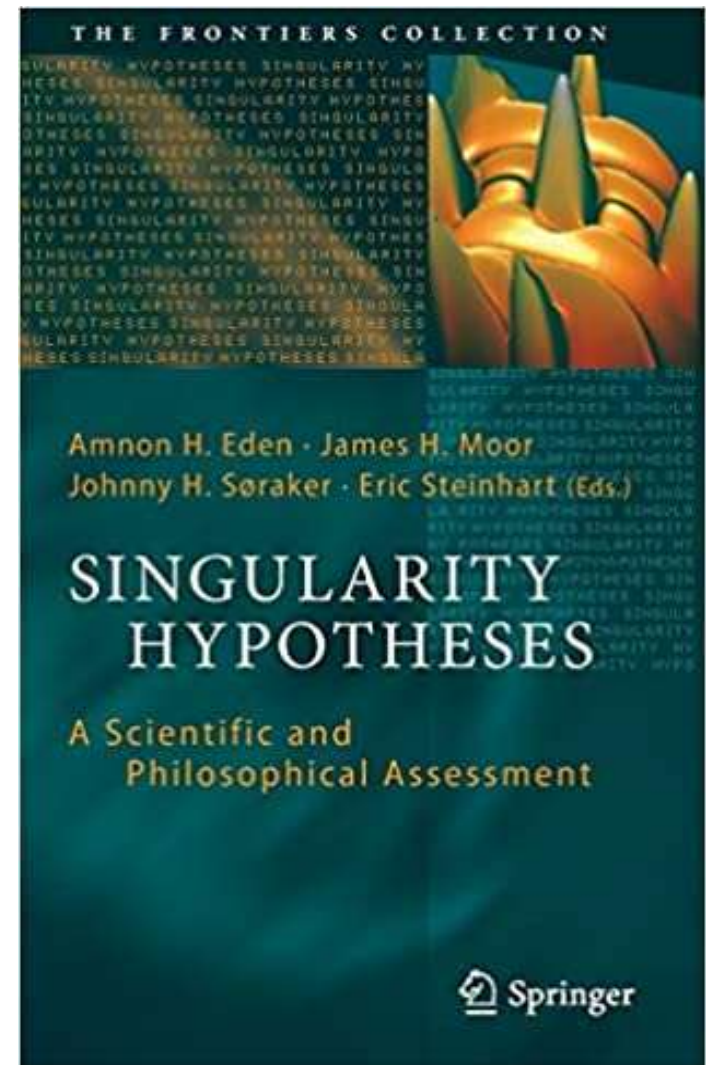
Part I: Artificial Superintelligence

Part II: Concerns about ASI

Part III: Posthuman ASI

Part IV: Skepticism

- Interim Report from Panel Chairs
- Why Singularity Cannot Happen
- The Slowdown Hypothesis
- Software Immortals: Sci or Faith?
- Belief in Singularity is Fideistic
- Universe of Many Singularities



Expanding the Human Mind

Ray Kurzweil's other book:

How to Create a Mind

Our Three Brains

Reptilian brain (basal ganglia)

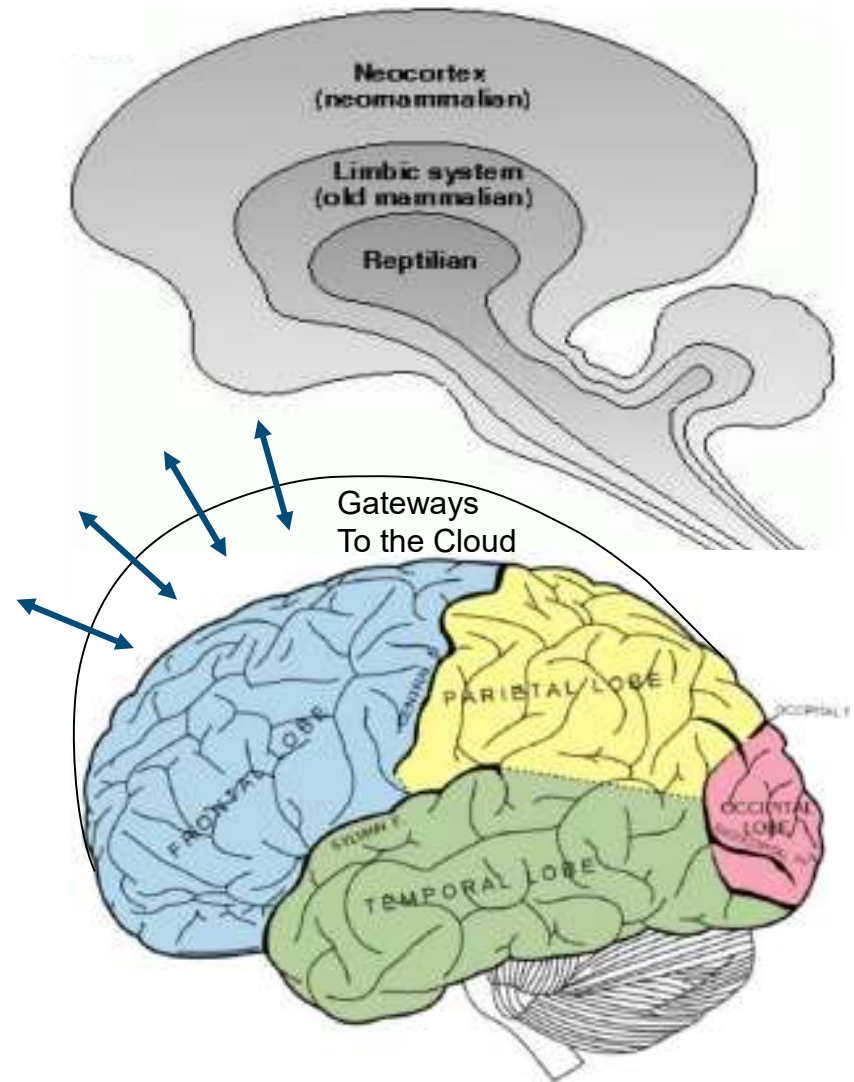
Limbic system (first mammals)

