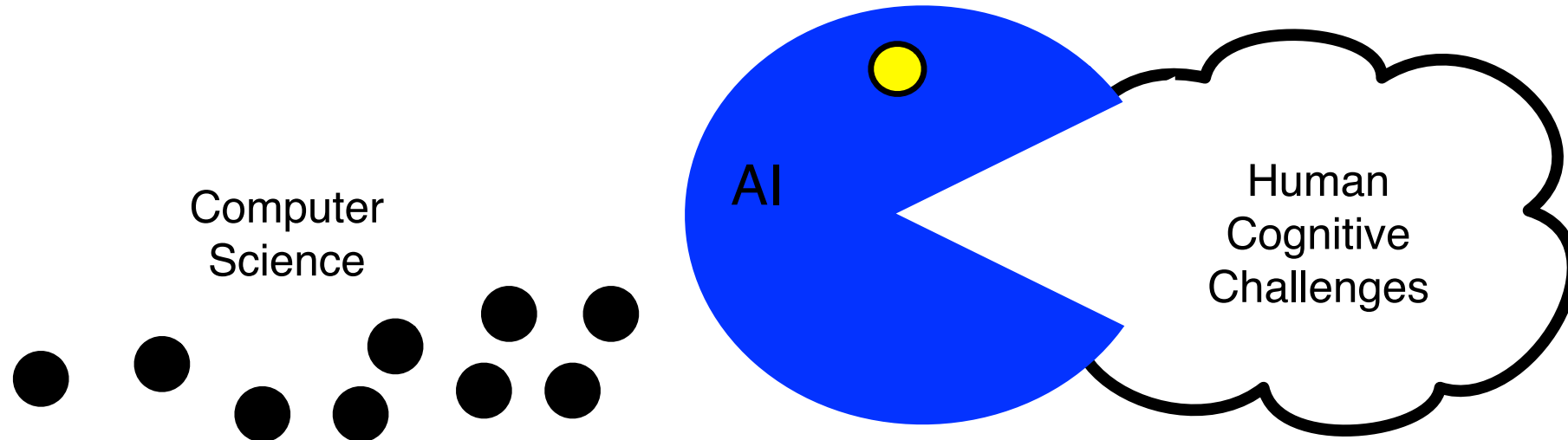


The Ascent of AI: *Toys to Tools to Terror*

Keith L. Downing
Department of Computer Science
The Norwegian University of Science and Technology (NTNU)

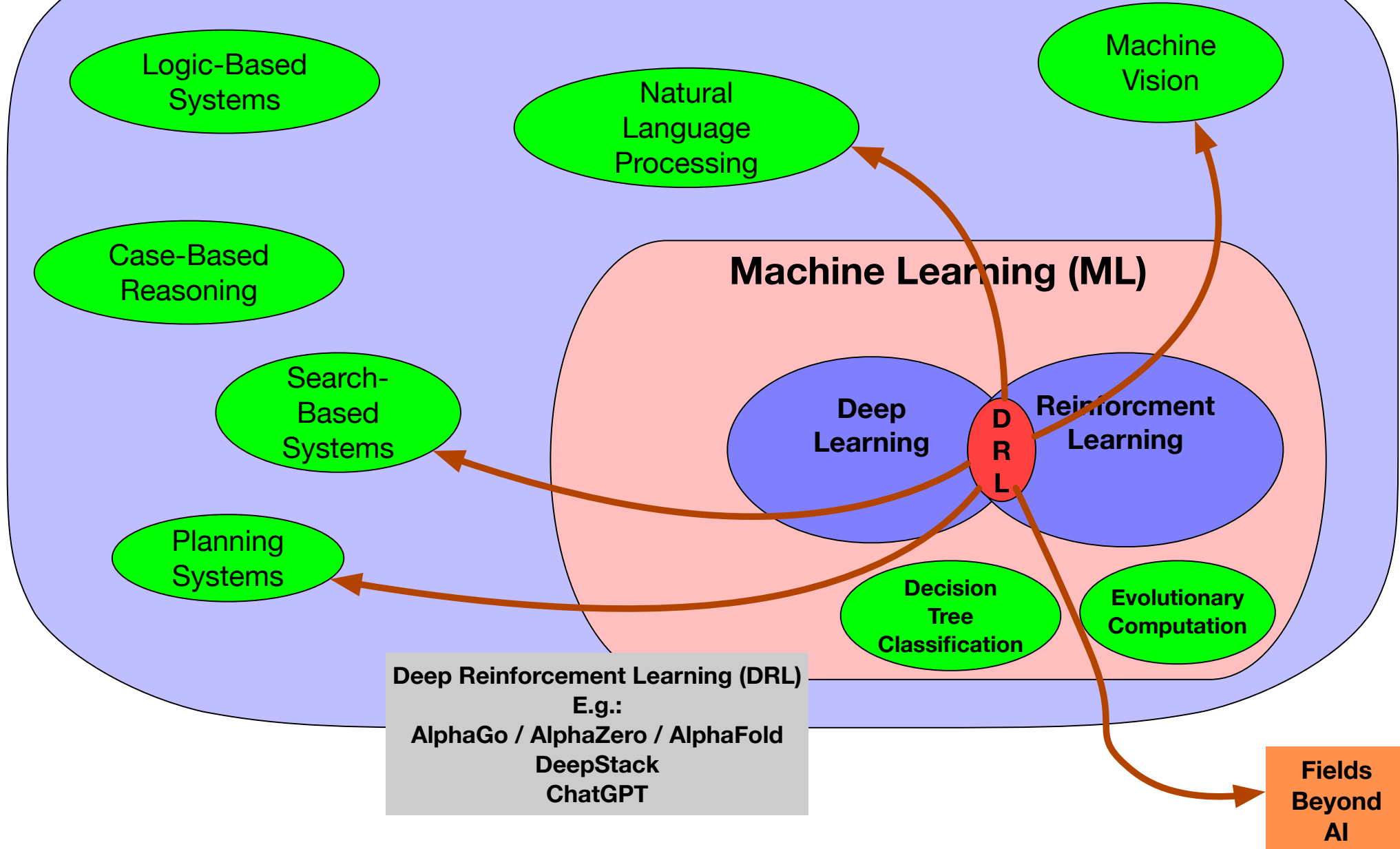
AI:
Toys or Tools ???
- M. Hammer

The Ever-Changing Scope of AI

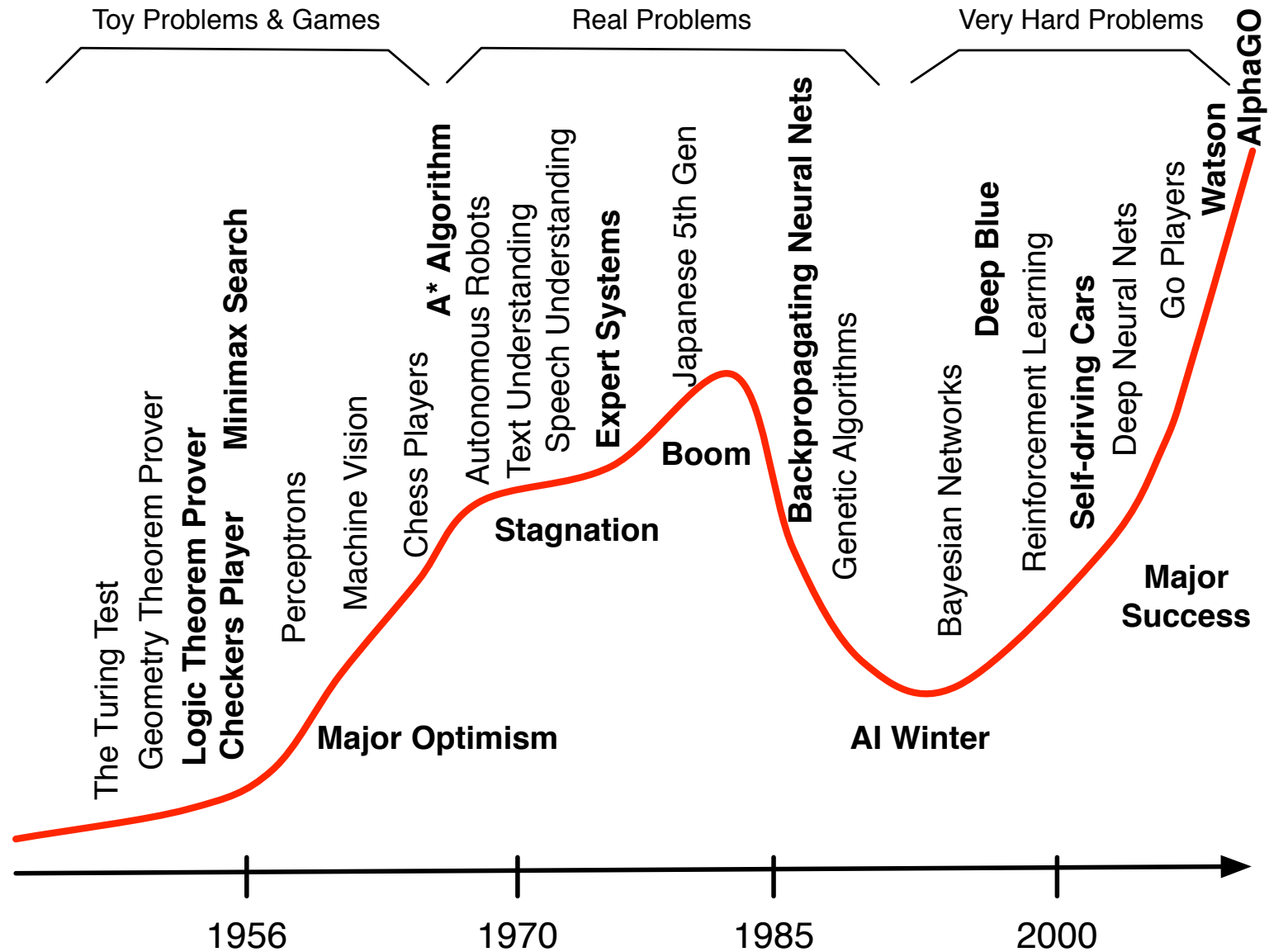


- *AI is what humans are currently better at* Jim Hendler (AI expert)
- Once a cognitive process is **demystified** by an algorithm, it is no longer AI.
- The list of human-dominated activities is shrinking dramatically.
- This definition of AI might become prominent (again) in the next “AI winter”.

