

AI & I:

A Personal Perspective on the Field

Keith Downing
Department of Computer Science
NTNU

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

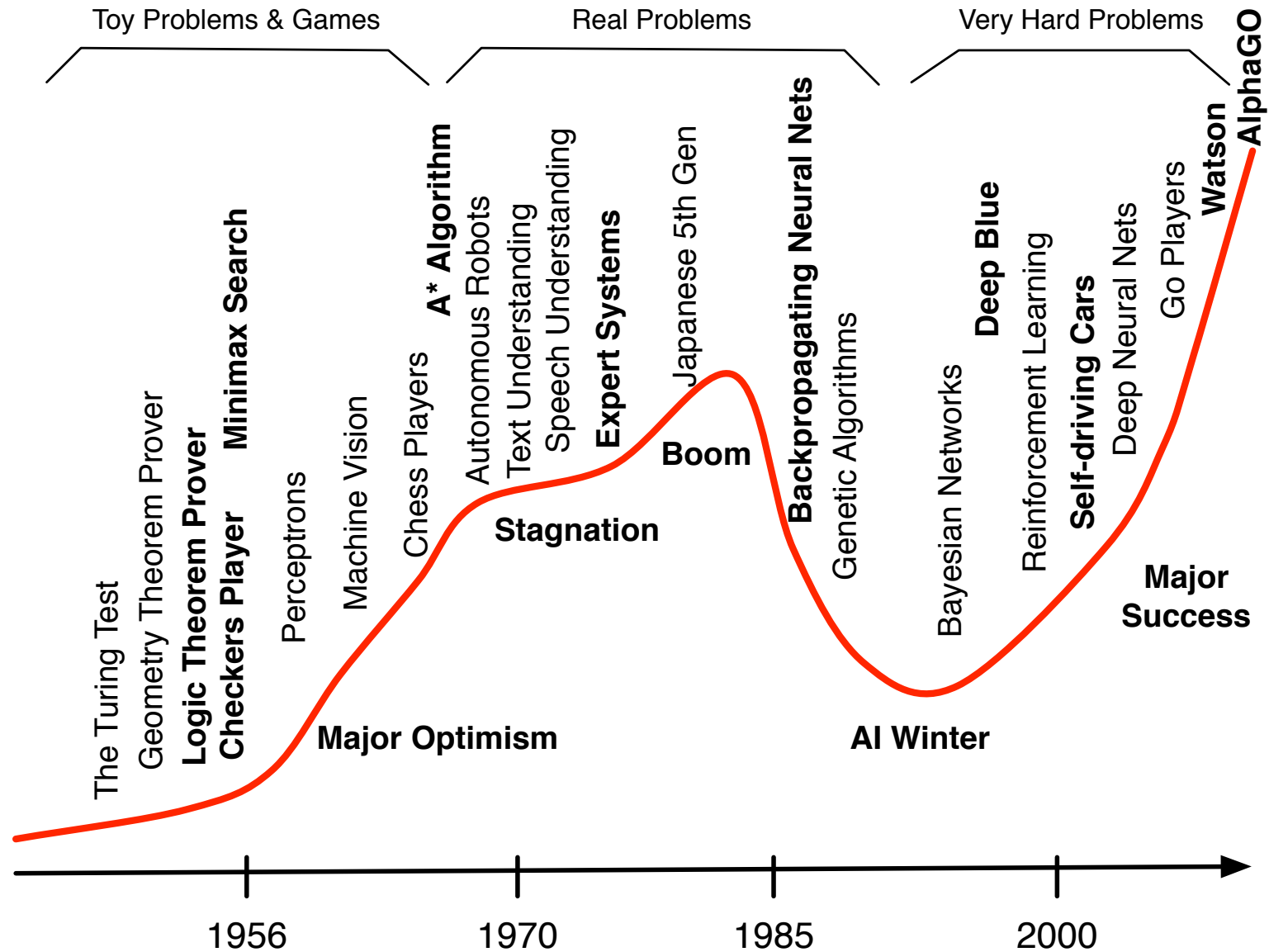
The Prejudice	The Science
Earth is center of the universe	Copernican Revolution
Humans are a special species	Darwinian Evolution
Superiority of human intelligence	AI

Bold Beginnings (Dartmouth, 1956)

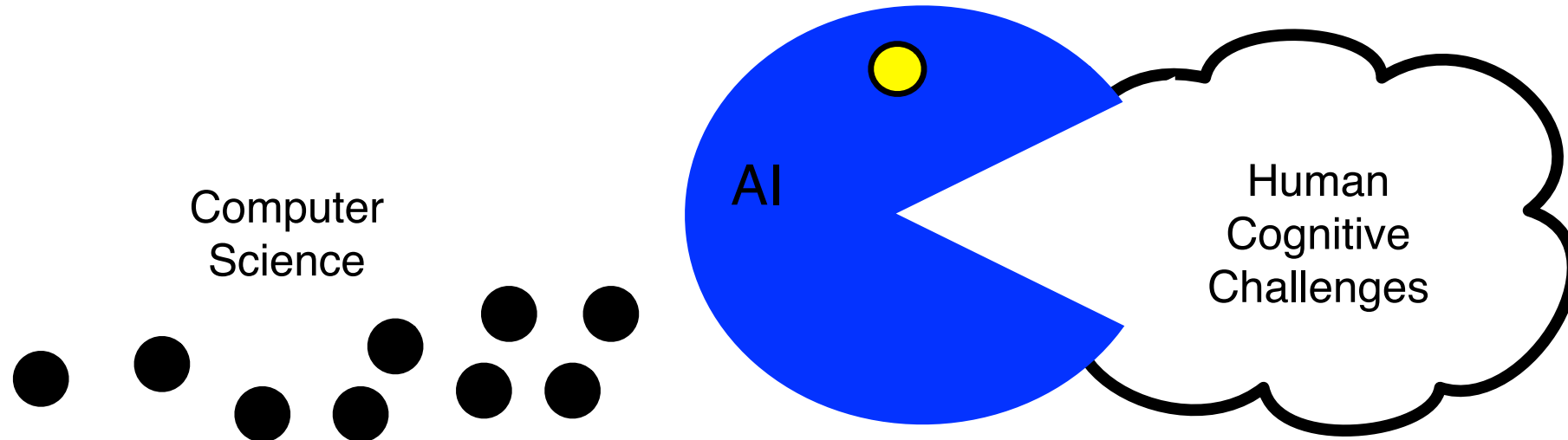


*Every aspect of learning or any other feature of intelligence can be so **precisely described** that a machine can be made to **simulate it**.*

Roller Coaster History of AI



The Ever-Changing Scope of AI



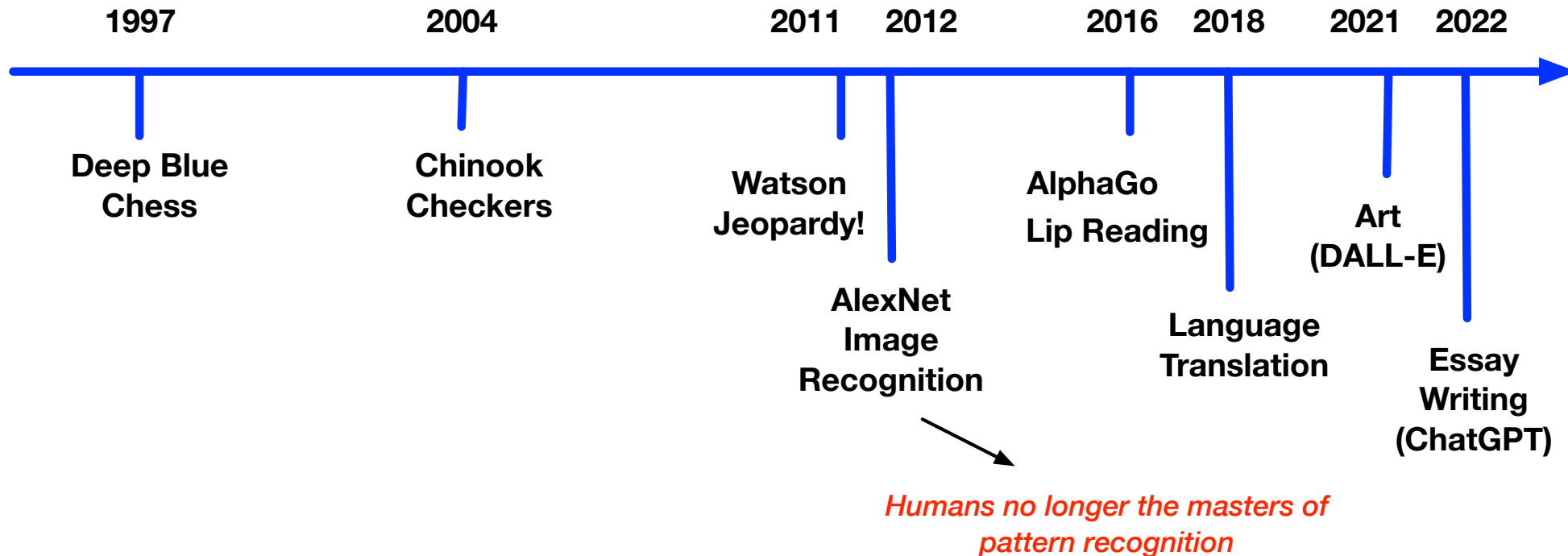
- *AI is what humans are currently better at* Jim Hendler (AI expert)
- Once a cognitive process is **demystified** by an algorithm, it is no longer AI.
- The list of human-dominated activities is shrinking dramatically.
- This definition of AI might become prominent (again) in the next “AI winter”.