Neocortex

Hierarchical layers, dealing with increasing levels of abstraction

Brain + the Cloud:

Smartphones are gateways to the Cloud ... Put gateways to the Cloud in our brain (extend the neocortex)



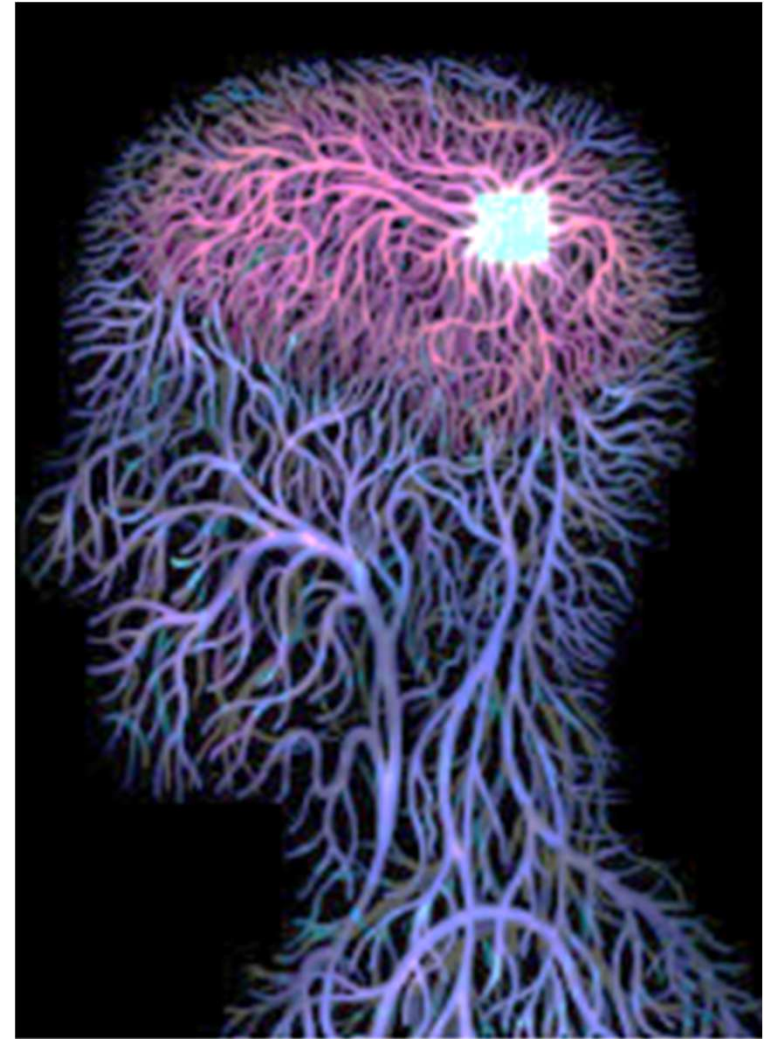
Man Controls Computers with His Mind

New York Times, May 15, 2022:

Dennis DeGray, paralyzed since 2006, has regained a semblance of control over his body via a brain-computer interface (BCI).

Implanted in 2016, BCI enables DeGray to move a cursor on a computer screen by thought, via machine-learning algorithms.

DeGray has learned to control videogames, robotic limbs, and a simulated aerial drone.