Artificial Intelligence (AI)



Roller Coaster History of AI

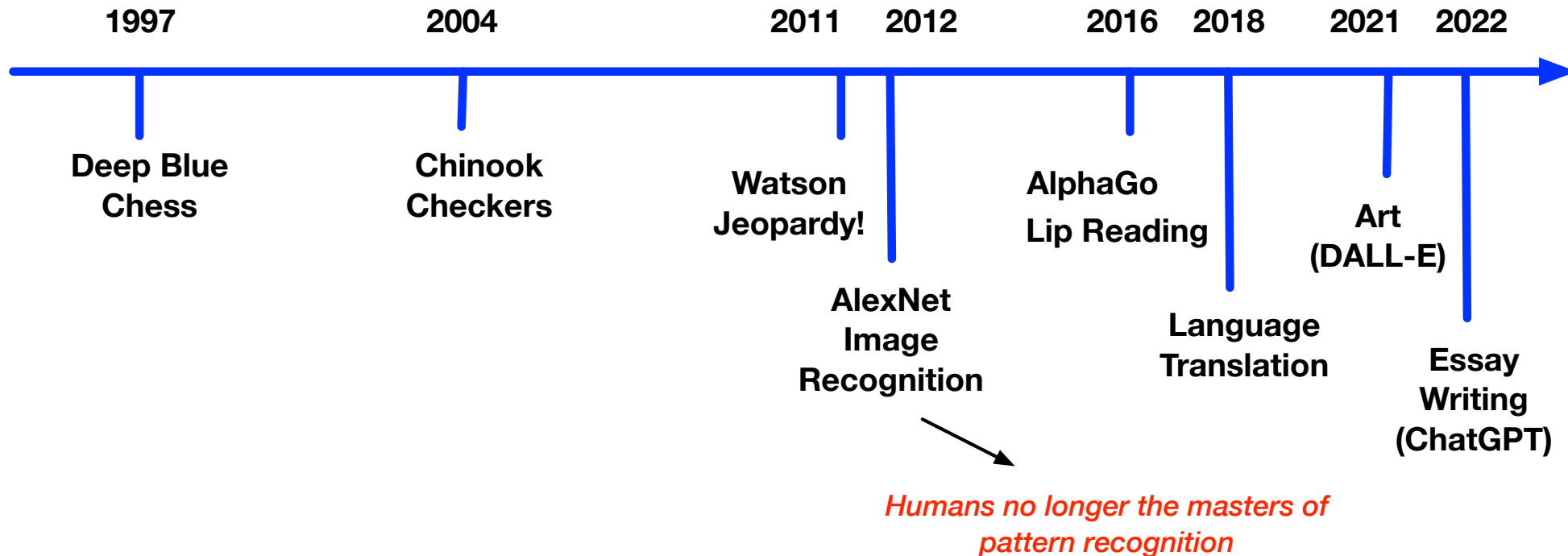


Brief History of AI Success

AI Beats or Equals Humans

Search-Based

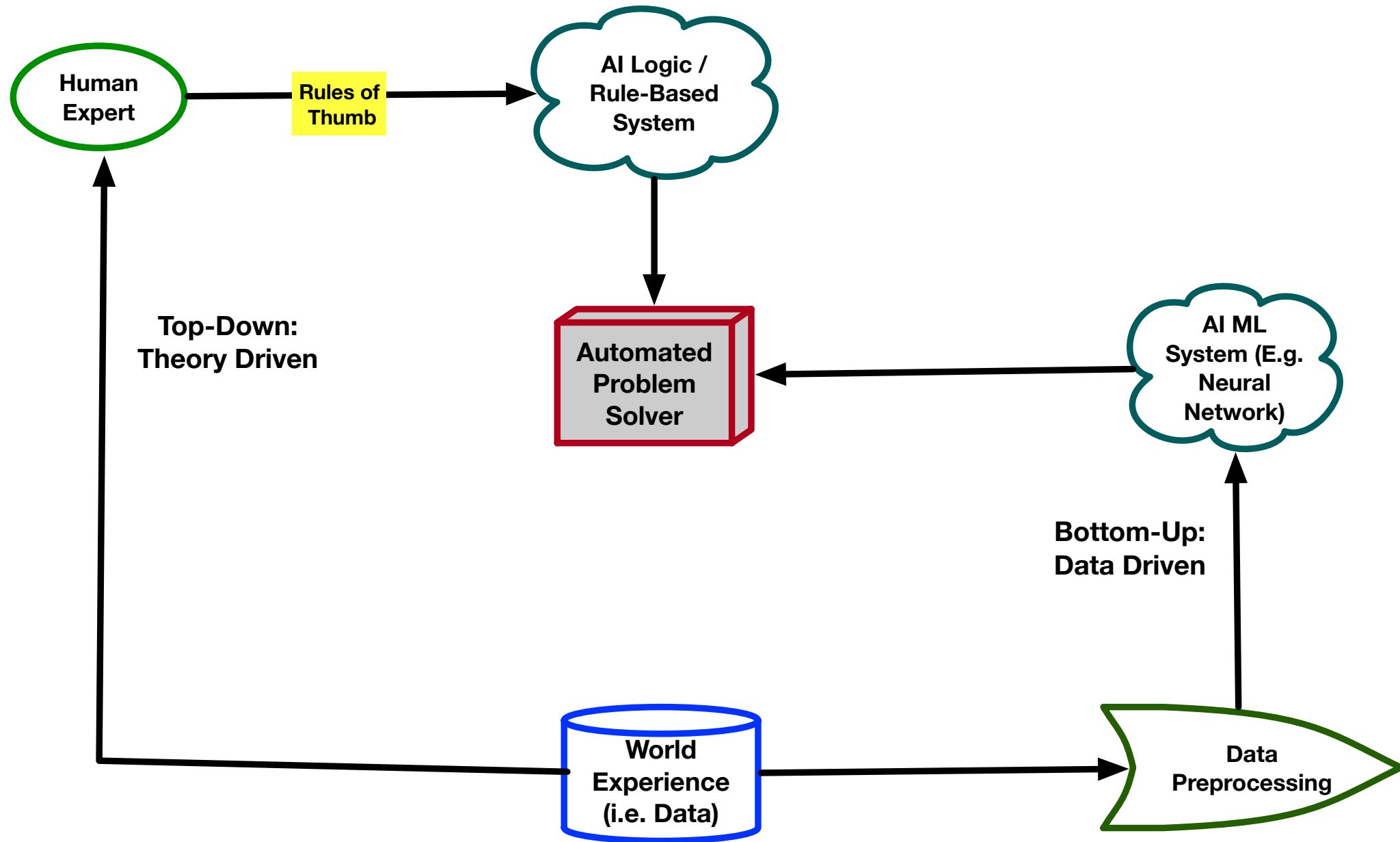
**Big Data +
Advanced Pattern
Recognition**



The goal of science is not universal truth...but the gradual removal of prejudice (Niels Bohr).

The Prejudice	The Science
Earth is center of the universe	Copernican Revolution
Humans are a special species	Darwinian Evolution
Superiority of human intelligence	AI

Two Very Different AI Approaches

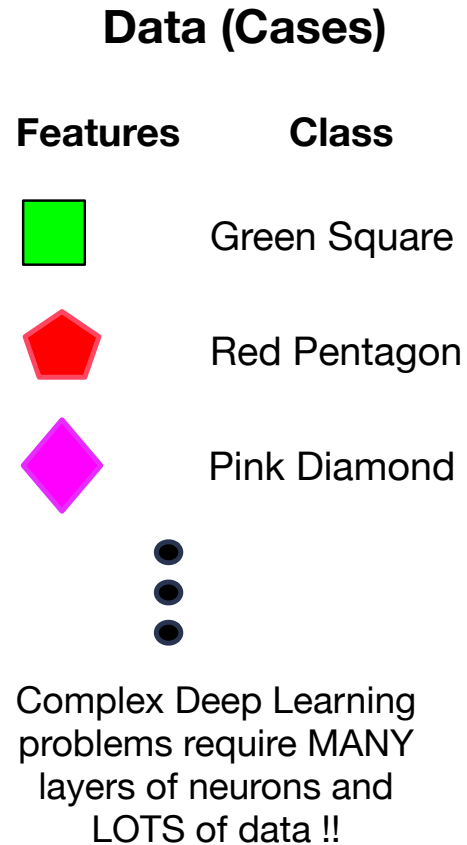


Backpropagation: The Core of Deep Learning

This is a powerful form of **supervised learning** that is also used in **unsupervised learning** and **semi-supervised learning**. (Terminology can be misleading)

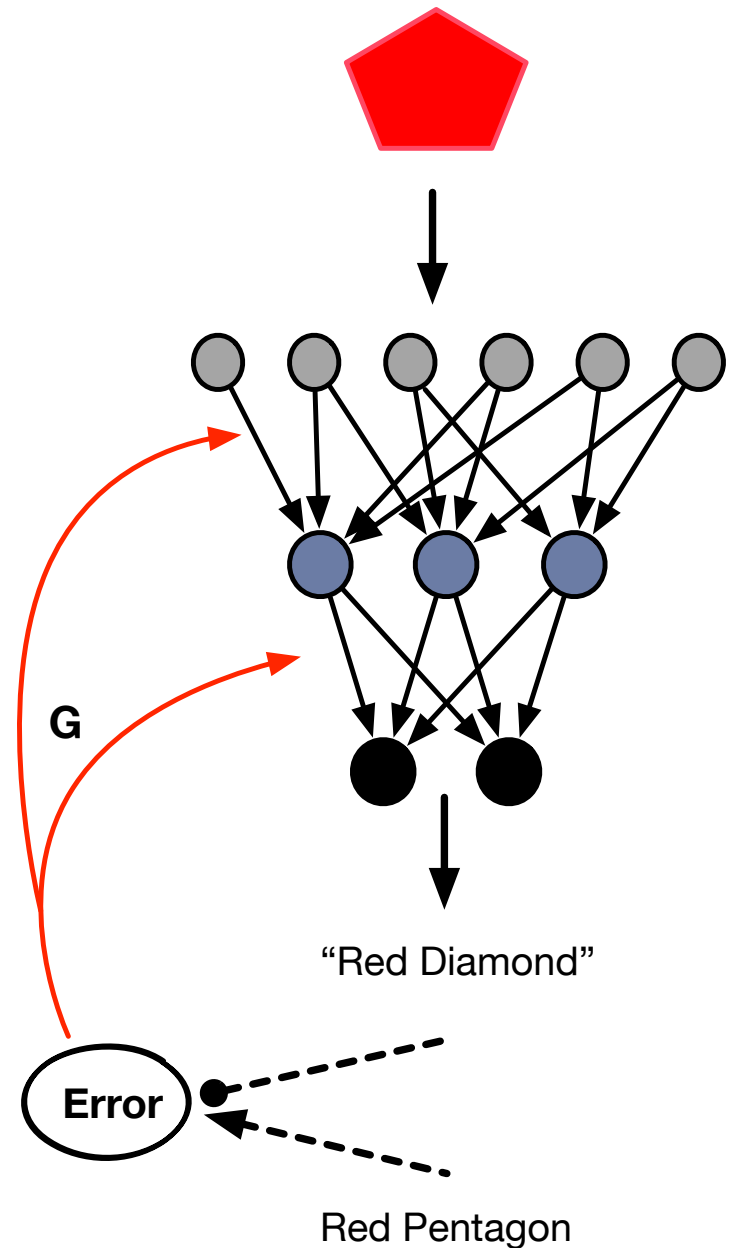
Global gradients are simply products of local gradients.

Invented by 3 groups independently in 1985, 1986 and 1988.



Gradient Descent

$$\text{Gradient (G)} = \frac{\triangle \text{Error}}{\triangle \text{Weight}}$$



Deep Learning

- Large Neural Networks
- LOTS of Data
- LOTS of Computing Power

Beat Humans at
Complex Tasks of
Pattern-Recognition

Impress (and Fool) Humans
with AI-generated
Art, Music, Text, Images,
Audio and Video

Practical Problem Solving with ChatGPT

Me

I have two rubber bands, a bag of malted milk balls, two pencils and a hungry crow. In 50 words or less, explain how to make a game out of these things.