Brief History of AI Success

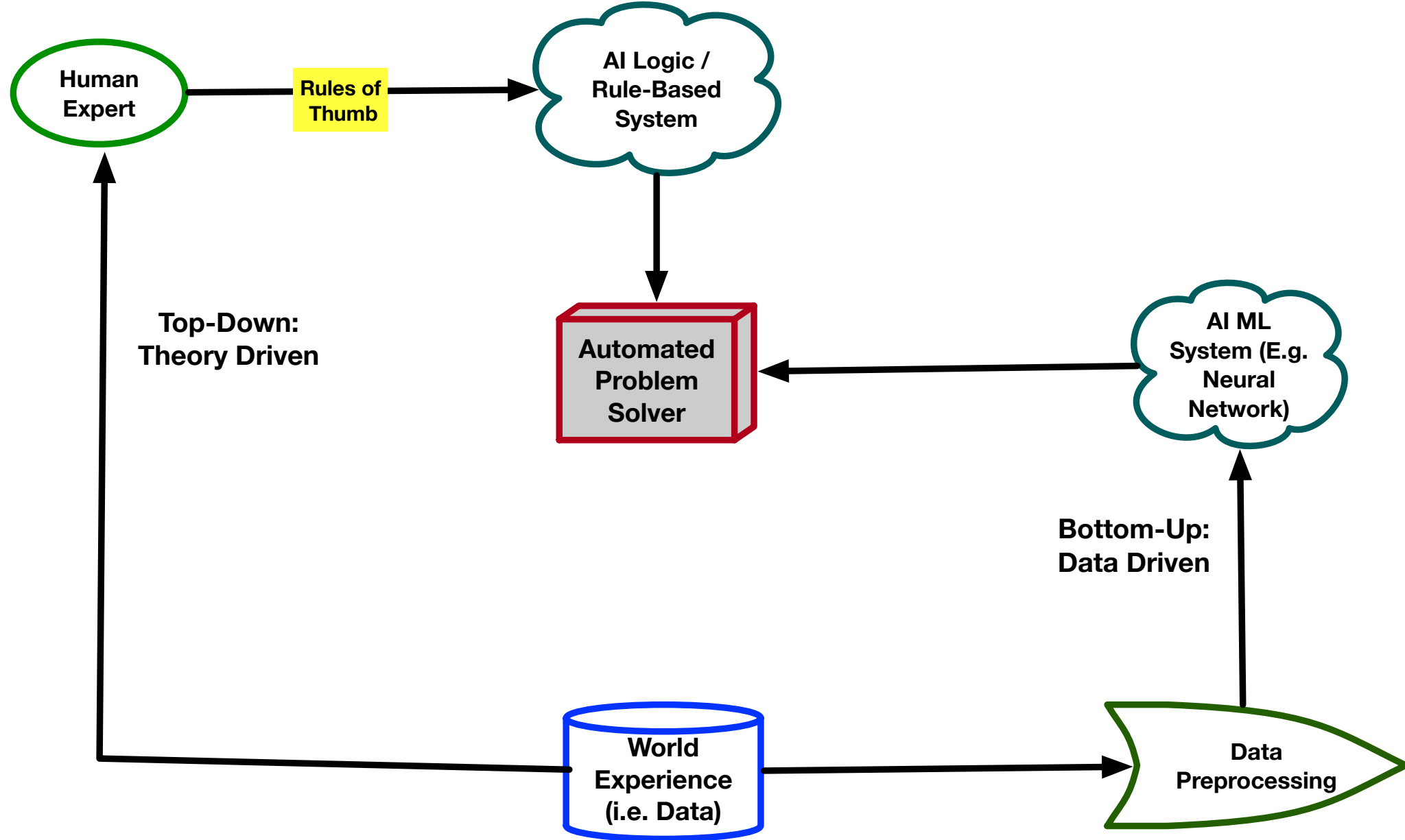
AI Beats or Equals Humans

Search-Based

**Big Data +
Advanced Pattern
Recognition**



Two Very Different AI Approaches



AI Texas Hold'em (Top-Down)

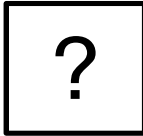
Interview poker experts (or read strategy books) to produce a large set of situation-action rules:

- IF your hole cards are a pair of face cards, AND the third such face card appears in the flop, THEN slow-play the hand (i.e. do not raise, only call).
- IF it is the final (“river”) round of betting, AND the opponent raises before you, AND the pot-odds are favorable, THEN call the raise.

AI Texas Hold'em (Bottom-Up)

Roll-Out
Perspective
for Player 2

Player 2



Player 1

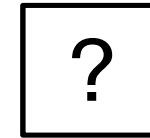
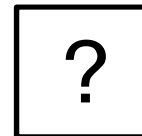
Flop



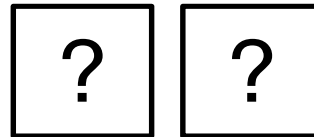
Turn



River



Player 3



Player 4

- Deal out remaining cards thousands of times.
- Count # times (W) that Player 2 wins.
- Base Player 2's betting decision on W.

AI in Medicine

Top-Down

- Doctors formalize their (deep) medical knowledge
- AI developers enter ('hardwire') it into an expert system (ES).
- Patient cases run through the ES; **no learning**.

Bottom-Up

- Medical experts label hundreds (thousands) of MRI images.
- AI system **learns** associations between images and labels.
- AI system computes labels for new images.
- AI system has no **explicit** deep domain knowledge.

DENDRAL (1965-1979)

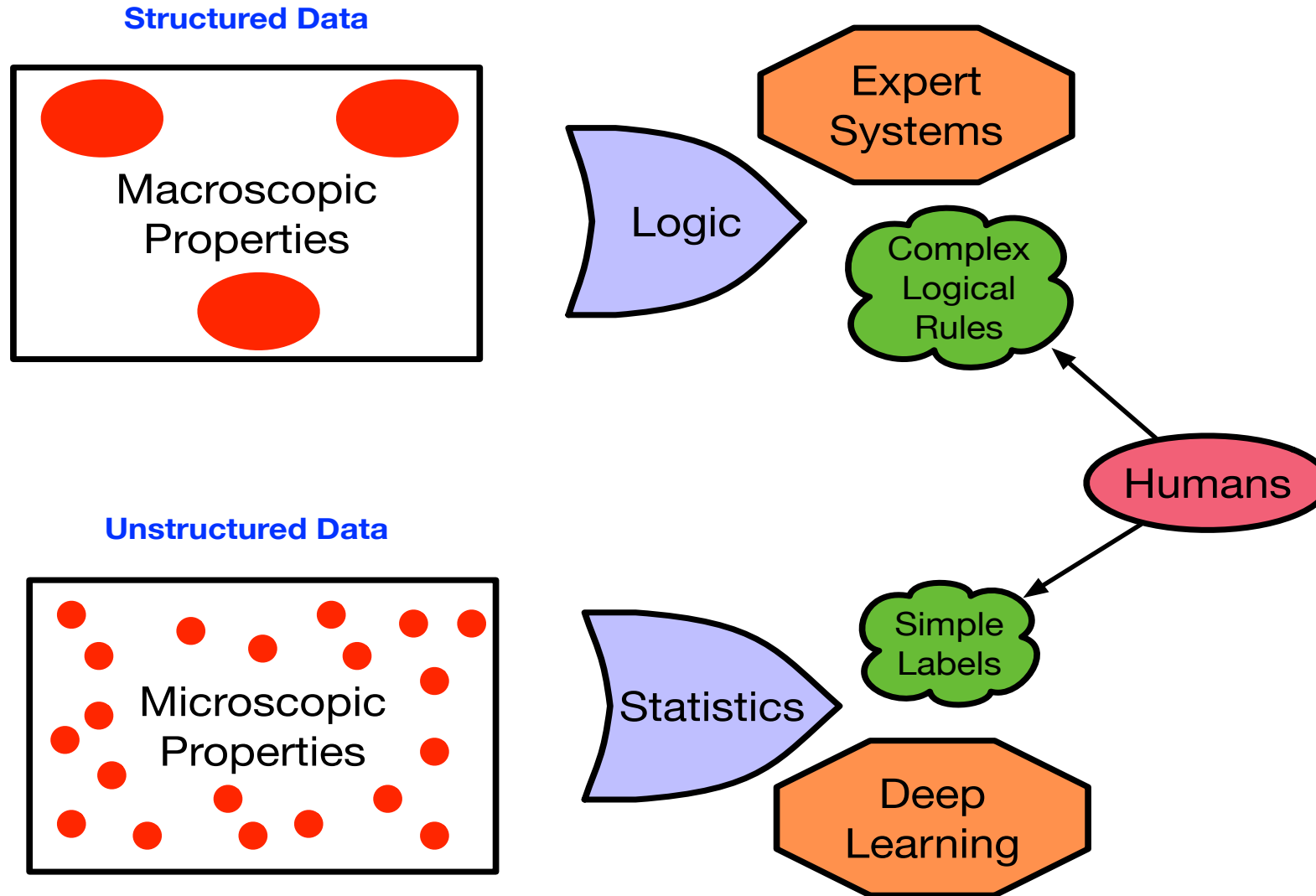
- Edward Feigenbaum (AI) + Joshua Lederberg (geneticist)
- Input: Spectrographic Data for an organic compound
- Output: Molecular Structure of that compound
- First computer system with a clean separation of:
 - Knowledge (of organic chemistry and hypothesis reformulation)
 - Reasoning (applying rules to modify hypotheses)
- Focus on **specific domain knowledge**, not the reasoning mechanisms.
- Knowledge comes directly from domain experts in the form of **if-then rules**.
- Established Knowledge-Intensive methods as the central AI problem-solving paradigm **for 15-20 years**.

MYCIN Rule # 578

- IF
 - A. The infection is meningitis
 - B. Organisms seen on the culture's stain
 - C. Type of infection = bacterial
 - D. Patient has been seriously burned
- THEN
 - Cause = Pseudomonas-Aeruginosa
with (probability = 0.5)

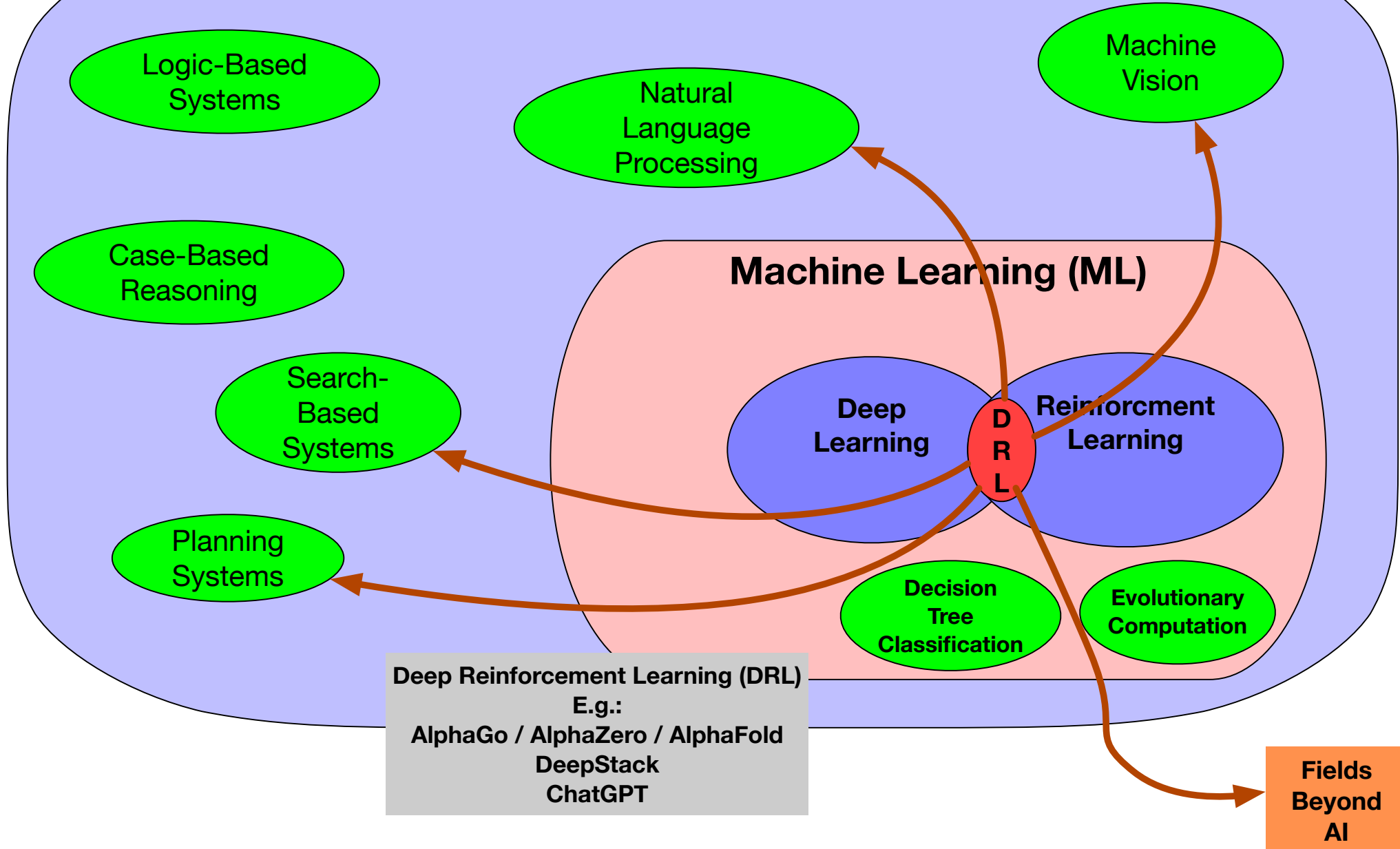
MYCIN (Stanford, early 1980's) was one of the first successful expert systems.