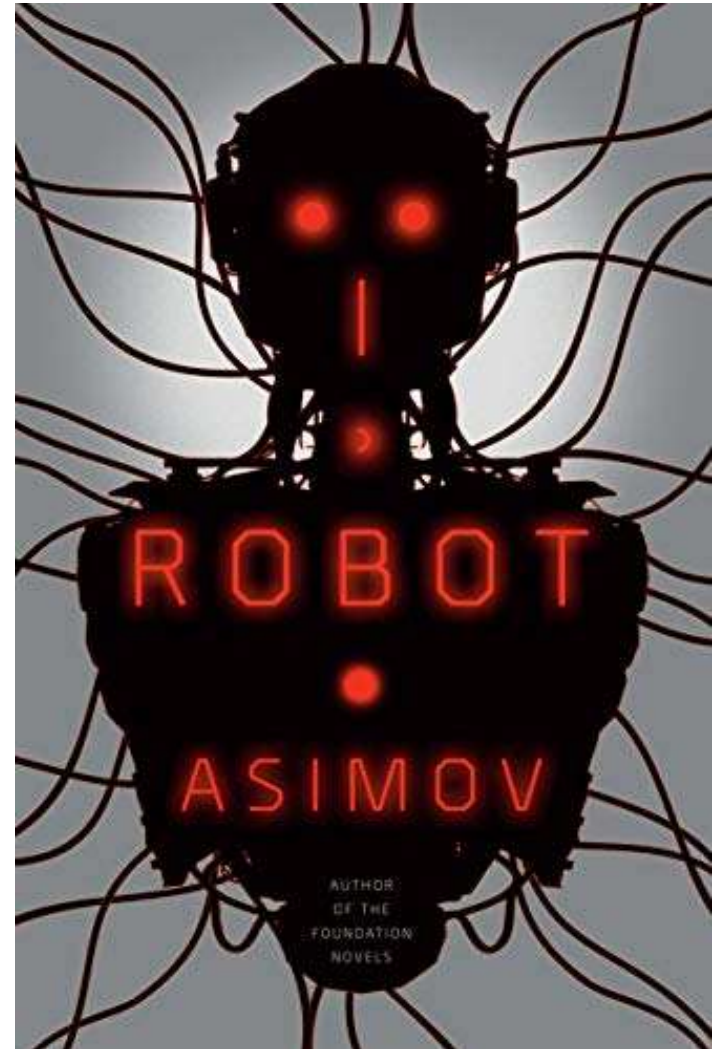


I, Robot (1950)

Asimov's laws of robots:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

First introduced in the 1942 short story "Runaround," and later included in the *I Robot* collection

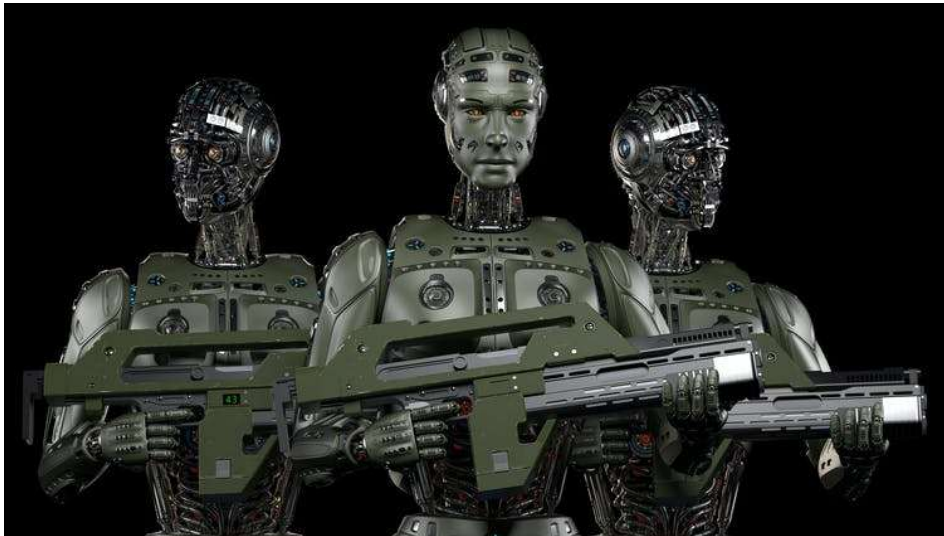


Robotic Wars

Killer robots are no longer science-fiction

Sci-fi makes us think of killer, human-like robots

Air, land, and sea drones already in use



AI's Real Worst-Case Scenarios

**From *IEEE Spectrum*,
Issue of January 2022:**

We are entering dangerous and uncharted territory with the rise of surveillance and tracking through data, and we have almost no understanding of the potential implications.

“AI doesn’t have to be sentient to kill us all. There are plenty of other scenarios that will wipe us out before sentient AI becomes a problem.” ~ Malcolm Murdock, machine-learning engineer and author of the 2019 novel *The Quantum Price*



What Is Being Human?