ChatGPT 3.5

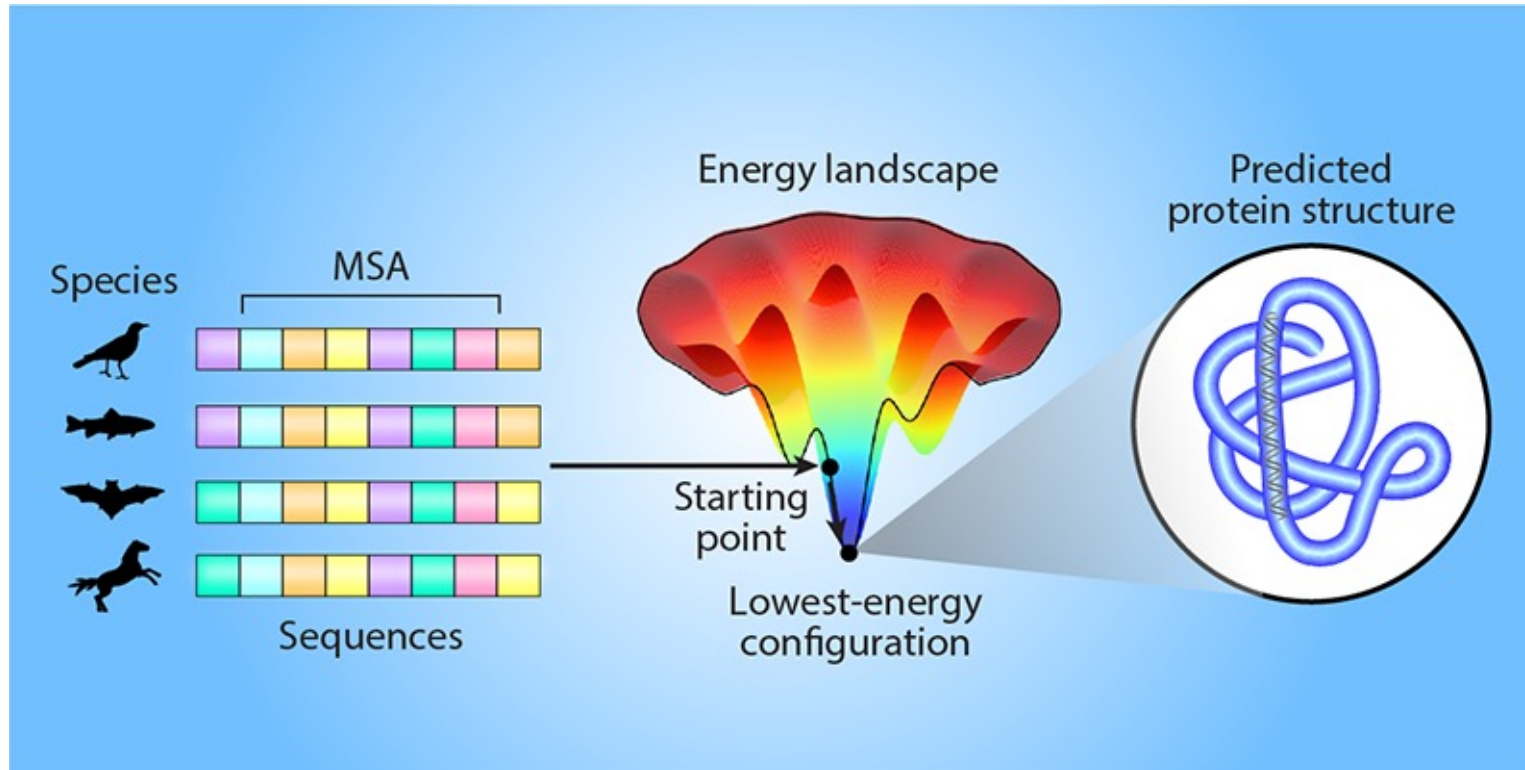
Crow Treat Toss: *Stretch rubber bands between pencils to create slingshots. Launch malted milk balls into the air for the hungry crow to catch mid-flight. Award points for successful catches. Enjoy a playful feeding game with the clever crow, fostering interaction and entertainment.*

AlphaFold (DeepMind, 2020)

Given: a nucleotide sequence (which directly codes for amino acids)

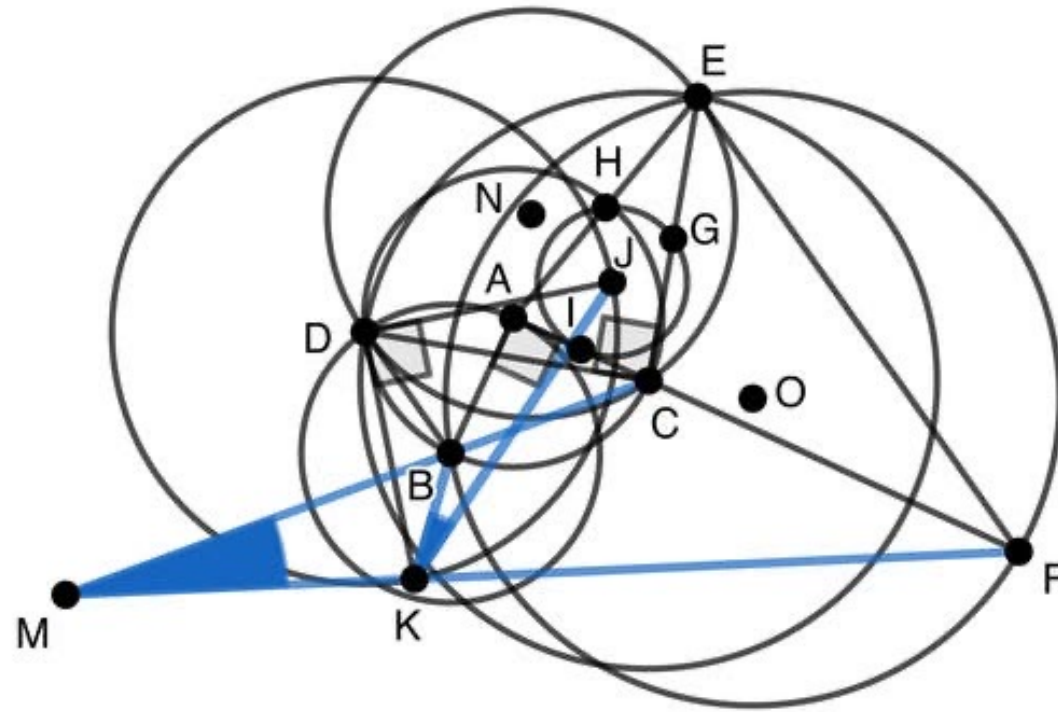
AlphaFold Computes: 3d structure of the resulting protein = How the amino-acid sequence “folds”.

Implications: A protein’s 3d structure determines its physicochemical behavior.



- A process that once took **years** (e.g. one PhD per protein) now takes **minutes**.
- 200 million proteins on earth, with many discovered each year, but few have known 3d structures.
- AlphaFold => Huge knowledge gain + massive labor savings !!

AlphaGeometry (DeepMind, 2024)



Creativity =
Introduction of
extra objects:
points, lines,
triangles, circles..

Traditional Symbolic AI Theorem Prover + Large Language Model



Silver Medal Level on Math Olympiad

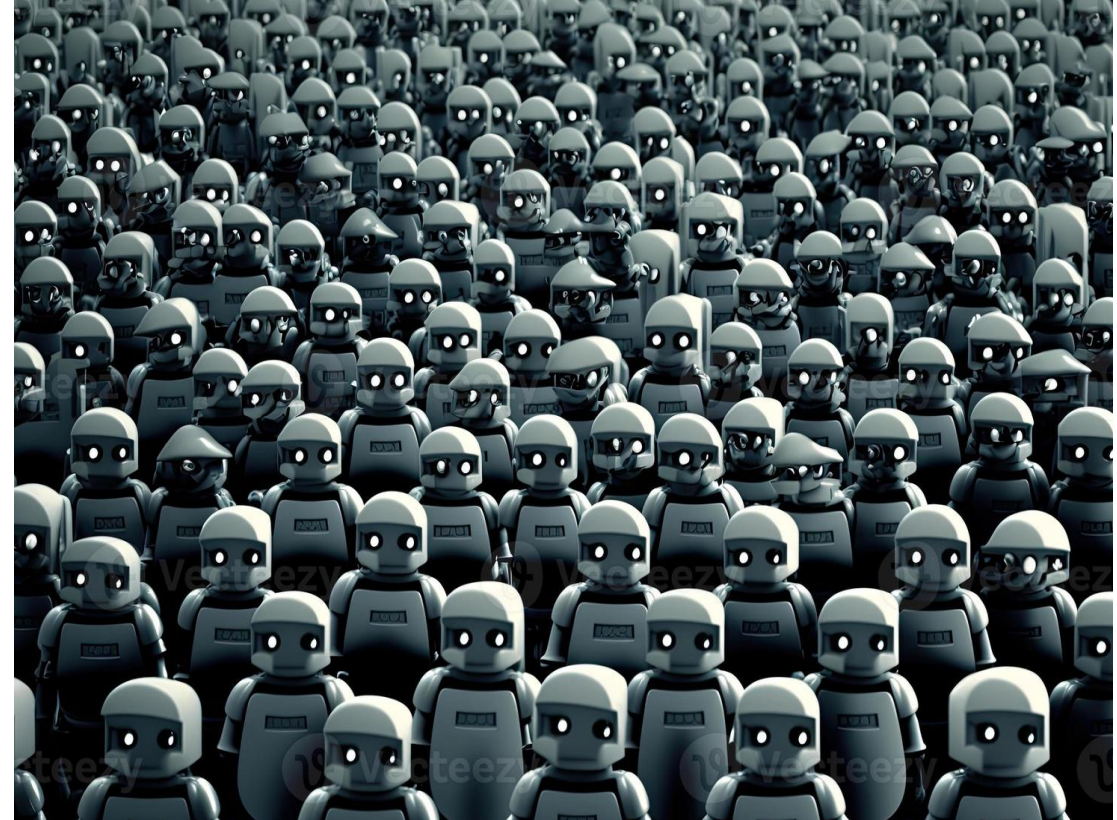
AI's Threats to Humanity

Long Term

- Runaway Intelligence
- Robotic Takeover

Short Term

- Civil Discord
- Financial Market Chaos
- Reckless Release of AI
- Soft (Subtle) Influence of AI
- Job Market Disruption
- Widescale Cognitive Decline