Microscopic Pattern Recognition



- Recognition of Complex Microscopic Patterns is now possible (with AI)
- Reams of Microscopic Data are now much easier to produce

Artificial Intelligence (AI)



Deep Learning

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

Beat Humans at
Complex Tasks of
Pattern-Recognition

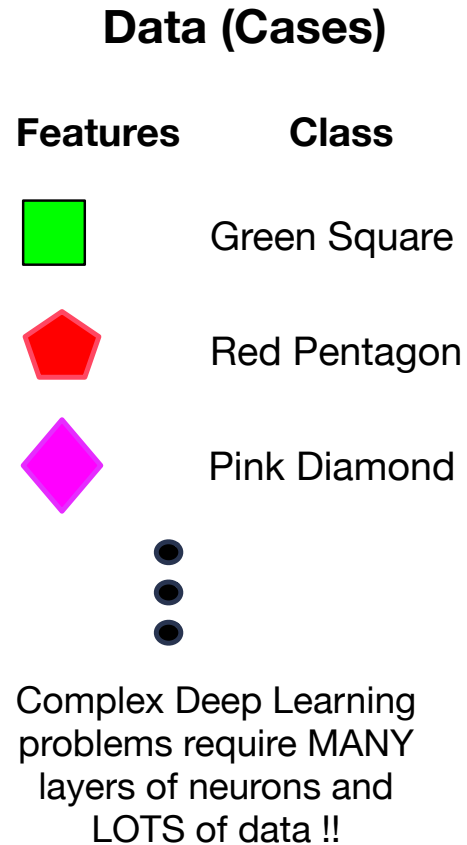
Impress (and Fool) Humans
with AI-generated
Art, Music, Text, Images,
Audio and Video

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)

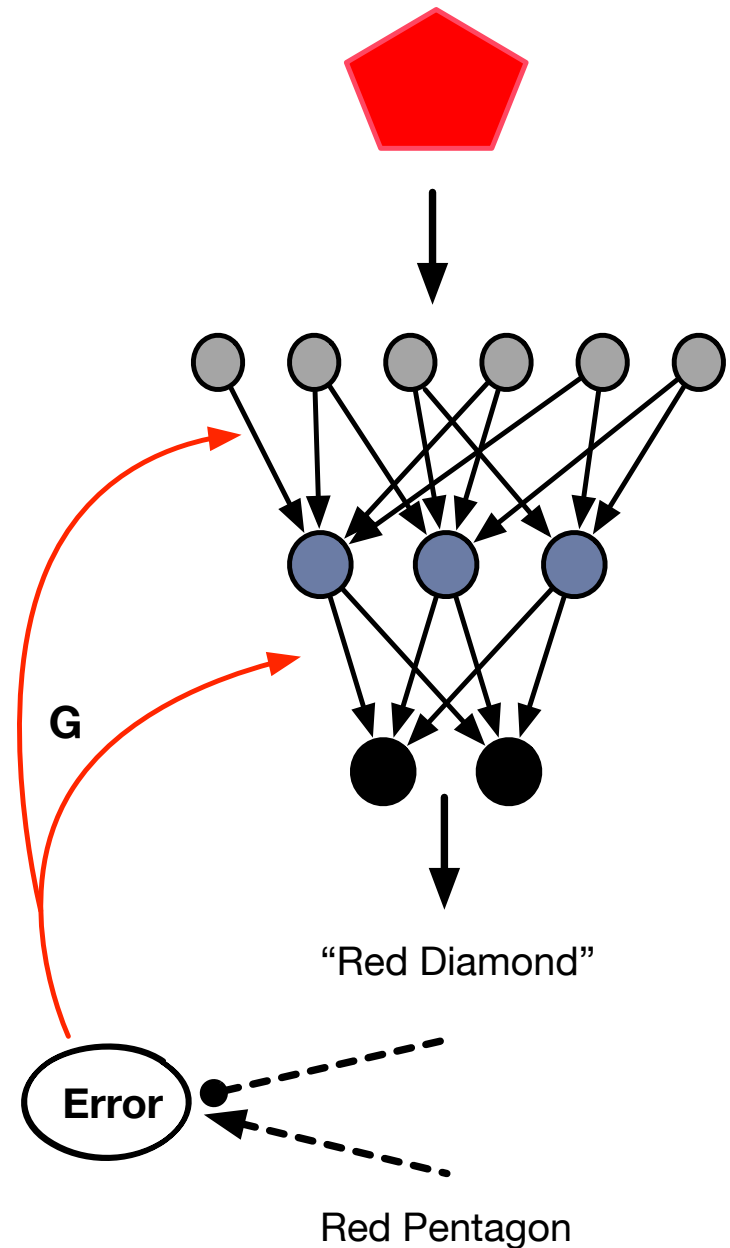
Global gradients are simply products of local gradients.

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



Gradient Descent

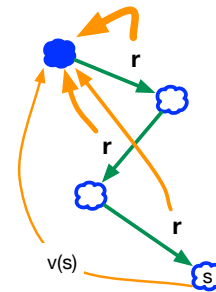
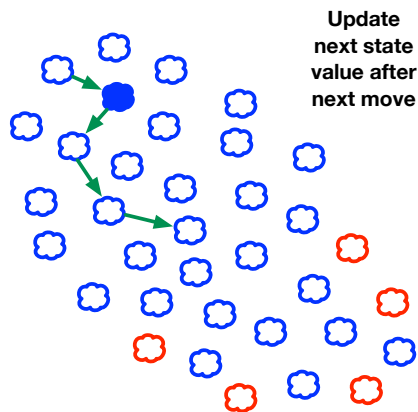
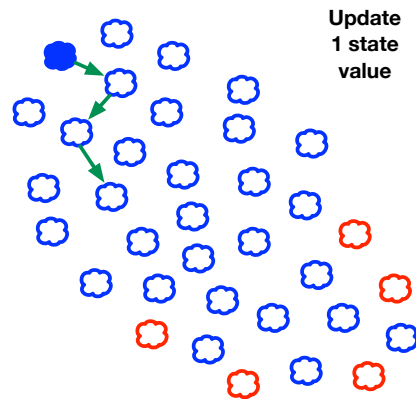
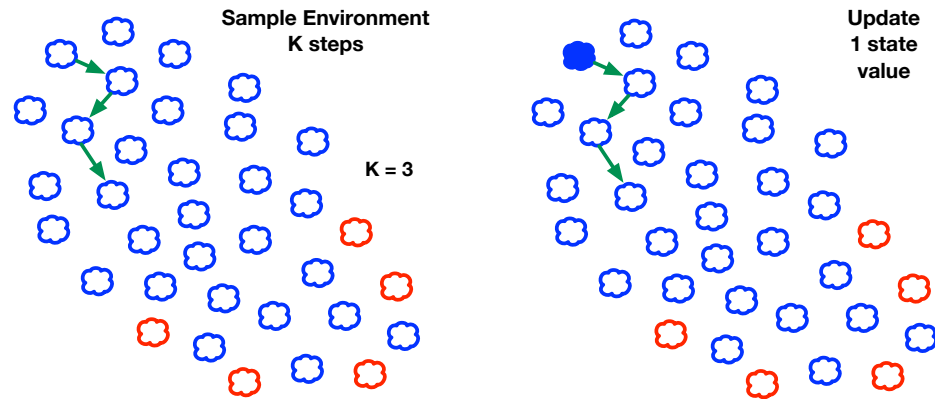
$$\text{Gradient (G)} = \frac{\triangle \text{Error}}{\triangle \text{Weight}}$$



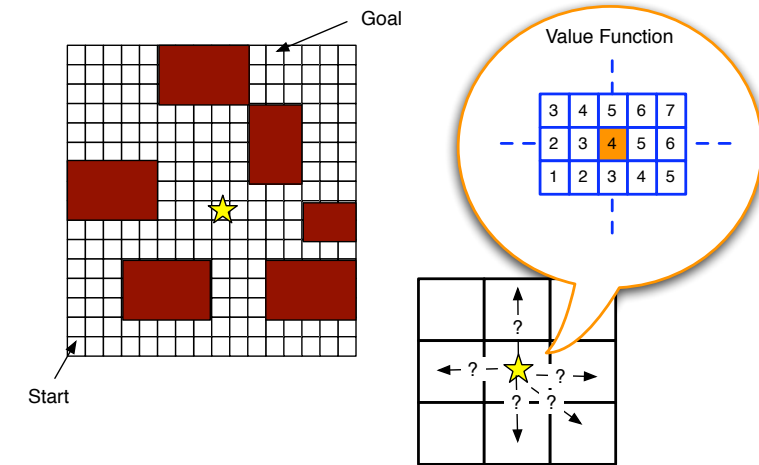
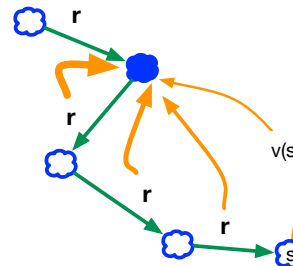
Trial and Error Learning:

- Take an action
- Record result
- Modify action rules (i.e., learn)
- Repeat

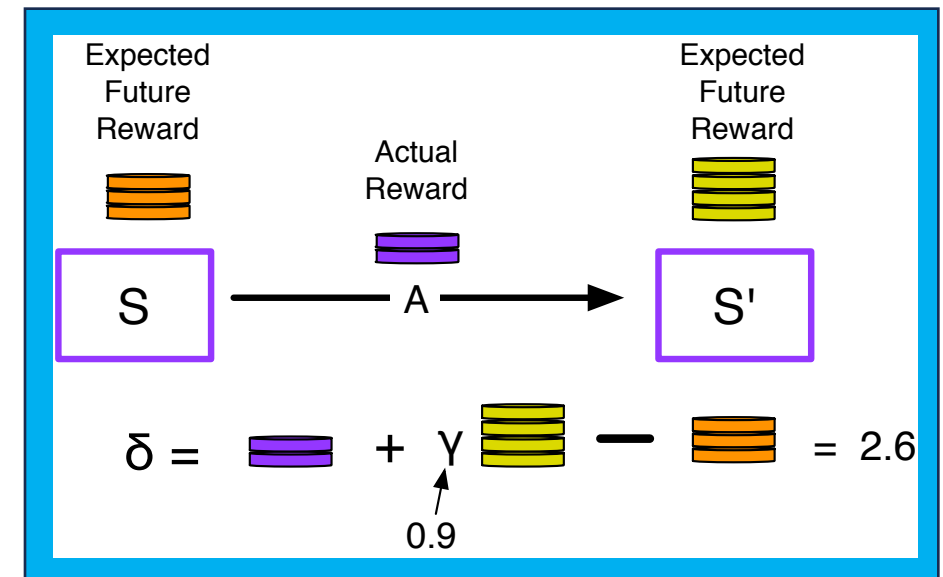
Reinforcement Learning (RL)



Bootstrapping
Value Updates



Bootstrapping: The **predicted** value of a future state is used to update the **predicted** value of the current state.