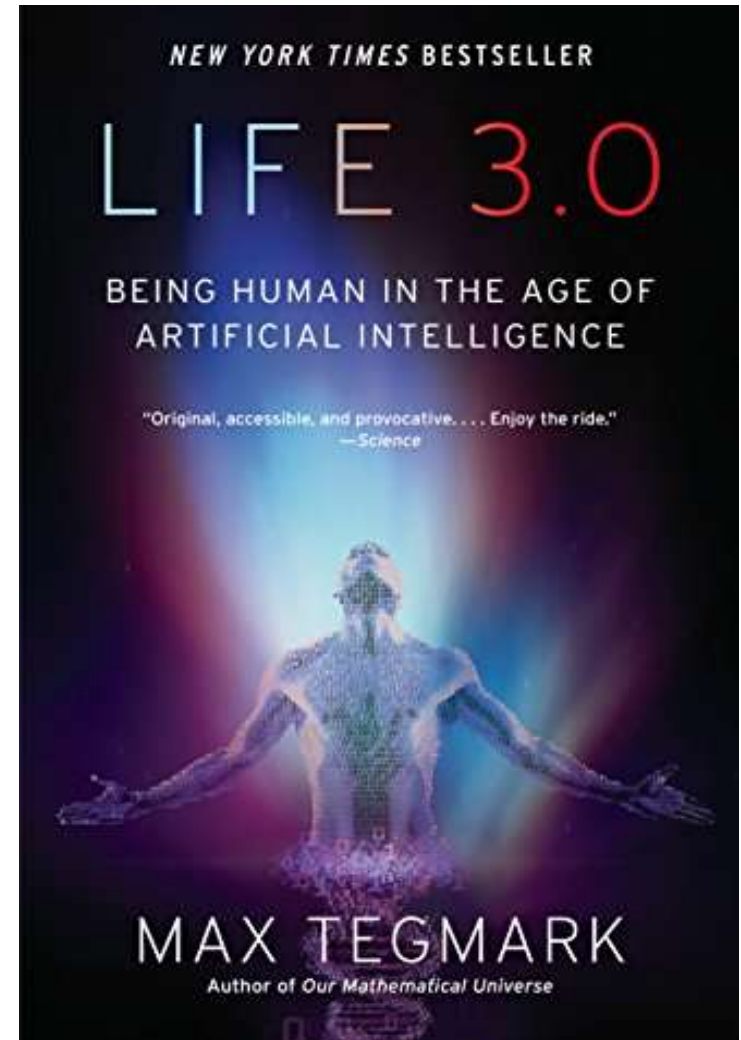
Life 1.0: Hardware & software modified through evolution (slowly)

Life 2.0: Hardware modified slowly, software modified/augmented at an accelerated pace through learning

Life 3.0: Both hardware & software modified/augmented by humans

“... why should our simplest path to a new technology be the one that evolution came up with, constrained by requirements that it be self-assembling, self-repairing, and self-reproducing?

Evolution optimizes strongly for energy efficiency because of limited food supply, not for ease of construction or understanding by human engineers.”



There Are Bad News and Good News

If you're uncomfortable with how much Google knows about you, imagine if it had your brain data as well!

FIRST WIN FOR THE NEURORIGHTS CAMPAIGN

Chile plans to regulate all neurotech and ban the sale of brain data

BY ELIZA STRICKLAND

The Five Neurorights:

From *IEEE Spectrum*,
Issue of January 2022

1. The right to mental privacy
2. The right to personal identity
3. The right to free will
4. The right to equal access to mental augmentation
5. The right to protection from bias

Ethics for Technology in General

Can you withdraw tech support when a bionic body part becomes obsolete?

Clinical ethics:

1. Beneficence
2. Nonmaleficence
3. Autonomy
4. Informed consent
5. Truth-telling
6. Confidentiality
7. Justice



<https://www.karger.com/Article/FullText/509119>

IEEE Spectrum magazine cover, March 2022

Sep. 2026



Realistic Assessment of AI's Benefits & Risks

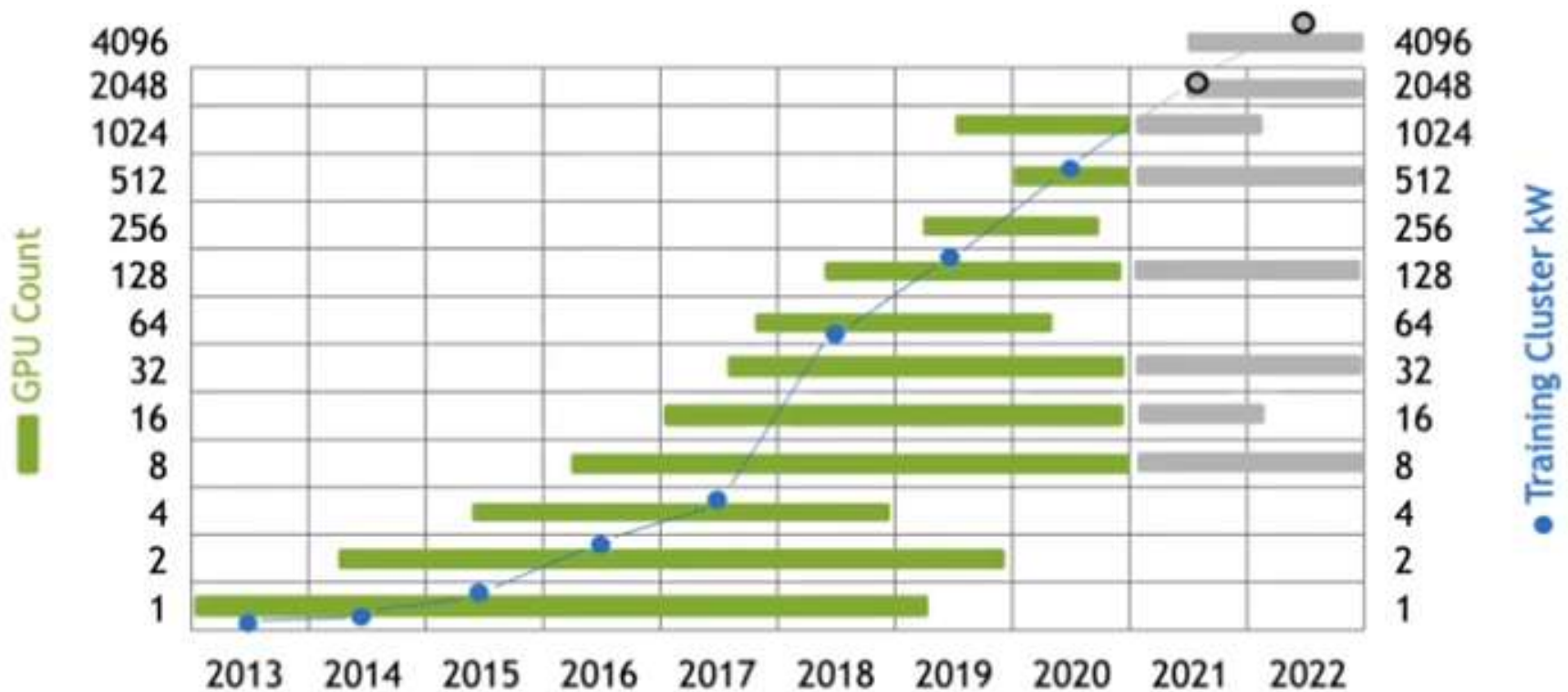


Slide 39

Resources for AI Training

According to Robert Ober (NVIDIA):

- Compute resources for AI training grow ~10x per year
- AI clusters use 100s to 1000s of synchronous GPUs



18-minute talk, "AI Cluster Trends": https://www.youtube.com/watch?v=uIOFL_KCctU

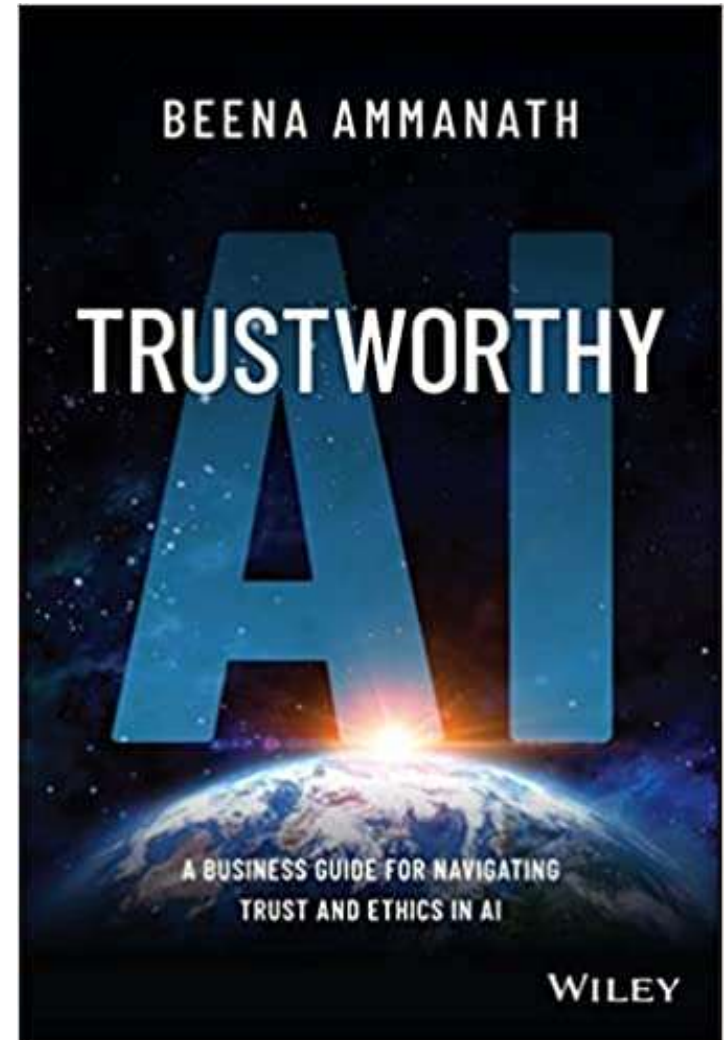
Trustworthy AI Systems

We need to build AI systems that we can trust

Components of trust include:

Transparency (explainability)
Fairness (absence of bias)
Reliability (availability, consistency)
Privacy (avoiding data leakage)
Safety (not doing harm to people)
Robustness (not easily thrown off)

Many researchers have shown that small changes in input data can lead to widely different decisions



Transparent or Explainable AI

Unless humans can understand how AI systems reach their decisions, systems can go awfully wrong

ImageNet competition debacle

Millions of images to be classified into hundreds of categories
There was much cheating by trying the challenge many times and submitting the best results
(like taking an exam many times and keeping the best grade)

Using the wrong criteria, but achieving correct results

Images of forest, possibly with animals in the foreground
Determine which images have animals in them
If the trees are blurred, then there is likely an animal in front

Can AI Systems Be Creative?

Like all the previous questions we dealt with in this talk, it depends (on your definition of creativity)

AI systems can produce works of art

Paintings that are comparable in “quality” to human art
Completion of Beethoven’s unfinished 10th Symphony

AI systems can prove theorems

Help with the solution of the four-color problem
Several new theorems that had eluded mathematicians

AI systems can generate new food recipes

Analyze recipes and combine ingredients in new ways

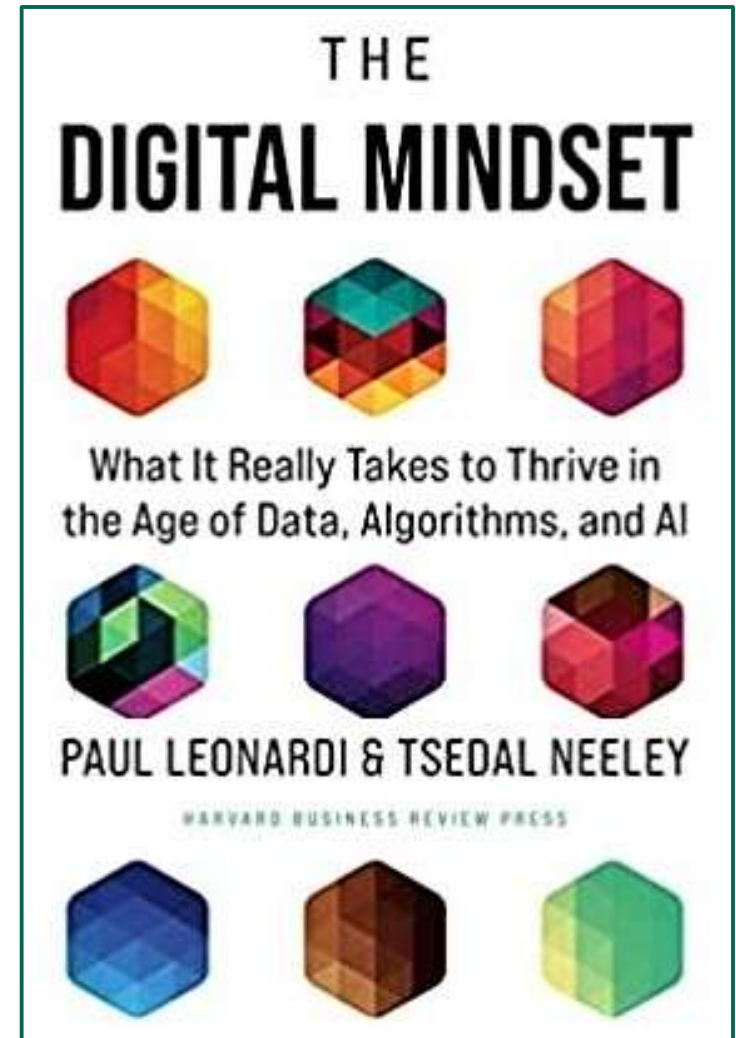
Business Angst for AI and Going Digital

New book, Sep. 2026:
Paul Leonardi (UCSB TMP)
Tsedal Neeley (Harvard Bus. Sch.)

Useful advice to businesses, but ...