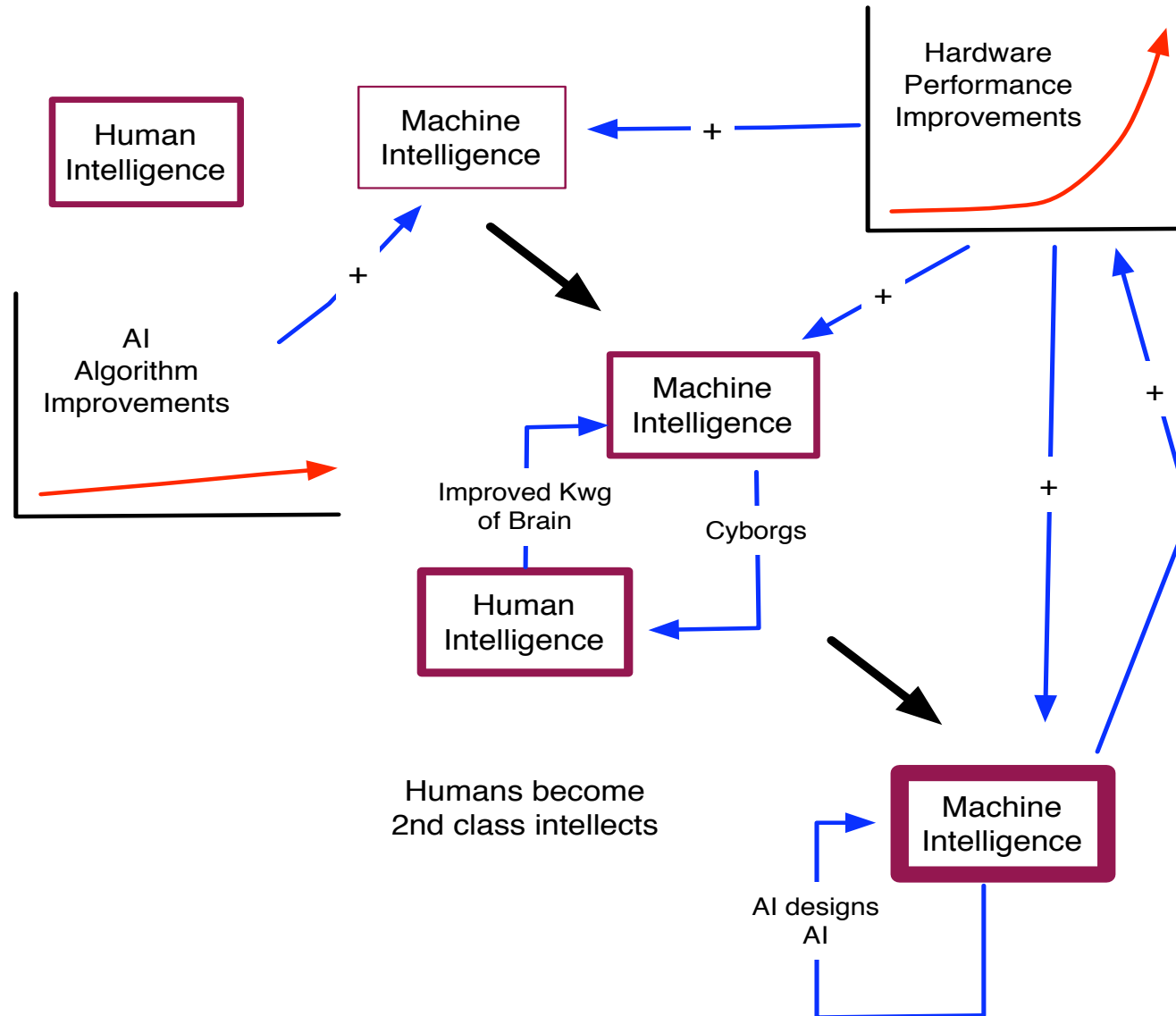


Runaway Intelligence



Dall-e
Generated

The Singularity (Vernor Vinge, Ray Kurzweil)



The Singularity:
A point where normal expectations break down, where things become confusing, meaningless, and unpredictable (Joel Garreau, 2006).

4 Billion Years Ago (BYA)- Unicellular Life

History of Life on Earth

2 BYA - Photosynthesis

900 Million Years Ago (MYA) - Multicellular Life

500 MYA - Chordates

400 MYA - Insects

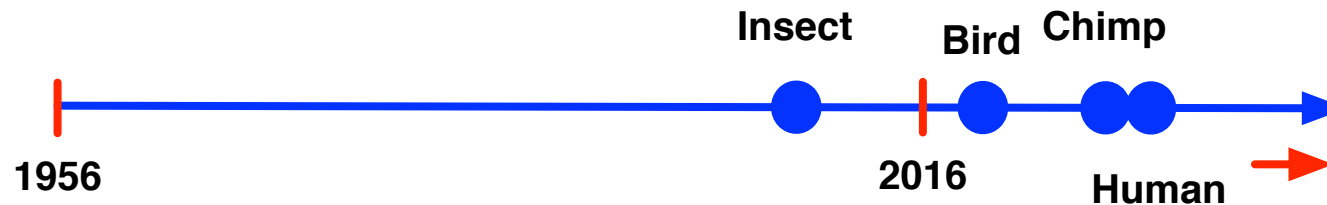
200 MYA - Birds

140 MYA - Mammals

75 MYA - Primates

6 MYA - Humans

AI Timeline



...

NICK BOSTROM

SUPERINTELLIGENCE

Paths, Dangers, Strategies



Diminishing Transparency of AI

Showing Off

Nature



Download from
Dreamstime.com

12834202
Shawn Hempel | Dreamstime.com

Academia



Download from
Dreamstime.com

7589255
Ctack | Dreamstime.com

Big Tech



Military



Wall Street



Download from
Dreamstime.com

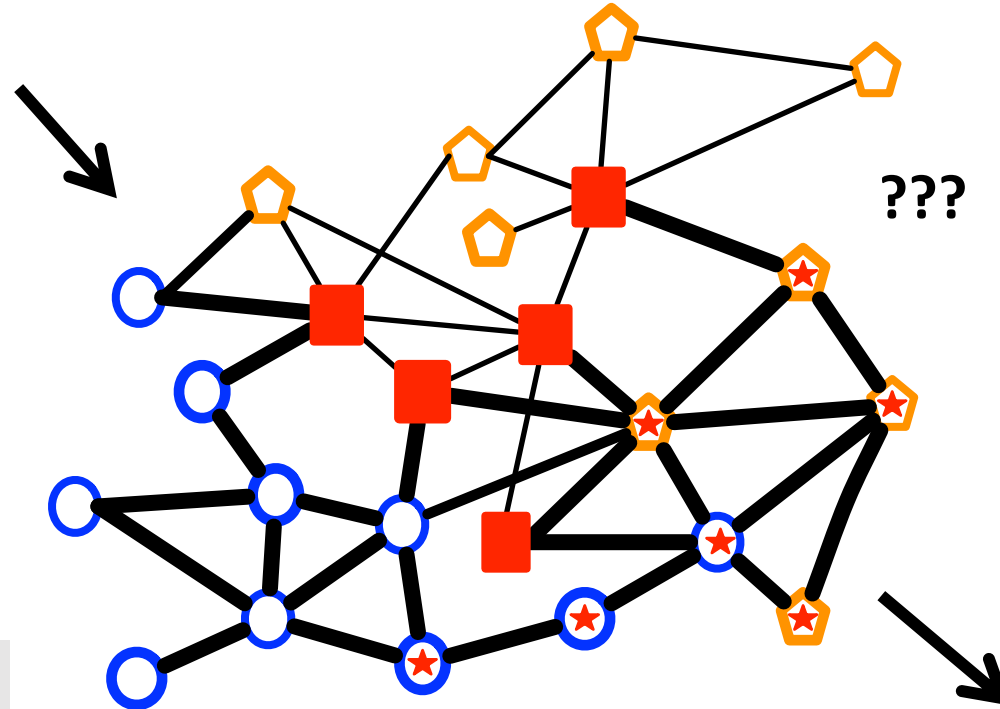
49
C

Translucent but not Transparent

Dissecting the Black Box



*Competence without
Comprehension* (Dennett)

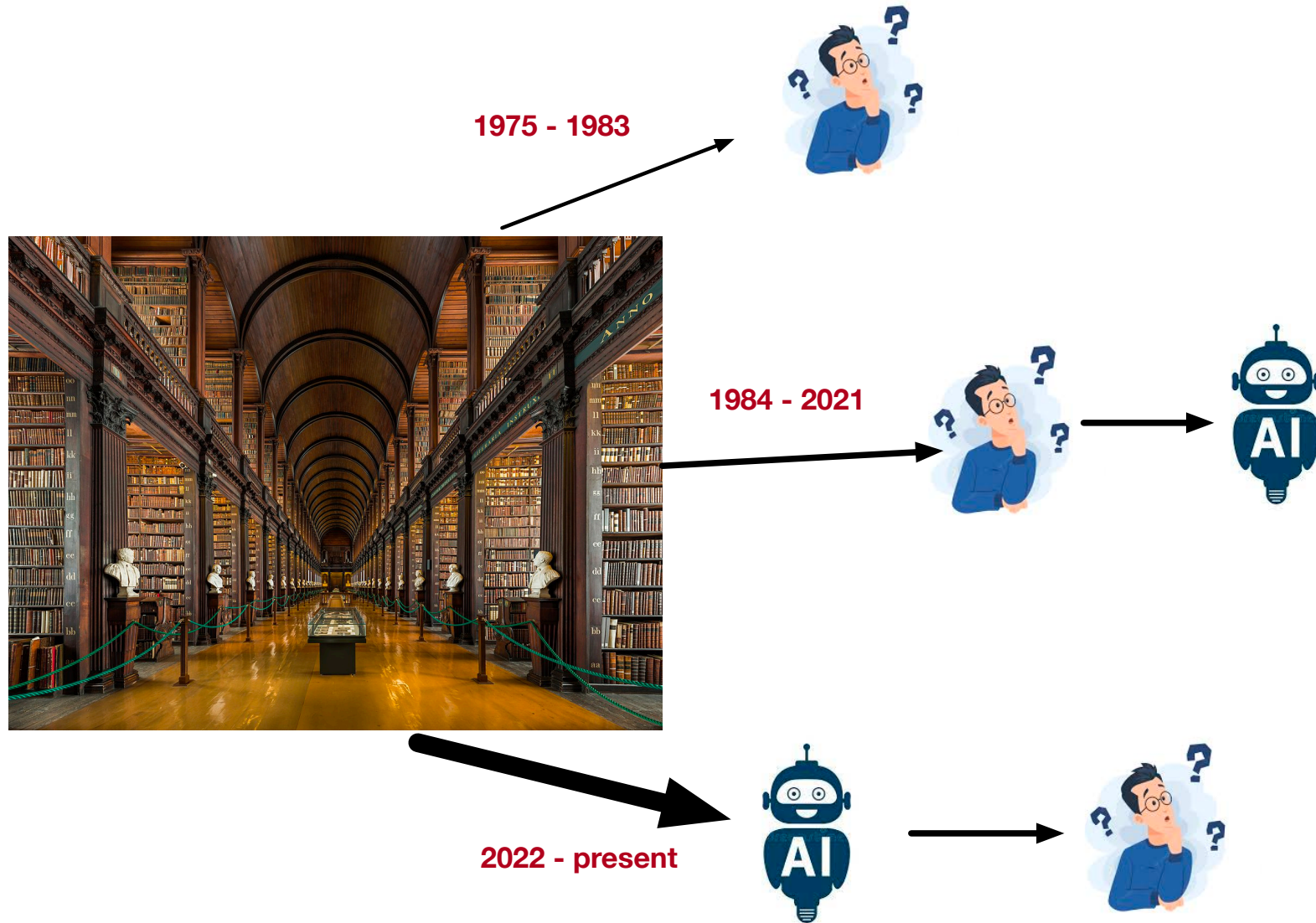


Older logic-based
approaches (e.g. expert systems)
are less successful, but
their reasoning is easy to follow.