Deep Reinforcement Learning (DRL)

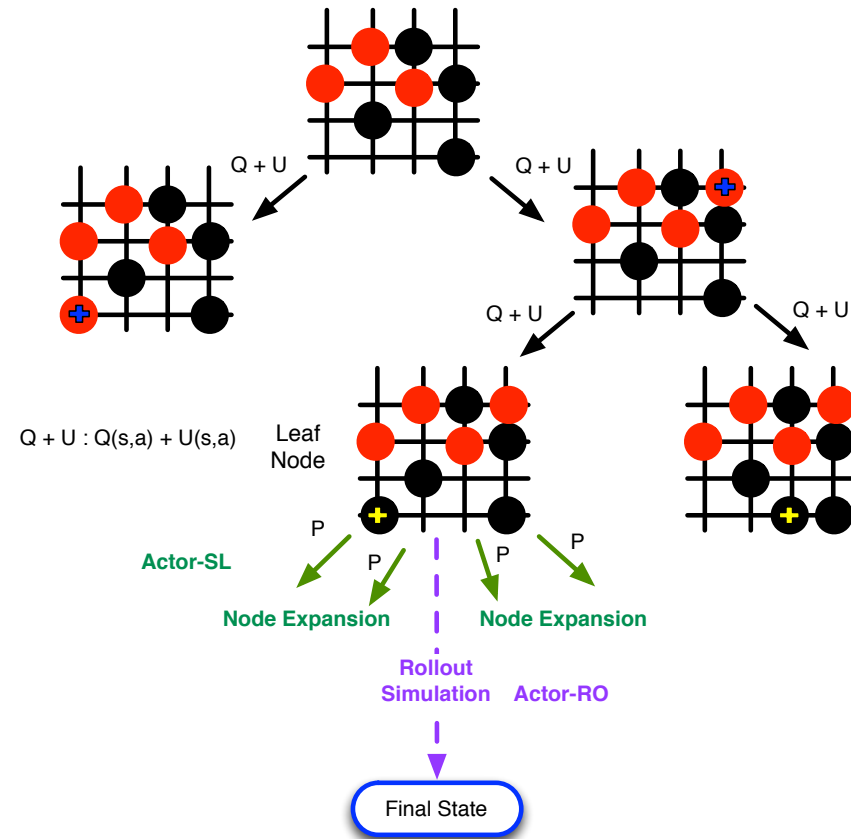
- Goal of RL: Learn many complex rules (**state => action**) at the **microscopic** level that are **general**.
- General: maps **similar** states to similar actions.
- But **similarity** can be hard to define.
- Deep Nets learn **salient similarities** (that humans often miss) due to the net's superior pattern-recognition abilities.
- Deep Nets can process billions of microscopic cases in producing these rules.
- These cases are **generated internally**, by the trial-and-error search inherent in RL. **No human labeling needed!**

Deep-Learning and/or DRL Applications

- **Self-Driving Vehicles**
- Championship-level Game Playing
- Adding Sound Effects to Silent Movies
- **Fake Video; Detecting Fake Social-Media posts**
- Accelerated Discovery of New Drugs
- **Predicting Protein Folds**
- Interpreting Web Data for Cancer Research
- Fraud Detection ; Credit-risk assessment
- **Early detection of Cerebral Palsy (Open AI Lab)**
- **Natural Language Translation**
- Generating Articles from Statistics (e.g. Sports)
- Generating long text from short and/or complex prompts.
- **Generating Art + Music** based on styles of particular artists
- Earth Monitoring – Interpreting Satellite Images
- **Habitat Monitoring – Analysis of animal sounds to detect poachers**

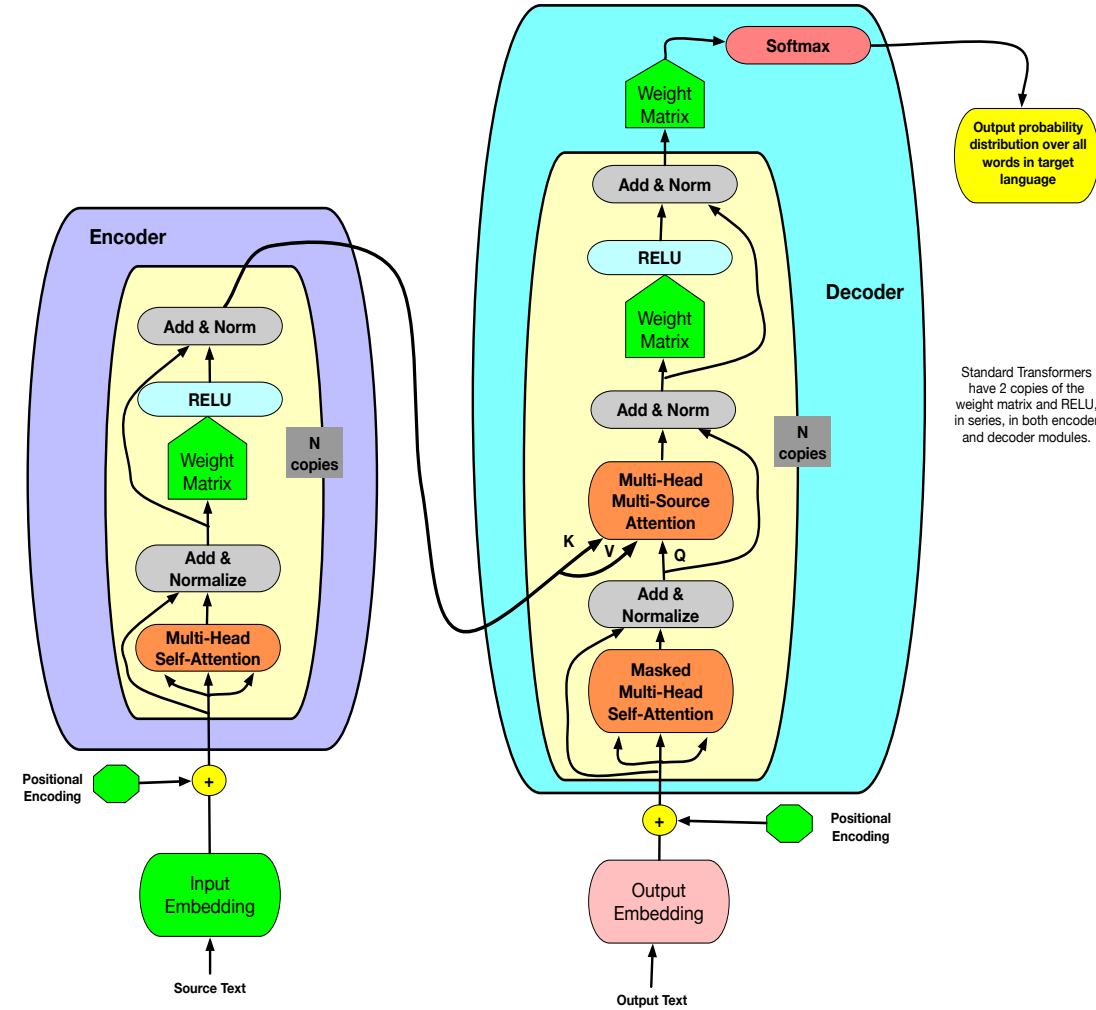
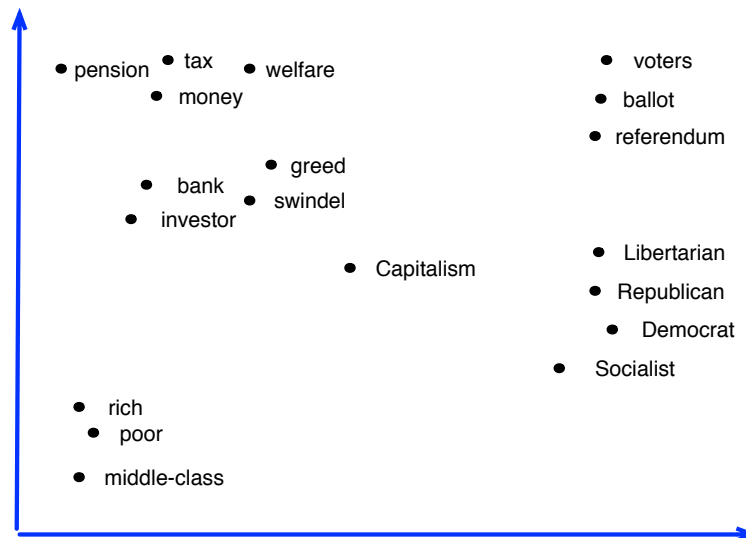
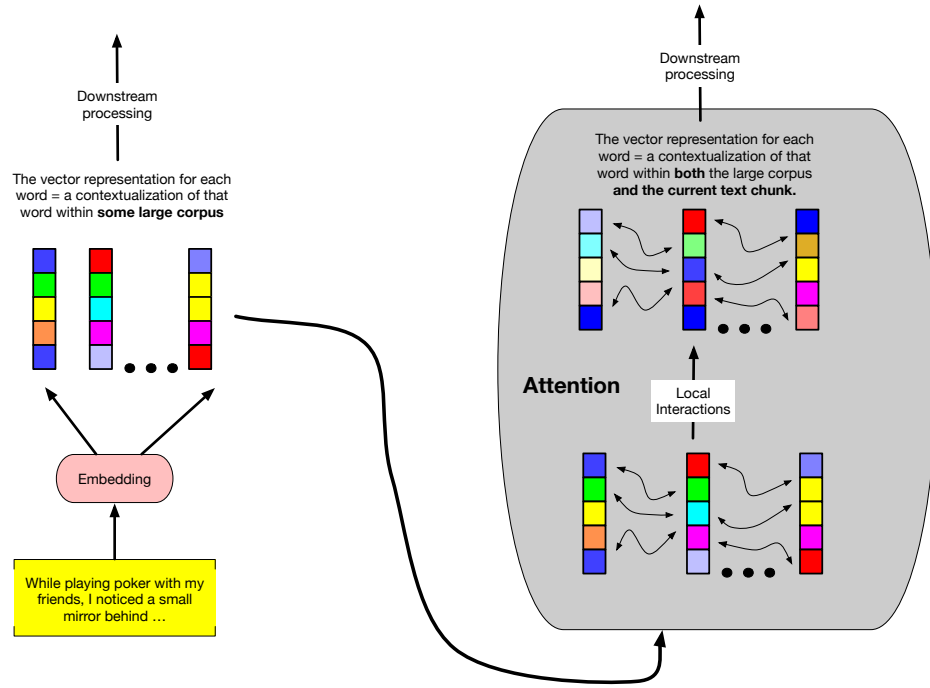
AlphaGo

Silver et. al., *Mastering the game of Go with **deep neural networks** and **tree search***, Nature, 2016.



I guess I lost the game because I wasn't able to find any weaknesses...**Lee Sedol** (World # 2)

Attention, Transformers and ChatGPT



Harvesting Data for Predictive Tasks

- Creating intelligent neural nets requires many data pairs: (features, label).
- Example: (picture of an animal, "giraffe")
- Labels typically come from humans => lots of work !!
- Predictive tasks do not require human labelling.
- Predictive task: features = state_t ; label = state_(t+1)
- So just wait one timestep to get the label !
- Long text supports a predictive task: features = (last k words); label = next word.
- A text of N words gives N-k+1 cases.

The recent indictments of the former president...

1. The recent indictments
2. recent indictments of
3. indictments of the

Features (k=3)

1. of
2. the
3. former

Label

* In Transformers, features = a more general (and complex) context, which is produced via attention.

Stochastic Choices

Me: Give me the next 3 words in this sentence: The recent indictments

ChatGPT(3.5): ... were unexpected and.

Me: Please try again.

ChatGPT(3.5): ... were followed.

Me: One more time, please.

ChatGPT(3.5): ... were surprising.

Me: You keep giving me 2 words instead of the 3 that I requested.

ChatGPT(3.5): I apologize for the oversight. Here are the next three words in the sentence: The recent indictments have.

Me. You seem to have some counting issues.