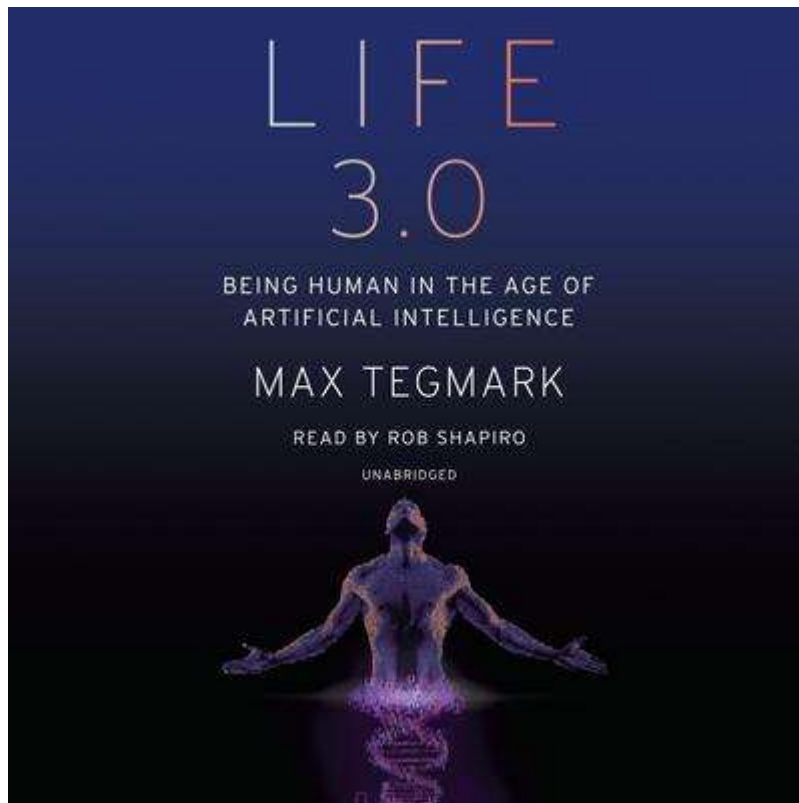
Small businesses may be hurt due to their inability to adapt to the new digital/automation landscape

Large businesses can afford to hire AI/ML specialists, data scientists, software/app developers, and cyber-security experts

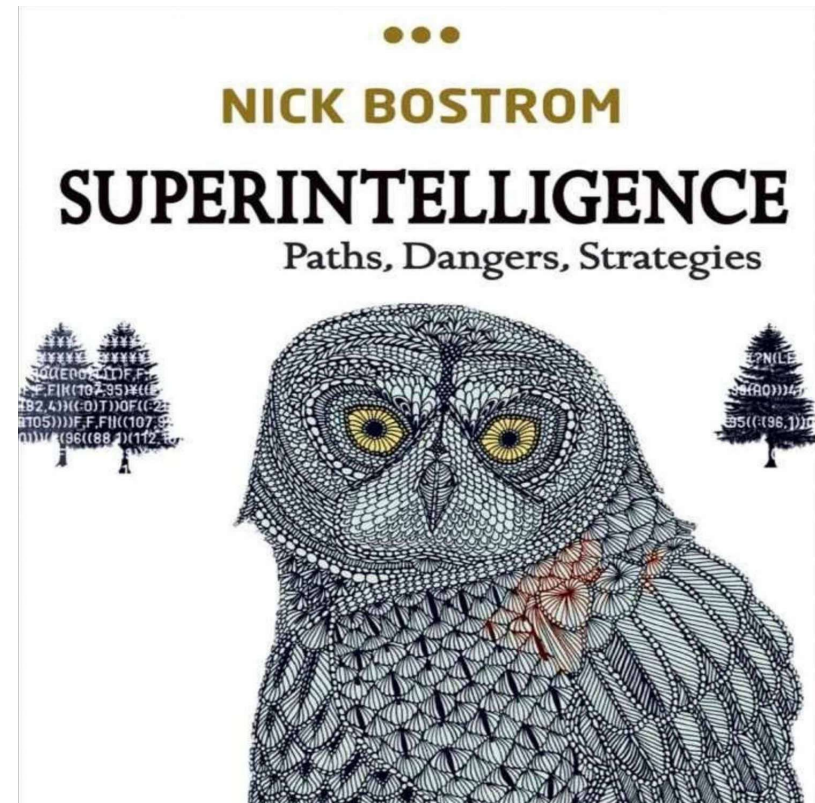


Humans in a World Dominated by AI

Life 3.0: Being Human in the Age of Artificial Intelligence
(Max Tegmark)



Superintelligence: Paths, Dangers, Strategies
(Nick Bostrom)



Summary and the Road Ahead

The Bottom Line:

- AI & ML have both been hyped by researchers & techies
- Separating the wheat from the chaff is quite challenging

Some Problems Being Worked on:

- Using the increasing computational power for better ML
- Developing better neural-network and other models
- Improving joint decision-making by humans + machines
- Ethics of AI & ML; ethics of technology, more generally

Going Forward:

- AI need not be human-like to be useful (factory robots)
- It's not humans vs. machines, but humans + machines
- The singularity isn't near (or if it is, no need to worry)

Further Reading

***Artificial Intelligence: A Guide for Thinking Humans,*
by Melanie Mitchell, Macmillan, 2019**

Table of Contents (16 chs.)