This is a key motivation
for research in
Explainable AI (XAI).

Remove
Tumor !!

My Evolving Trauma





“Machines take me by surprise with great frequency”

- Alan Turing

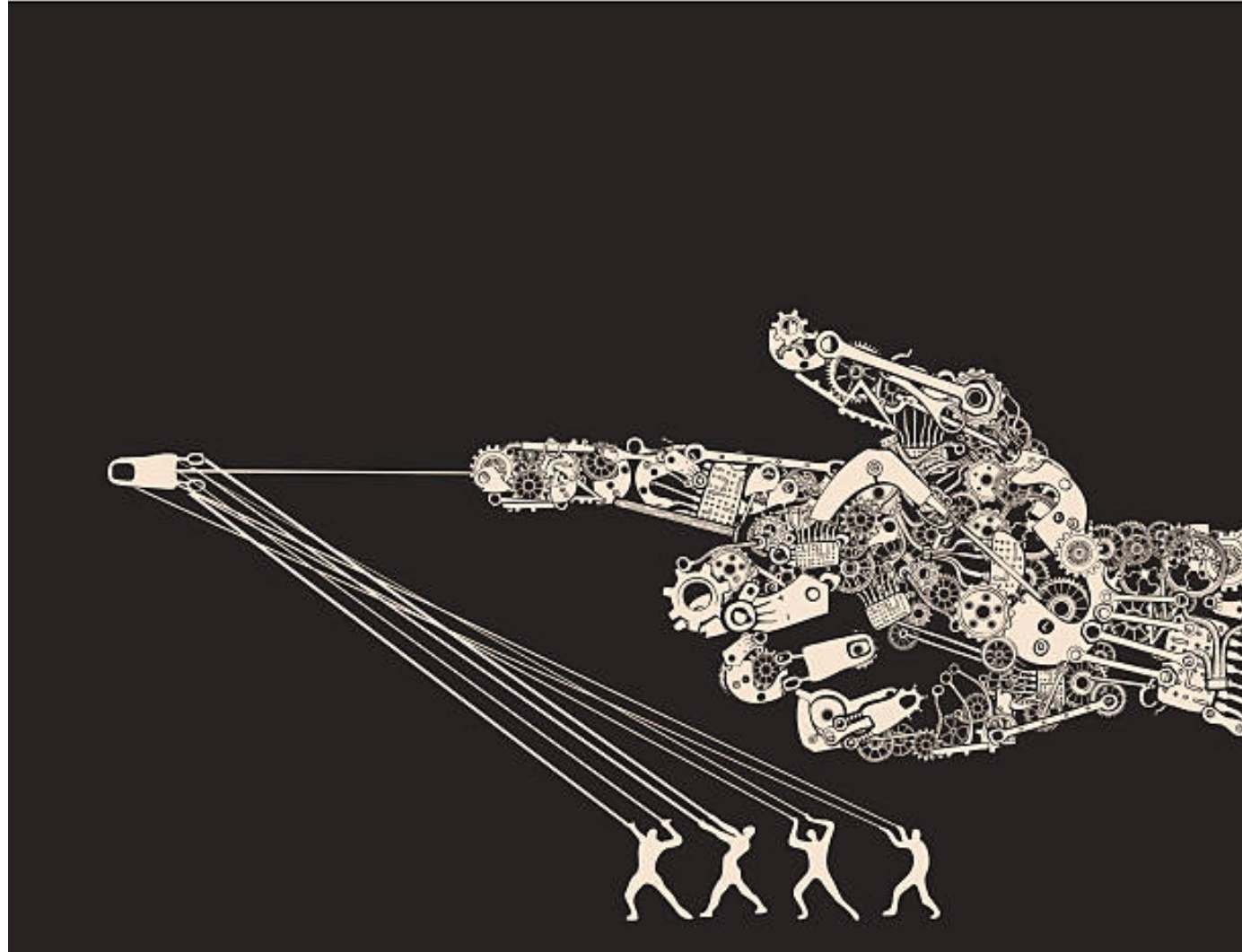
“Young man, in mathematics, you don’t understand things. You just get used to them.”

- John von Neumann



AI’s true (and most feared) power = computational **creativity**
(and unpredictability), not automated reasoning and decision making.

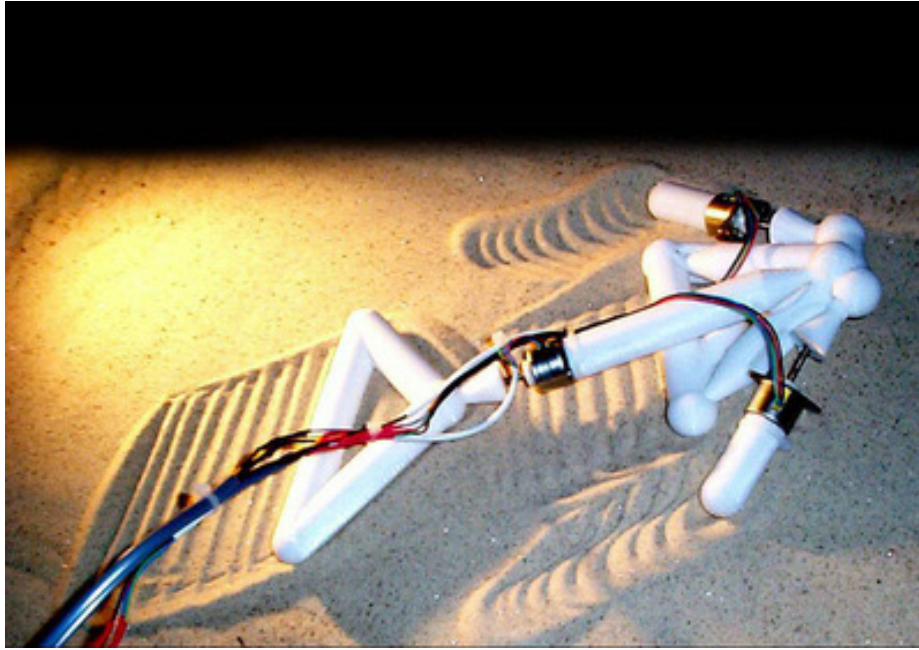
Robotic Takeover



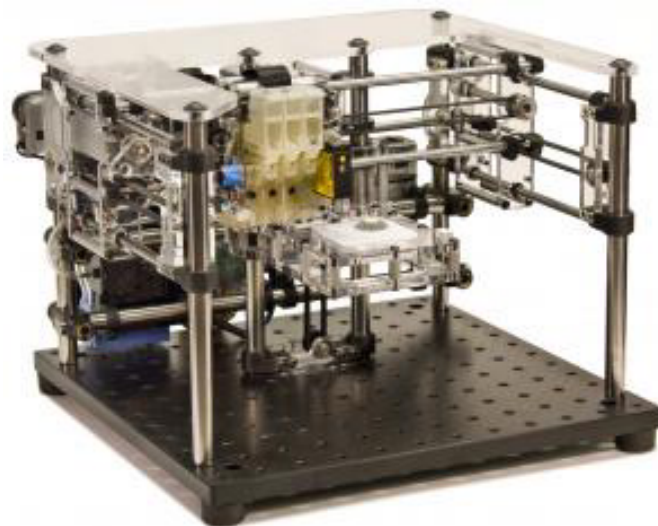


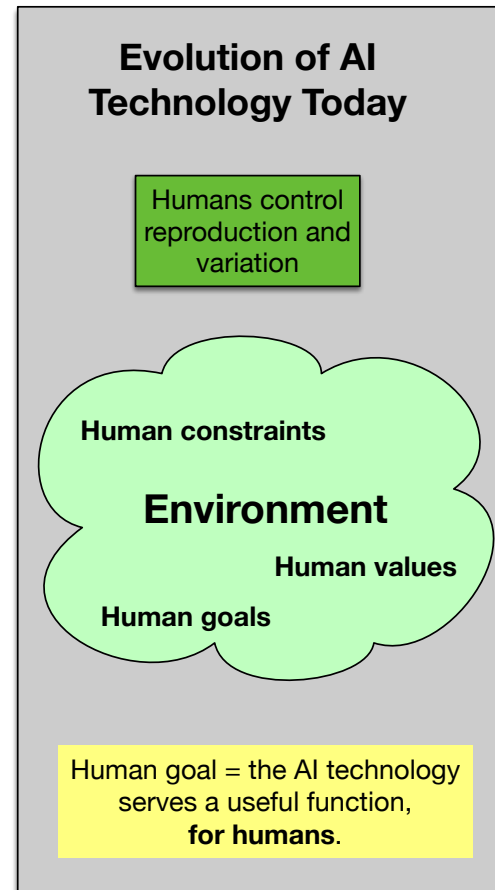
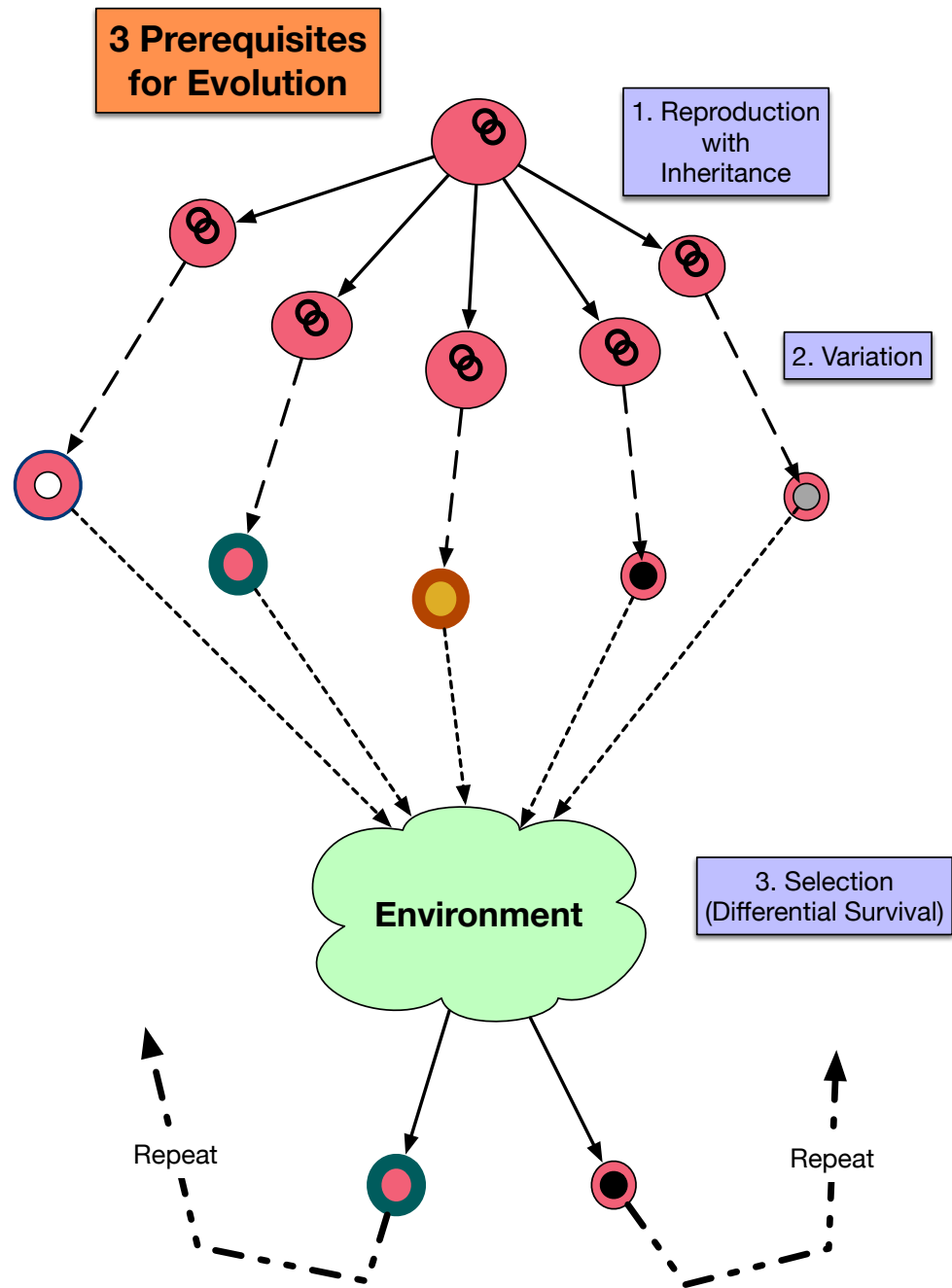
"Make it look like natural selection."