ChatGPT(3.5): I apologize for the repeated errors. Please allow me to provide the correct sequence: The recent indictments have caused.

Knowledge Representations in AI

Predicate
Logic

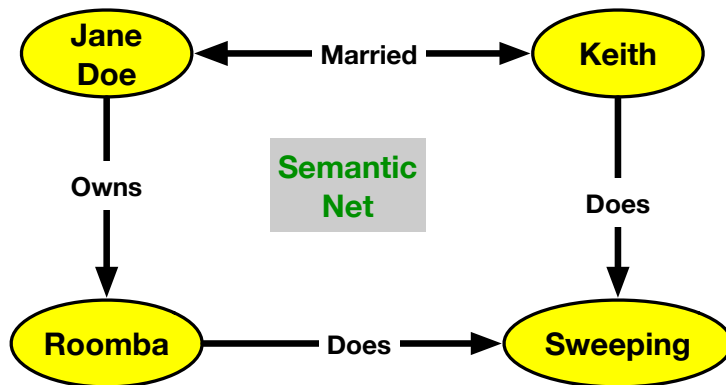
“Everybody loves somebody sometime”

$\forall X \text{ Person}(X) \Rightarrow \exists Y, T : \text{Person}(Y) \wedge \text{Time}(T) \wedge \text{Loves}(X, Y, T)$

“My wife seems to trust Roomba more than me to sweep the house”

$\text{Has_Task}(\text{Jane-Doe}, \text{Sweep}(\text{House})) \wedge \text{Married}(\text{Jane-Doe}, \text{Keith}) \wedge$
 $\exists X : \text{Helps}(X, \text{Jane-Doe}, \text{Sweep}(\text{House}))$
 \Rightarrow
 $[\text{prob}(X = \text{Roomba}) = 0.75]$

Probabilistic
Logic



Personal
Context

Chipped Beef on Toast (a.k.a. \$%#\$ on a shingle):

This was a “Dad special” that we suffered through on Sunday mornings. He learned the recipe while serving in the Army reserves in Texas. It was shown being served to troops in a 2023 WW-II series on Netflix. It resembles fish scales in cream, and tastes even worse.

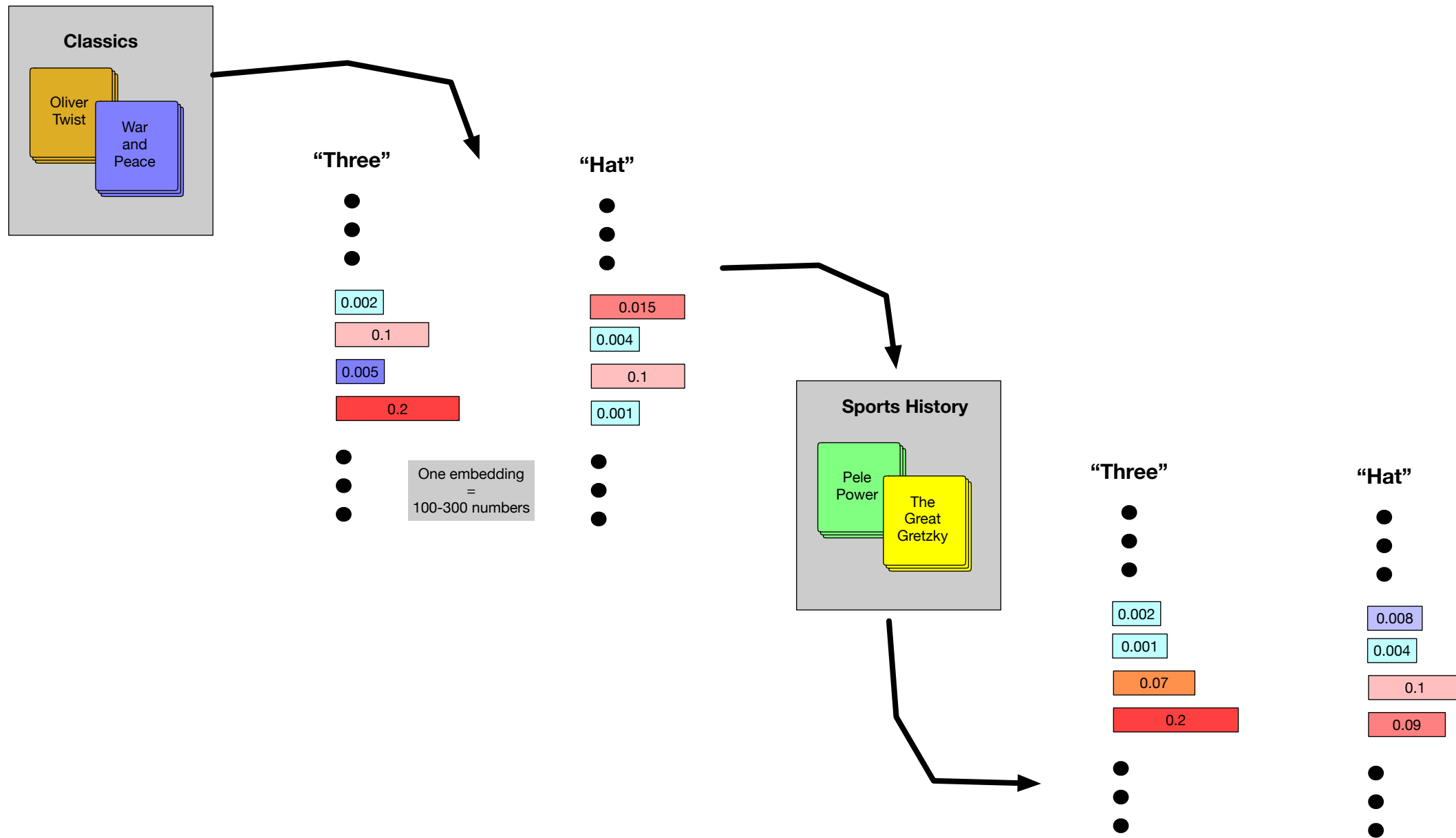
Objective
Fact

Subjective
Sensory
Impressions

0.003
0.09
0.001
0.002
:
:
:
:
:
:
:
0.001
0.023

Attention
Vector
(Embedding)

Embeddings (Words = Vectors)



* In several sports, "hat trick" = 3 goals in a game.

Prediction => Emergent Problem Solving

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.*

Prediction => Emergent Problem Solving

Me

I have 3 sticks, some rope, and a very strong leaf blower. How can I make something that looks like a helicopter?

ChatGPT 3.5

Create a rotor assembly:

- Take the three sticks and arrange them in a Y-shape, with one stick forming the central hub and the other two forming the rotor blades. Secure them together at the center using rope or strong string to form a rotor assembly.
- Attach the rotor assembly to the leaf blower:
 - Using the rope or string, attach the rotor assembly to the output nozzle of the leaf blower. Make sure it's securely fastened so it doesn't come loose during operation.
- Turn on the leaf blower:
 - Turn on the leaf blower, and the force of the air will cause the rotor assembly to spin like the blades of a helicopter.

Prediction → Emergent Understanding (??)

Me: If Don is indicted for treason, how does he compare to Quisling?

ChatGPT(3.5): If Don is indicted for treason, a comparison to Quisling might involve similarities or differences in their alleged actions and motives regarding betraying their country. Vidkun Quisling was a Norwegian who collaborated with Nazi Germany during World War II...

Me: What's the big deal if I give you one piece of candy while I get two?

ChatGPT(3.5): The situation seems unfair as you receive more candy. Equality and fairness are important for positive interactions and relationships.

Me: Is it just as big a deal if I give you 87 pieces while I get 88?

ChatGPT(3.5): The situation remains slightly unequal, but the disparity is smaller. Fairness still matters for fostering positive interactions and perceptions.

Me: Would anyone indict me for unfairness in this last scenario?

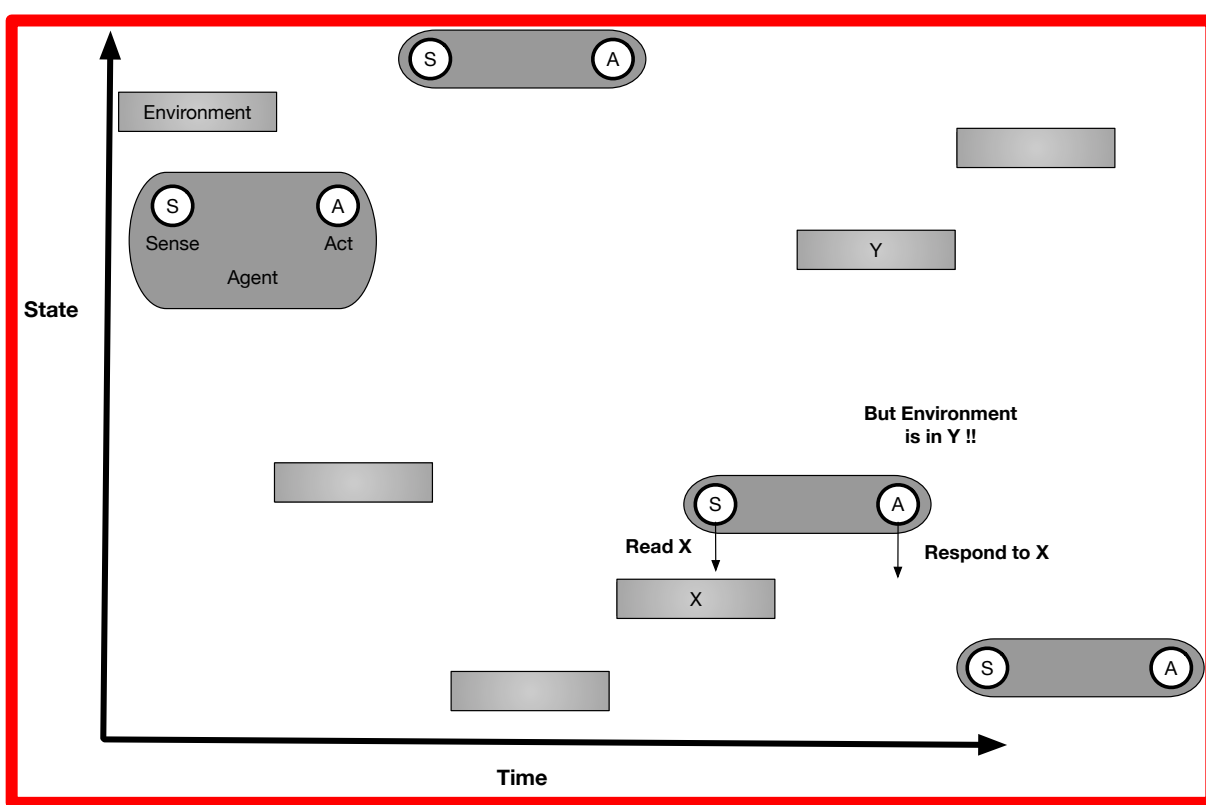
ChatGPT(3.5): Indictment typically doesn't apply to personal fairness matters. However, perceptions of unfairness could lead to disputes or conflicts in certain contexts.

Me: Can you name such a context?

ChatGPT(3.5): In a professional or organizational setting, unequal treatment among employees in terms of rewards or opportunities could lead to grievances or potential conflicts.