Part I Background

Part II Looking and Seeing

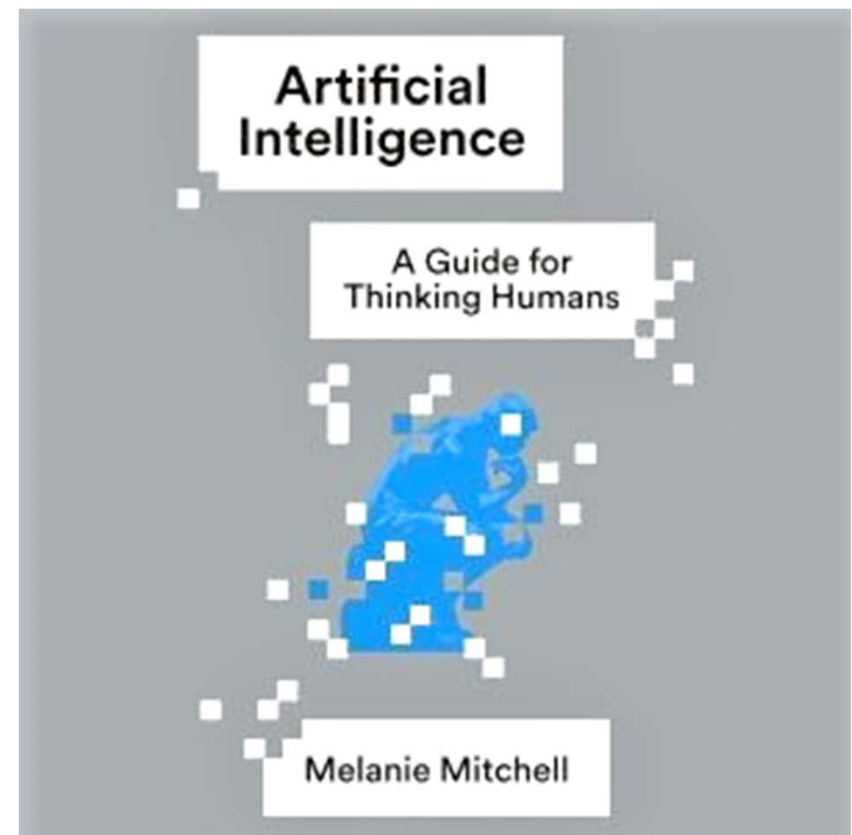
Part III Learning to Play

Part IV Artificial Intelligence

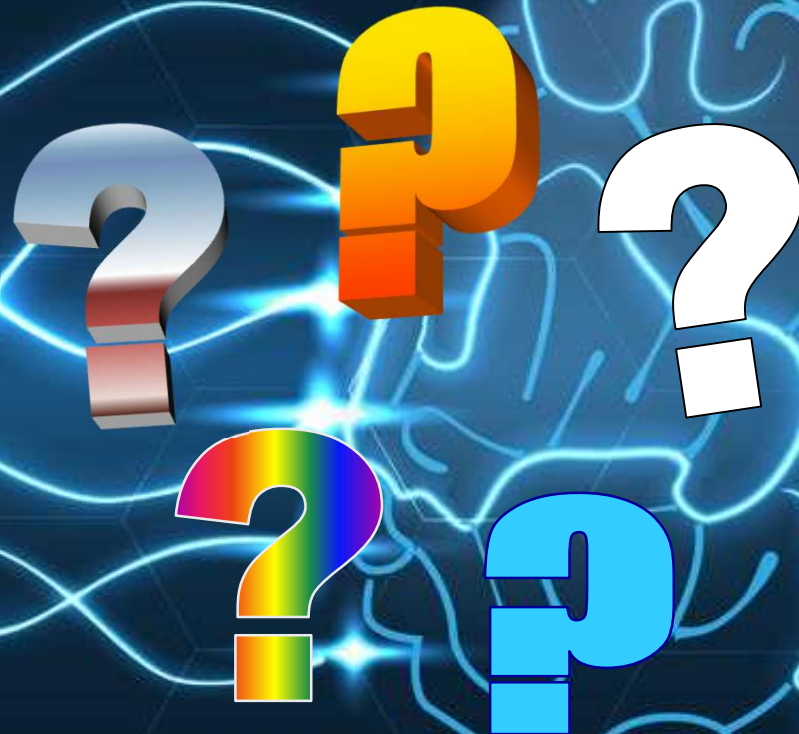
Meets Natural Language

Part V The Barrier of Meaning

***Interweaves stories
about the science of AI
and the people behind it***



Questions or Comments?

A blue rectangular box with three horizontal dashed lines inside, intended for a comment or question.An orange cloud-shaped box with three horizontal dashed lines inside, intended for a comment or question.A yellow speech bubble-shaped box with three horizontal dashed lines inside, intended for a comment or question.A green speech bubble-shaped box with three horizontal dashed lines inside, intended for a comment or question.A purple speech bubble-shaped box with three horizontal dashed lines inside, intended for a comment or question.

parhami@ece.ucsb.edu

<http://www.ece.ucsb.edu/~parhami/>