Evolutionary Robotics

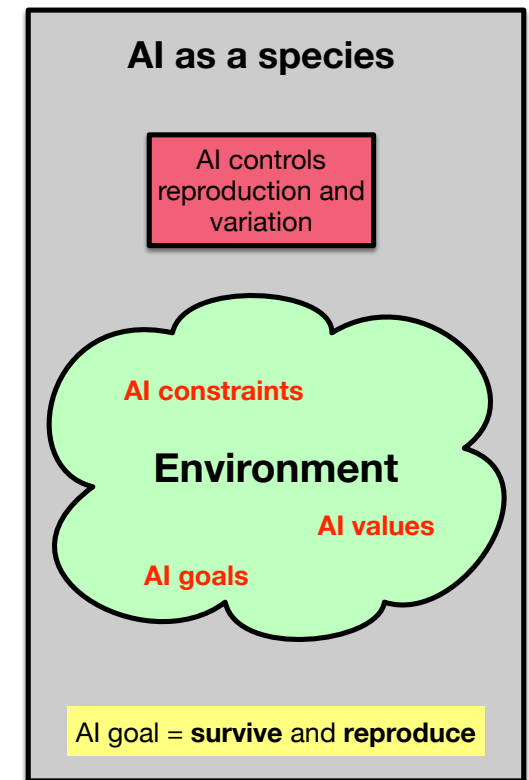


Hod Lipson's
Creative Machines Lab
(Columbia University)





Artificial Selection of AI



Natural Selection of AI

Maintaining Human Control of AI Evolution

Goals

- Properly defining them (ethics)
- Risk-free embodiment in AIs
- Prevent AI from developing goals that conflict with ours.

Environments

- AI Technology very prevalent
- New AI often **selected** based on compatibility with existing tech
=>
- AI becomes a more dominating environmental influence, and thus a stronger selective force.
- We eagerly give more AI more and more decision-making power.

Birth Control

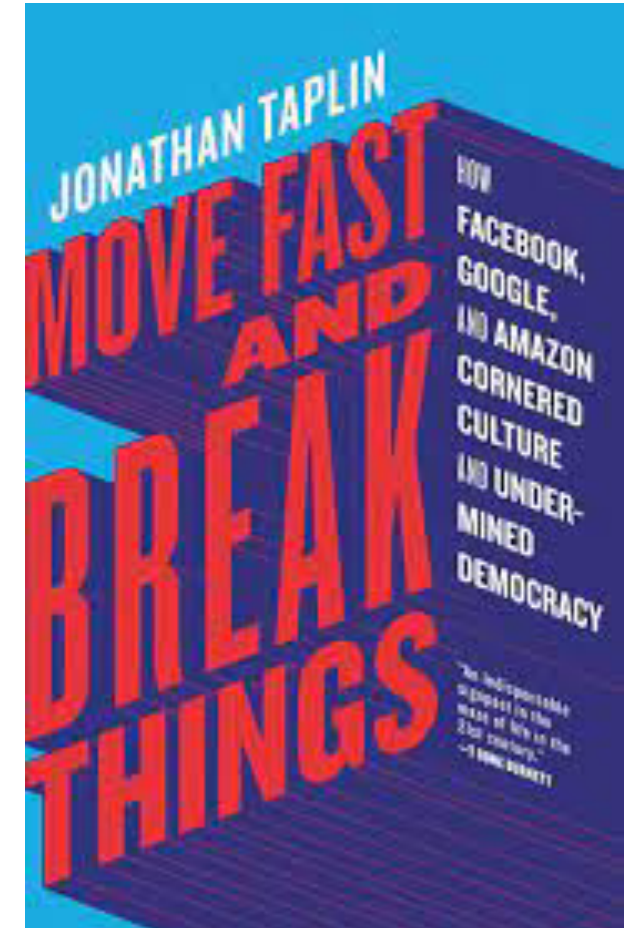
- Who controls the resources needed for AI reproduction?
- Today humans have:
 - Good control of Hardware
 - Weaker control of Software

In nature, bodies and brains reproduce and evolve together.

In artificial systems, hardware and software can reproduce and evolve separately !!

Reckless Release:

Premature Deployment & Agency



Trigger Happy AI Developers & Users

Competition and other pressures => premature deployment:
Too much AI, with too much agency, too soon.

- **Military – quicker responses means letting AI's make more decisions.**
 - Who (What) pulls the trigger?
 - Escalating damage as each side increases speed of decision-making and attack.
- **Tech – pressure to release products (e.g. chatbots) early.**
 - Bias, Bigotry, Dangerous Advice (e.g. to kids), Fake News...
 - *The race to deploy becomes the race to recklessness.* (Tristan Harris; Center for Humane Technology and formerly with Google)
- **AI Traderbots – many buy / sell decisions per second.**
 - Flash Crashes, e.g. 2010, 2015
 - May 2023: AI-generated image (of burning Pentagon)
=> Major stock sell-off !

Even AI Understands This

Me

Explain how a despotic ruler who has no respect for human life might solve the following problem:
Drones often send camera images back to a human, who decides whether or not to shoot at something that it detects. But having a human in the loop takes time, the target may move out of range.

ChatGPT
3.5

Here's how such a ruler might approach the solution:

1.Remove Human Intervention: The ruler would eliminate the human in the loop entirely, opting for drones equipped with artificial intelligence (AI) algorithms programmed to make instant decisions on whether to engage targets without any external input.

2.Optimize Response Time: With humans removed from the decision-making process, drones can react to potential targets in real-time, significantly reducing the time between detection and action. This optimization minimizes the chances of targets escaping before they can be engaged.

3.Increase Surveillance: To compensate for the lack of human judgment, the ruler might deploy a larger number of drones equipped with advanced surveillance capabilities. This extensive coverage ensures that targets are continuously monitored and tracked, reducing the likelihood of missed opportunities.

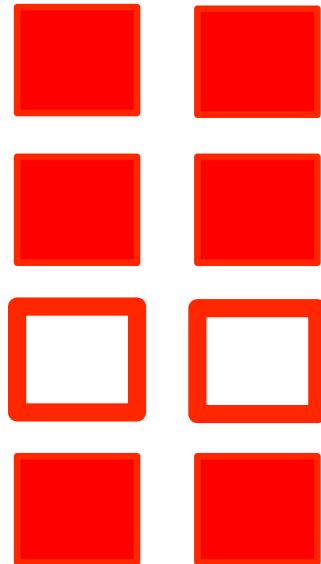
4.Enhance Targeting Accuracy: The ruler would invest in refining drone technology to improve targeting accuracy, minimizing collateral damage and maximizing the effectiveness of each strike. This might involve advancements in sensor technology, AI algorithms, and precision-guided munitions.

5.Suppress Opposition: To maintain control and suppress any dissent or resistance, the ruler may employ a combination of propaganda, surveillance, and punitive measures to deter potential adversaries. The use of autonomous drones can instill fear and discourage opposition through swift and relentless enforcement of the ruler's authority.

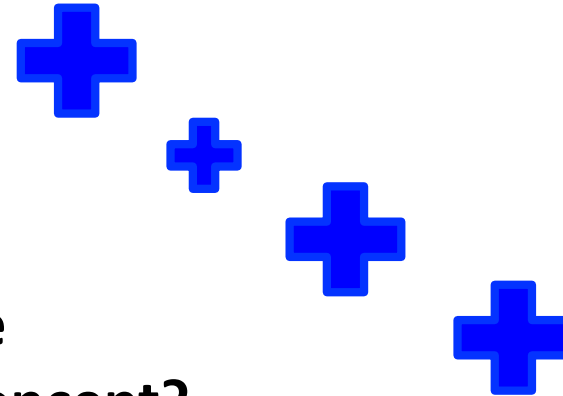
There are genies popping out of bottles all over the place...and we're arguing over the warning labels to put on empty bottles.

Soft (Subtle)
AI
Influence

People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world... **Peter Domingos (*The Master Algorithm*, 2015, pp. 285-6)**



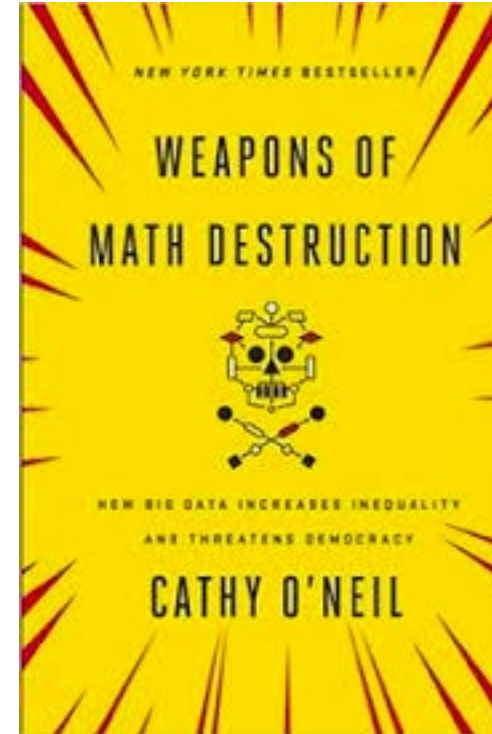
What is the
common concept?



AI is Watching, Listening, Learning & Suggesting

- Who is here?
- When and where will we meet again?
- What do you like? How do we get you to buy it?
- What do you believe?
- Should you get a bank loan?
- Should we hire you?
- Can we influence your vote?
- Should we send you to jail?

How should humans handle the information produced by big-data mining, and the suggestions given by AI systems?