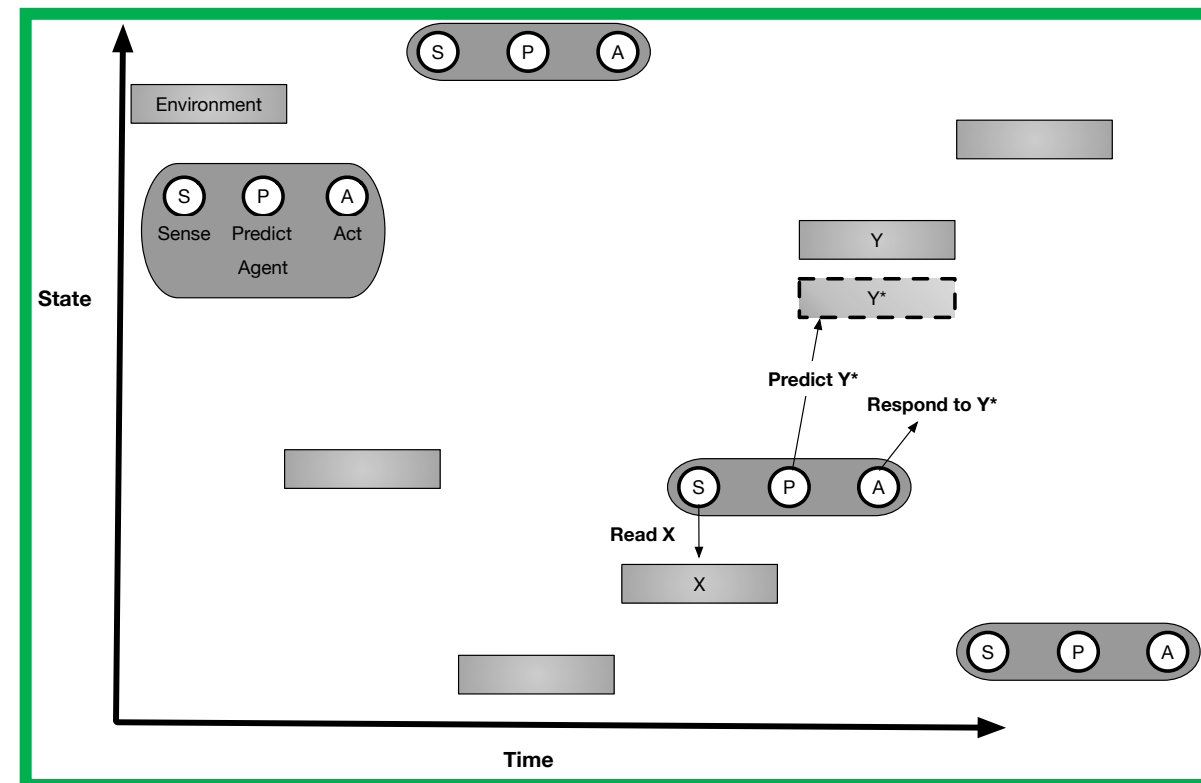
Prediction = The Core of Intelligence

- The capacity to **predict** the outcome of future events - critical to successful movement – is likely, the ultimate and most common of all global brain functions. (Llinas, *I of the Vortex*, 2001)
- The mystery is, and remains, how mere matter manages to give rise to thinking, imagining, dreaming, and the whole smorgasbord of mentality, emotion and intelligent action...But there is an emerging clue...summed up in a single word: **prediction**. (Clark, *Surfing Uncertainty*, 2016).
- The cortex is an organ of **prediction**. If we want to understand what intelligence is, what creativity is, how your brain works, and how to build intelligent machines, we must understand the nature of these **predictions** and how the cortex makes them. (Hawkins, *On Intelligence*, 2004)



Without
Prediction

With
Prediction



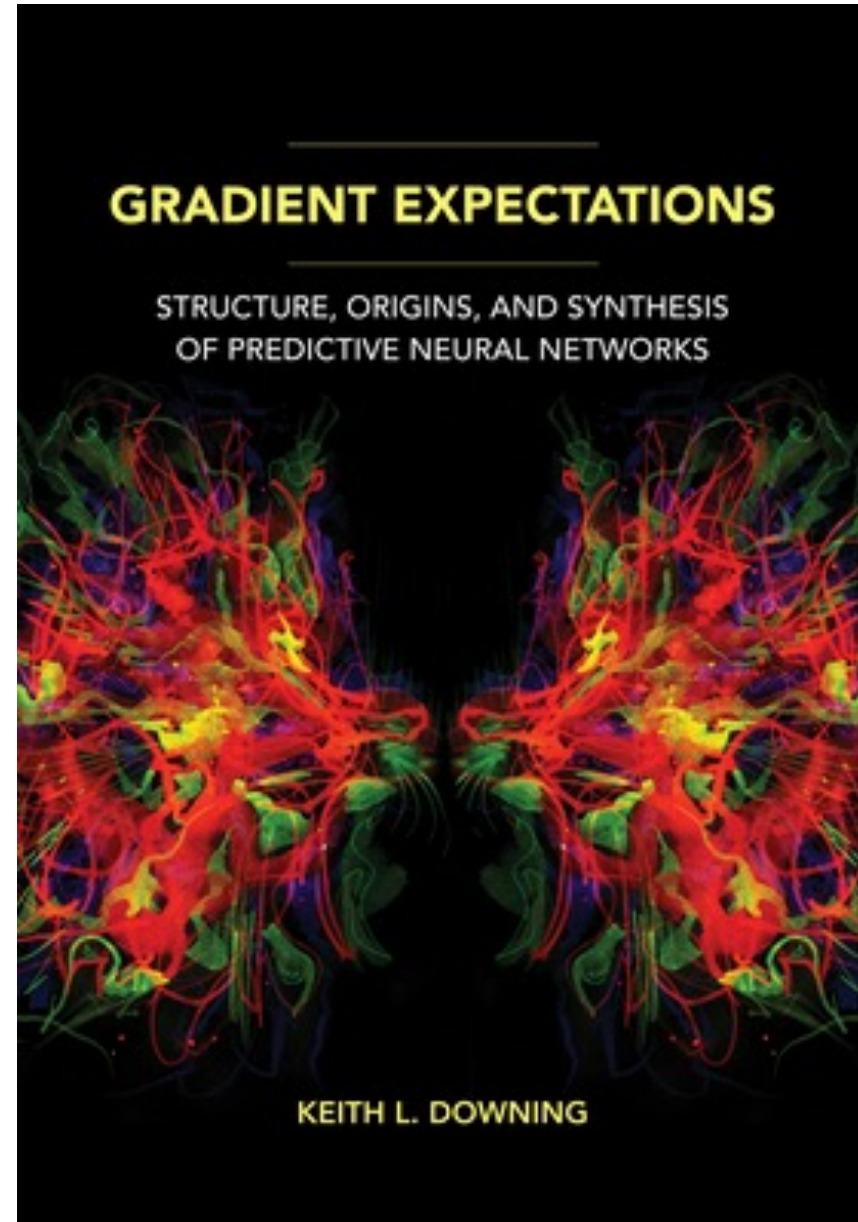
Charting a path from early artificial neural networks to the contemporary vision of the predictive brain, with rich forays into biology and evolution, this book explains the buzz about brains as engines of prediction.

(Andy Clark, University of Sussex)

Downing's reach is omnidirectional. He connects the roots and new growth of deep learning with math, neuroscience, evolutionary biology, ethology and computer science to show **how intelligence emerged in animals and is emerging in machines.**

(Josh Bongard, University of Vermont)

... not so easy



**MIT Press
2023**

Diminishing Transparency of AI

Showing Off

Nature



Download from
Dreamstime.com
This watermarked comp image is for previewing purposes only.

12934202
Shawn Hempel | Dreamstime.com

Military



Download from
Dreamstime.com
This watermarked comp image is for previewing purposes only.

44349832
Xi Zhang | Dreamstime.com

Financial Sector



Download from
Dreamstime.com
This watermarked comp image is for previewing purposes only.

44349832
Xi Zhang | Dreamstime.com

Academia



Download from
Dreamstime.com
This watermarked comp image is for previewing purposes only.

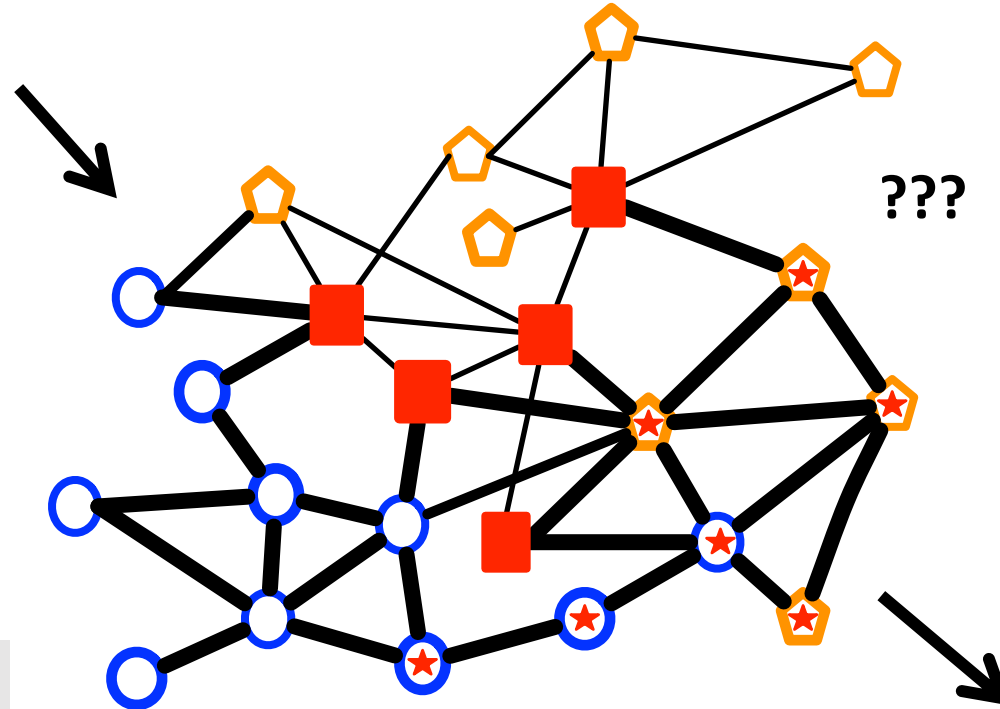
7580956
Ctack | Dreamstime.com

Translucent but not Transparent

Dissecting the Black Box



*Competence without
Comprehension* (Dennett)

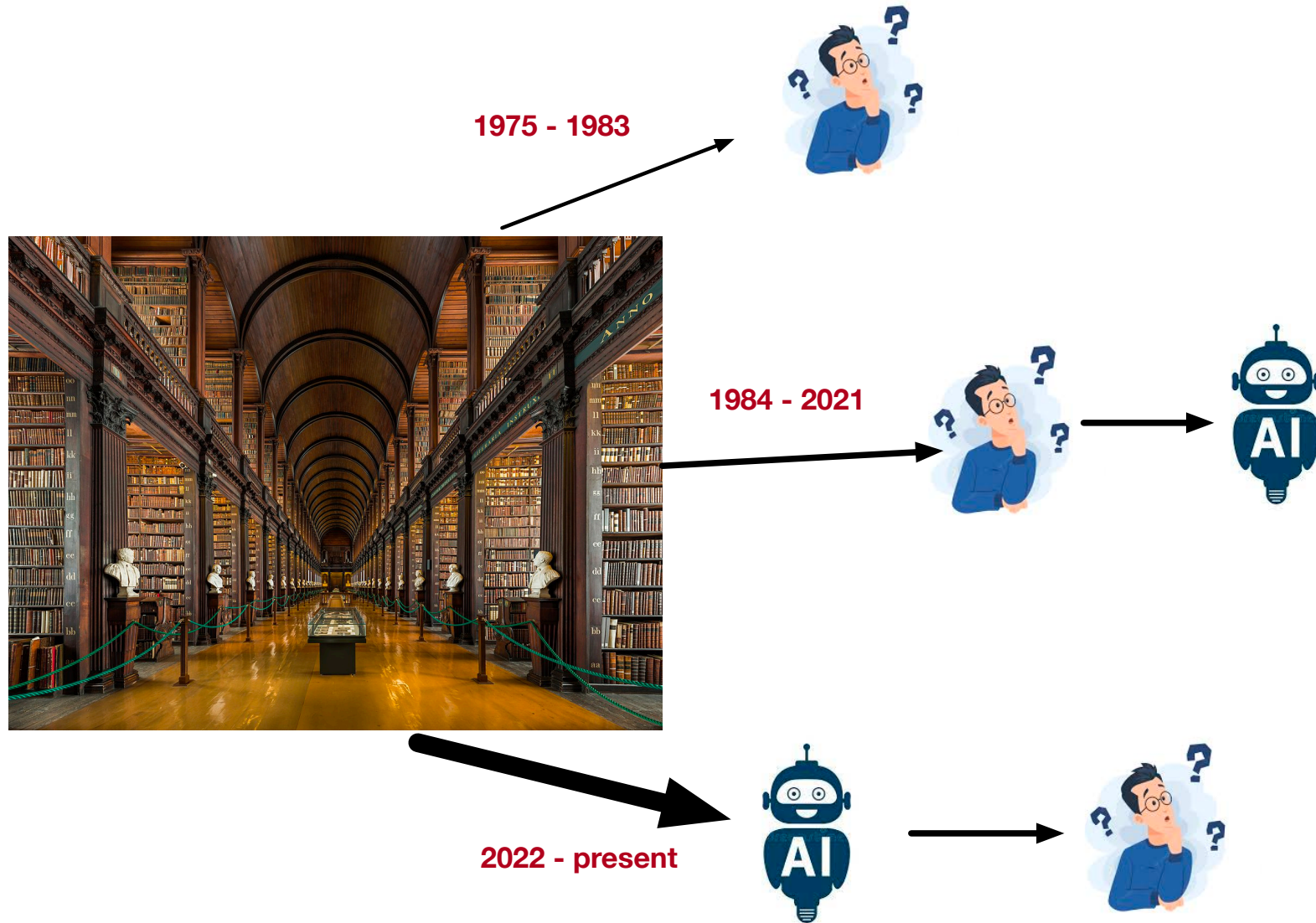


Older logic-based
approaches (e.g. expert systems)
are less successful, but
their reasoning is easy to follow.

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

Remove
Tumor !!

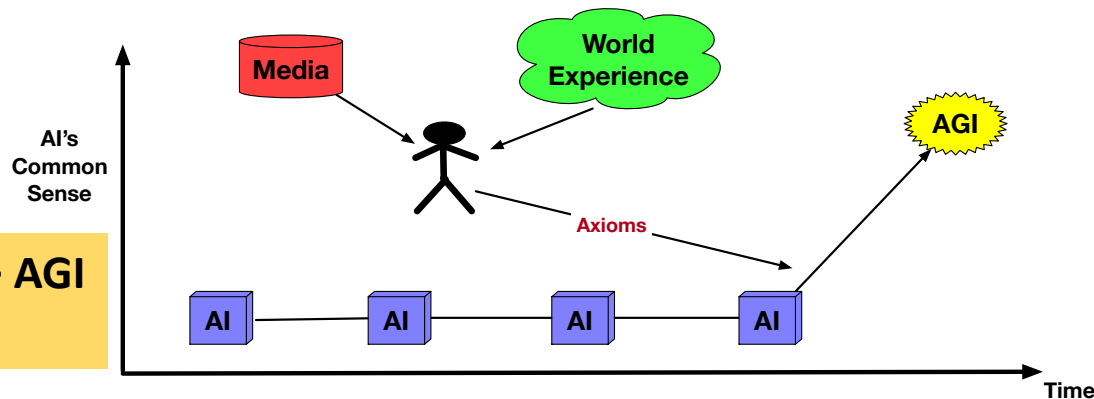
My Evolving Trauma



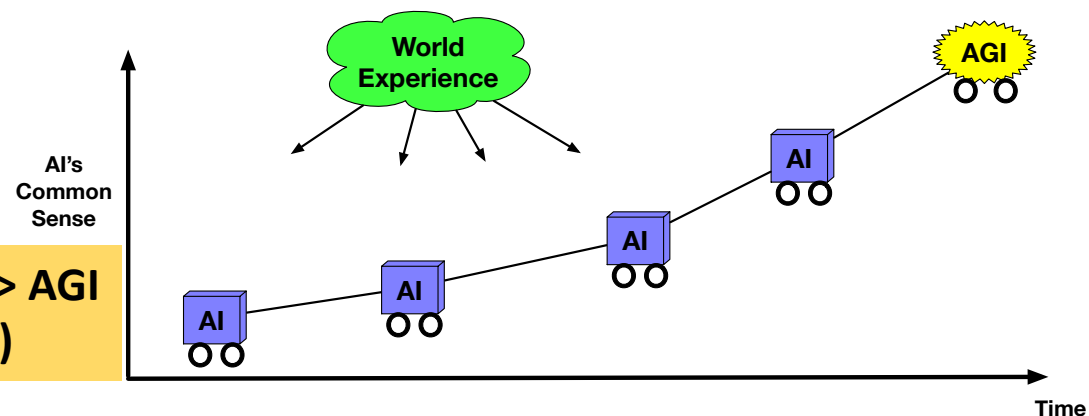
Alternative Approaches to Common Sense and Artificial General Intelligence (AGI)

Common Sense = primary deficiency of most AI systems.

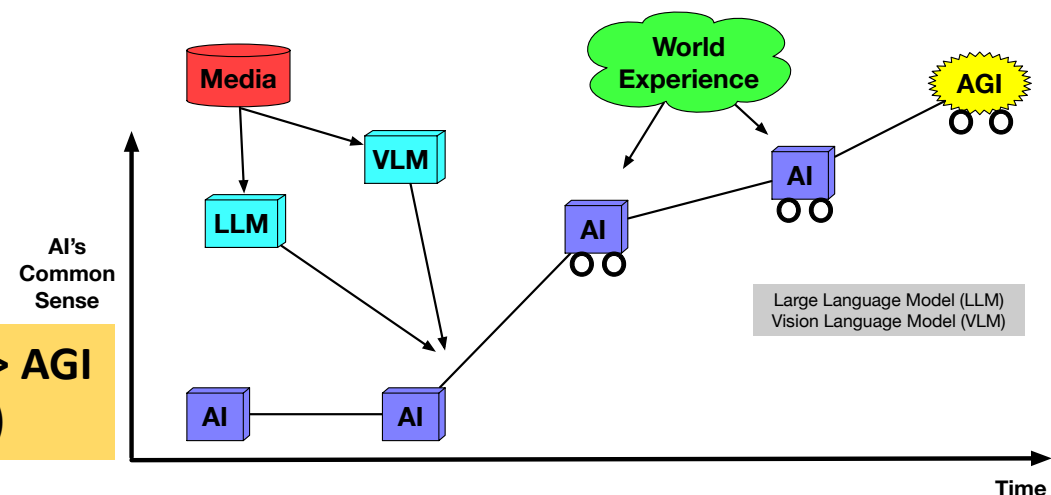
**Logic => AGI
(1980's)**



**ALife => AGI
(1990's)**



**LLMs => AGI
(2020's)**



Axiomatization of Common-Sense (1980's)

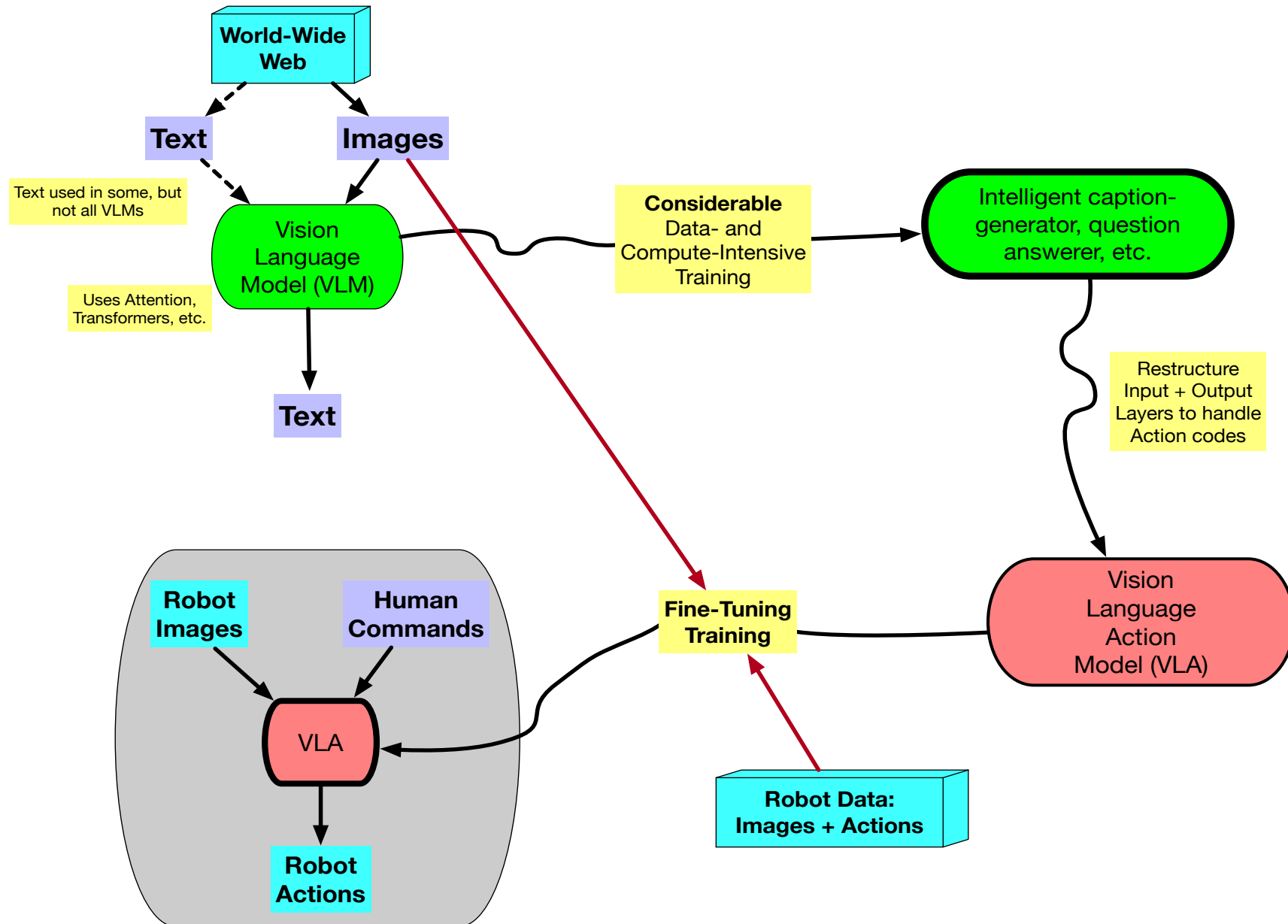
- Qualitative Reasoning (P. Hayes, J. Allen, K. Forbus, J. de Kleer, B. Williams. B. Kuipers..)