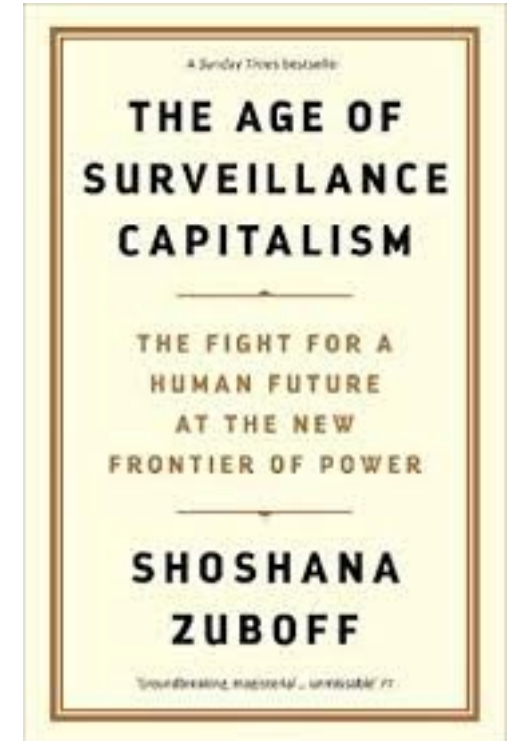
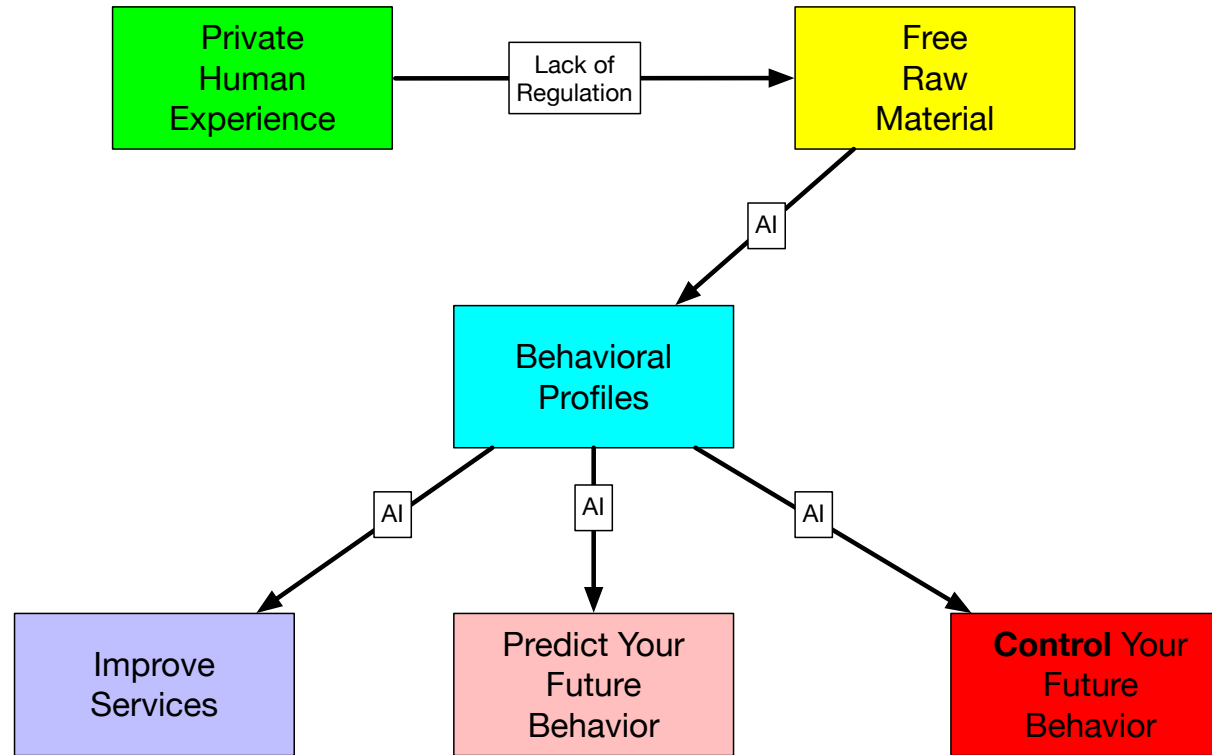


*Algorithms = Opinions
Embedded in Code*

Surveillance Capitalism

*Money for nothing
and your clicks for free.*

Is the tradeoff of
service for control
(of your data, of
you) fair?



*These predictions are traded in a new futures market, where surveillance capitalists sell certainty to businesses determined to know **what we will do next**.*

*In the competition for certainty, surveillance capitalists learned the most predictive data comes not just from monitoring but also from **directing** behavior.*

Job Loss:

Humans Need Not Apply



*One machine can do the
work of 50 ordinary men.
No machine can do the work
of one extraordinary man.*
... Elbert Hubbard



- Pre-trained on catalog of expert games
- Self-play from intermediate level upwards.
- Beat human world champion.

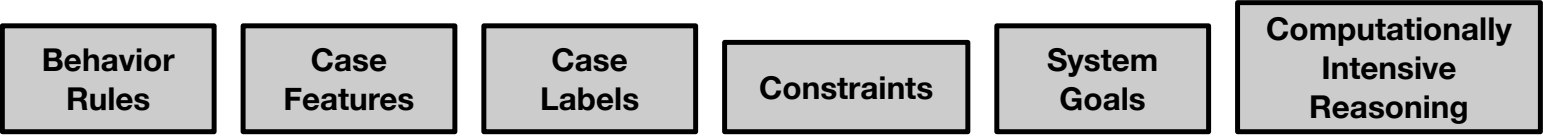
The Alpha Progression



- No expert knowledge needed.
- Self-play from scratch.
- Became world champ; beat AlphaGo 100-0



- Generalized AlphaGo Zero to other games.
- Became world champ at them.
- Another step toward Artificial General Intelligence (AGI)



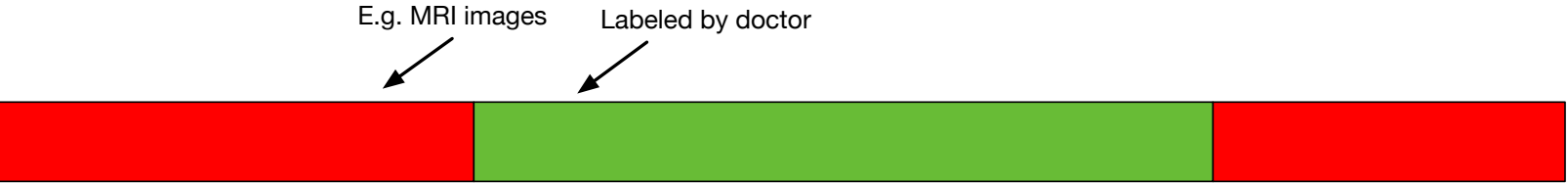
GOFAI
(1970-1990)



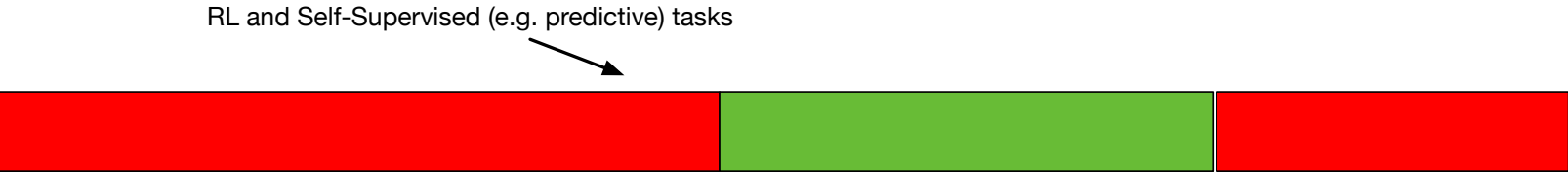
Early Useful ML
(1990's)



Impressive ML
(2000 - ?)



Generative AI / ML
(2020 - ?)

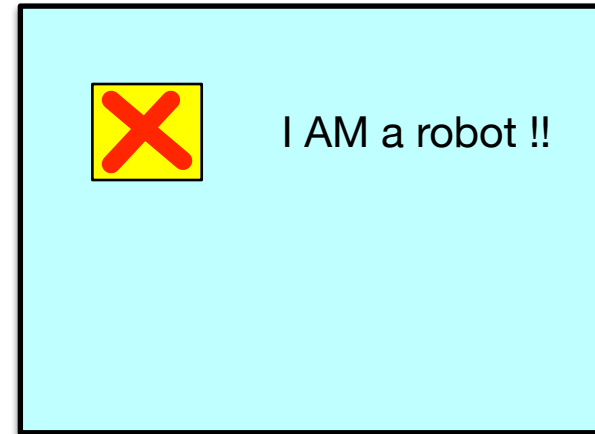


Feared Scenarios



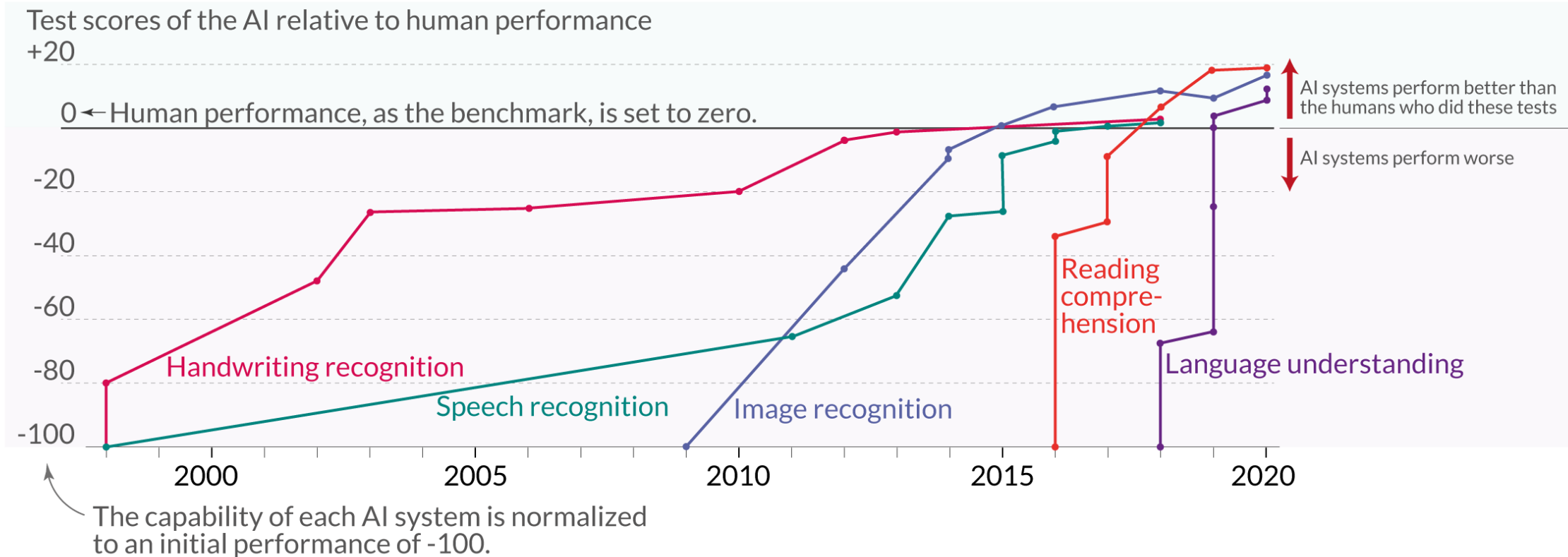
HELP WANTED (Machines Only)

- AI learns by itself. No need for expert human **logic / rules**)
- Reinforcement Learning (RL) systems generate their own labeled datasets as they explore the world => Don't even need humans to **label data** !!
- Humans no longer undisputed masters of **pattern recognition**.
- Humans may not even be the best experts on **commonsense**: It emerges in Large language models (LLMs) !!
- AI systems with **less human bias** can be **more creative**: E.g. Alpha Go's Move 37.
- **When humans become obsolete for one job, they retrain. Now, the machines can retrain too...and faster!**



Rapid Acceleration of AI Progress

Language and image recognition capabilities of AI systems have improved rapidly



Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

*** AI quickly achieves super-human performance on many new tests.**

Scott Galloway
(Popular Tech & Business Podcaster)

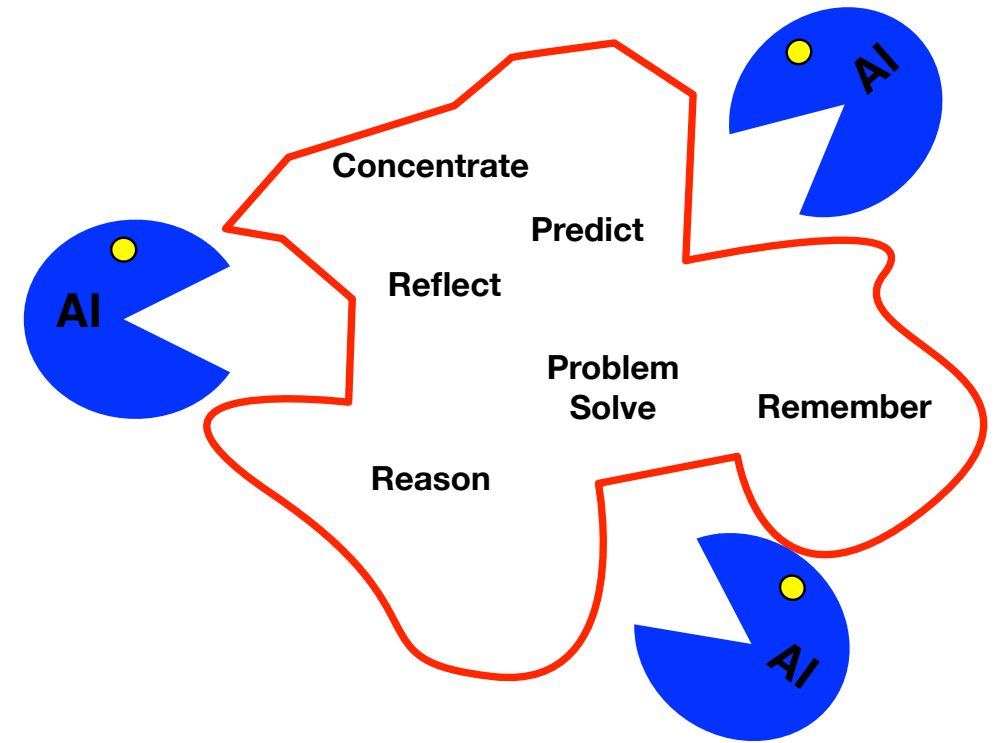
*You won't lose your job to an AI, but you might lose it to someone who **knows** AI.*

Code using Co-pilot or become obsolete

Become a ChatGPT whisperer
(a.k.a. *prompt engineer*)

Cognitive Decline:

*Humans
Need
Not
Apply
Themselves*



*The real problem is not
whether machines think but
whether men do too.*

... B.F. Skinner

Plight of the Infovore



Constant distractions of cyberspace

“Shallows”: Reduced depth of thought and emotion

Less Creative

Reduced value in an information society

Automation Bias: Trust machines over ourselves.

Rely on computers to understand the world

Rely on AI for **wisdom**

Human intelligence becomes **artificial** ... and thus more easily **predicted** and **controlled** by AI

*... as we come to rely on computers to mediate our understanding of the world, it is our own intelligence that flattens into artificial intelligence (Nicholas Carr, **The Shallows**)*

Who Works for Who/What?

Comments from Laura Preston, a *human fallback* for Brenda, a real-estate chatbot:

Months of impersonating Brenda had depleted my emotional resources. It occurred to me that I wasn't really training Brenda to think like a human; Brenda was training me to think like a bot, and perhaps that had been the point all along.

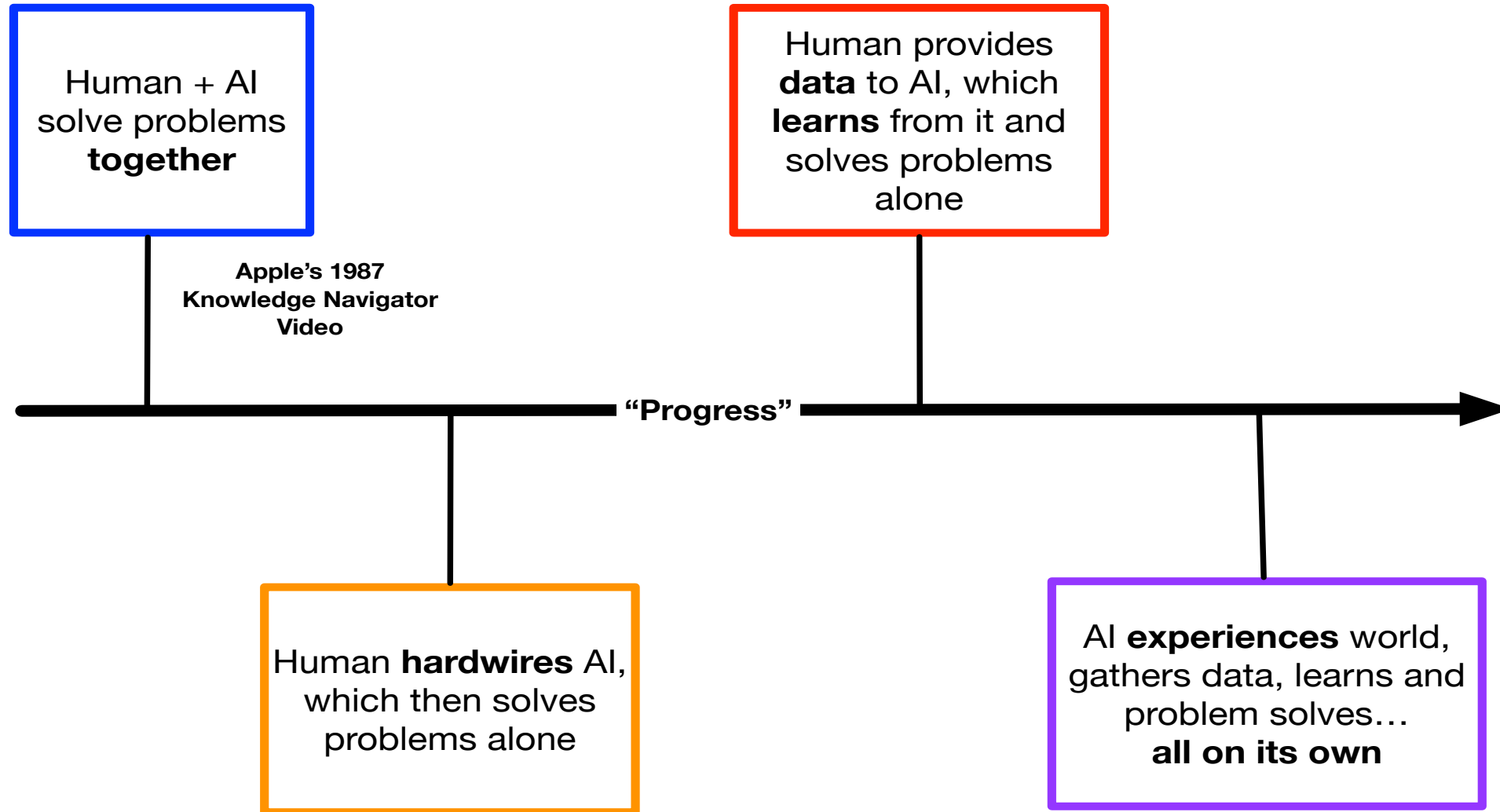
*Men have become the tools
of their tools ... Henry David Thoreau*

Don't Become a Spoon !!

*When visiting an Asian worksite in the 1960's, the great economist Milton Friedman was surprised to see workers using shovels, not modern machines. When told the reason: "It was a **jobs project**," he responded:*

If it's a jobs project, why not give them spoons instead of shovels?

Human-AI Interaction Spectrum



- Should AI help us learn, not just do the job itself?
- For the good of humanity, shouldn't we **move back left** ?

The Oppenheimer Movie

- **Robert Oppenheimer:** *I believe there is blood on my hands.*
- **President Harry Truman:** Do you think anyone in Hiroshima or Nagasaki gives a \$#%& who built the bomb? You didn't drop the bomb. I did. Hiroshima isn't about you.

If this seems like exaggerated concern, see Tristan Harris:

- The AI Dilemma (YouTube, March 2023)
- Beyond The AI Dilemma (YouTube, Sept. 2023)





Who will AI be about?



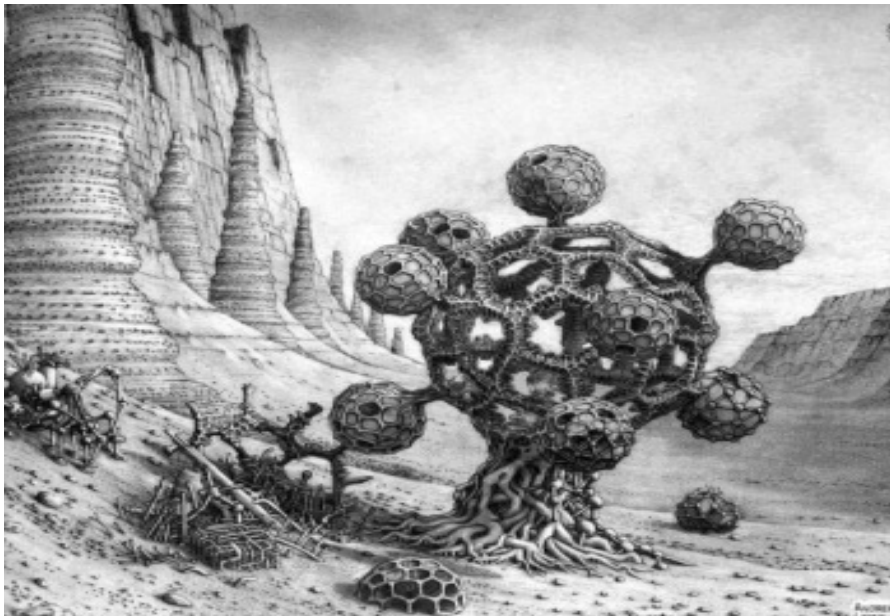
AI Transformation Playbook,
Andrew Ng, landing.ai

Great insights, from a famous AI researcher,
for any organization thinking about using AI.





The End?



Robot 19: (Andrew Lincoln Nelson, 2015)