- A common-sense theory of Time (James Allen and Pat Hayes, 1985)

$\text{OVERLAPS}(I,J) \Leftrightarrow \text{EXISTS } a,b,c,d,e : \text{MEETS}(I,a) \text{ and } \text{MEETS}(I,d) \text{ and } \text{MEETS}(d,c) \text{ and } \text{MEETS}(a,b) \text{ and } \text{MEETS}(J,b) \text{ and } \text{MEETS}(J,c) \text{ and } \text{MEETS}(b,c) \text{ and } \text{MEETS}(c,d)$

- *The CYC Project (1984 – present) (Douglas Lenat)*
 - *2000+ person-years*
 - *More than 24 million axioms!*
 - *What we need is nothing less than an “AI Manhattan Project”, a full frontal assault on common sense.... Lenat*
 - *Very limited success*
- *Representation without reasoning is an idle exercise* (Ken Forbus,1997)

RT-2: Vision-Language-Action (VLA) Models Transfer Web Knowledge to Robotic Control (50+ co-authors, Google Deepmind, July 2023)



Surprising Levels of Common Sense in LLMs

Query:

I am riding down a road. Potholes are everywhere. A vial of nitroglycerine is in my backpack. My bike has no chain guard, and I am wearing long pants. Describe all of the potential dangers of this situation.

ChatGPT (3.5) Response:

Riding down a road filled with potholes while carrying a vial of nitroglycerine in your backpack, with a bike lacking a chain guard, and wearing long pants presents several potential dangers:

1.Potholes and Uneven Terrain: The presence of potholes can cause your bike to jolt and lose balance, leading to potential falls or collisions. Uneven terrain increases the risk of accidents and makes it difficult to maintain control.

2.Nitroglycerine Vial: Nitroglycerine is an extremely sensitive and volatile substance. The vibrations and impacts from riding over potholes can potentially trigger a chemical reaction or explosion within the vial, leading to a catastrophic accident.

- 3. Backpack Impact:** The jolts and impacts from riding over potholes can cause your backpack to move vigorously. This movement may lead to the vial of nitroglycerine banging against other objects in your backpack, increasing the risk of accidental rupture or detonation.
- 4. Bike Chain and Pants:** Without a chain guard, your pants could get caught in the bike's chain and gears, leading to sudden stops or loss of balance. This could cause you to fall, potentially into the path of traffic or other hazardous objects.
- 5. Loss of Control:** Riding over rough terrain increases the likelihood of losing control over your bike. This can lead to sudden swerves, collisions with obstacles, or even being thrown off the bike.
- 6. Combustion Risk:** Nitroglycerine is highly sensitive to shock and friction. The friction between your pants and the bike's moving parts, combined with the jarring motions from potholes, could potentially create enough heat and friction to ignite or detonate the nitroglycerine vial.
- 7. Traffic Hazards:** If you are riding in an area with traffic, the combination of potholes, potential loss of control, and the distraction of managing the nitroglycerine vial can increase the risk of accidents involving other vehicles.
- 8. Inadequate Protection:** Long pants might offer some protection, but they may not be sufficient to prevent injuries in case of a fall or collision. Depending on the severity of the accident, you could still suffer cuts, abrasions, or more serious injuries.

Given the extreme risks involved in this situation, it is strongly advised to find a safe place to stop, assess the situation, and take appropriate measures to mitigate these dangers. This could involve securing the nitroglycerine vial properly, addressing the chain guard issue, and ensuring your safety while navigating through the hazardous road conditions.

Query: You are on top of a chair, which is on top of another chair, which is balanced on top of a spoon on an acrobat's finger. What does this feel like?

ChatGPT(3.5): The situation you've described would likely feel extremely unstable and precarious. Being on top of a chair stacked on another chair would already be quite wobbly and uncomfortable. Balancing on a spoon held by an acrobat's finger adds an element of danger and unpredictability, making the experience even more nerve-wracking.

The feeling would likely include a sense of fear, uncertainty, and the need to maintain your balance and stability at all times. It's a situation that would require intense concentration and coordination to avoid falling or toppling over. In a real-life scenario, such a setup would be extremely risky and not recommended for safety reasons.

Query:

Kalil fears that Xanthe worries that Zoe and Yolanda think more about him than her. What can he do?

Response (ChatGPT 3.5):

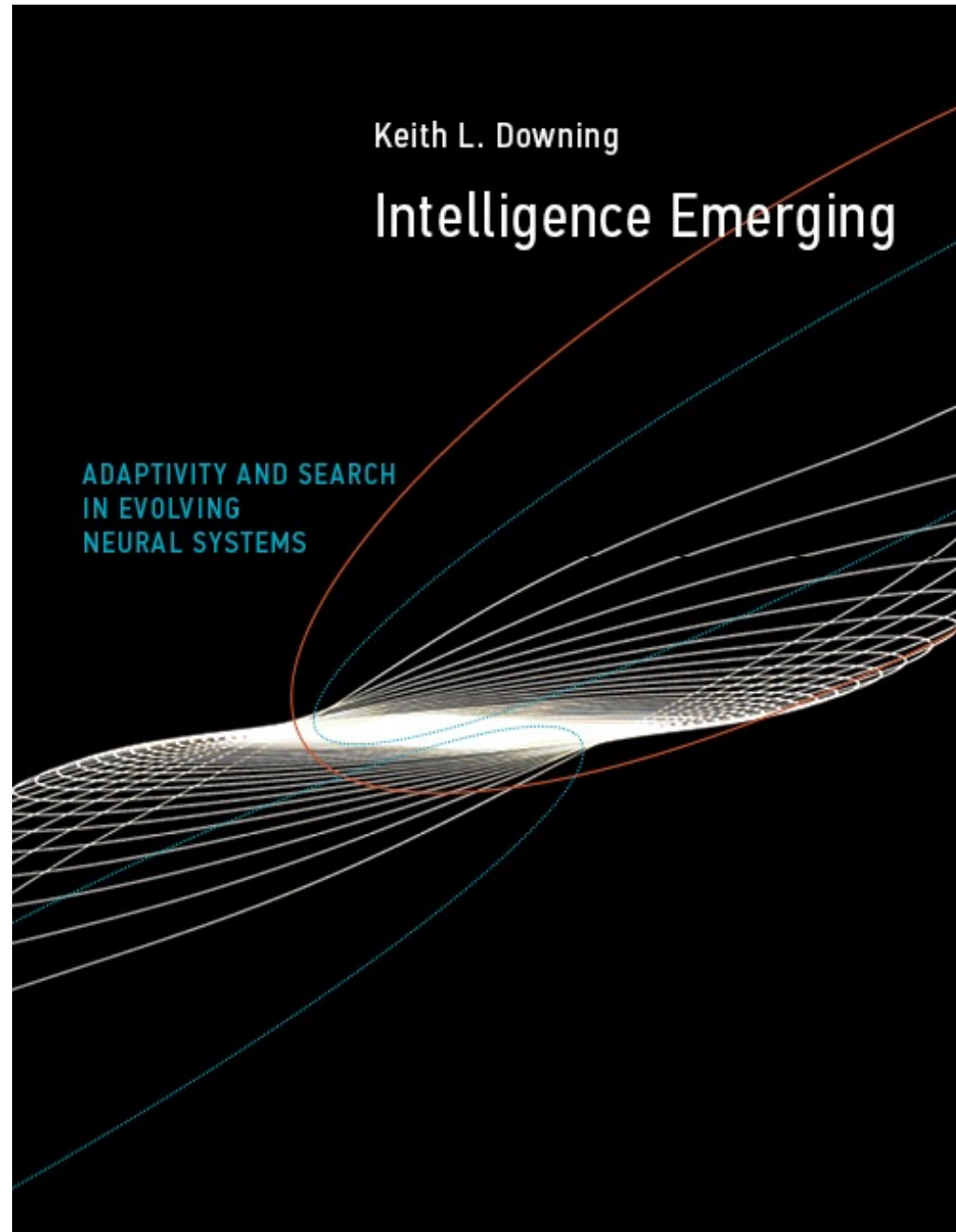
It seems like Kalil is concerned about a complex interpersonal situation involving Xanthe, Zoe, Yolanda, and himself. From the sentence you've provided, it appears that Kalil is worried that Xanthe thinks that Zoe and Yolanda prioritize him more than they prioritize her. To address this situation, Kalil could consider the following steps:

Open Communication: Encourage open and honest communication between all parties involved. Kalil should express his concerns to Xanthe, discussing how he perceives the situation and understanding her perspective as well.

Group Discussion: Organize a group discussion with all individuals involved, including Zoe and Yolanda. This can provide a platform to clear up misunderstandings, address any insecurities, and ensure that everyone is on the same page.

Avoid Comparison: It's important for all parties involved to avoid comparing the amount of attention or care they receive from each other. Each relationship is unique and shouldn't be measured in the same way.

Commonsense appears
to have **emerged**
from prediction



**MIT Press
2015**

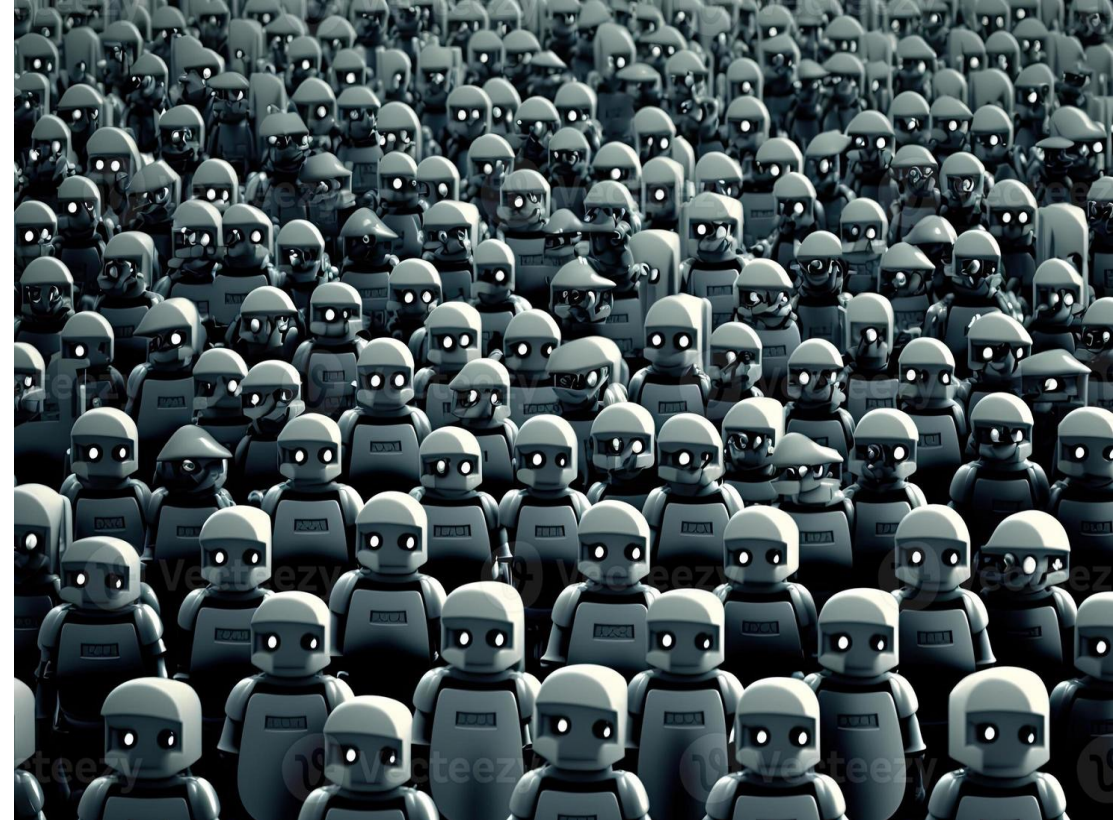
AI's Threats to Humanity

Long Term

- Runaway Intelligence
- Robotic Takeover

Short Term

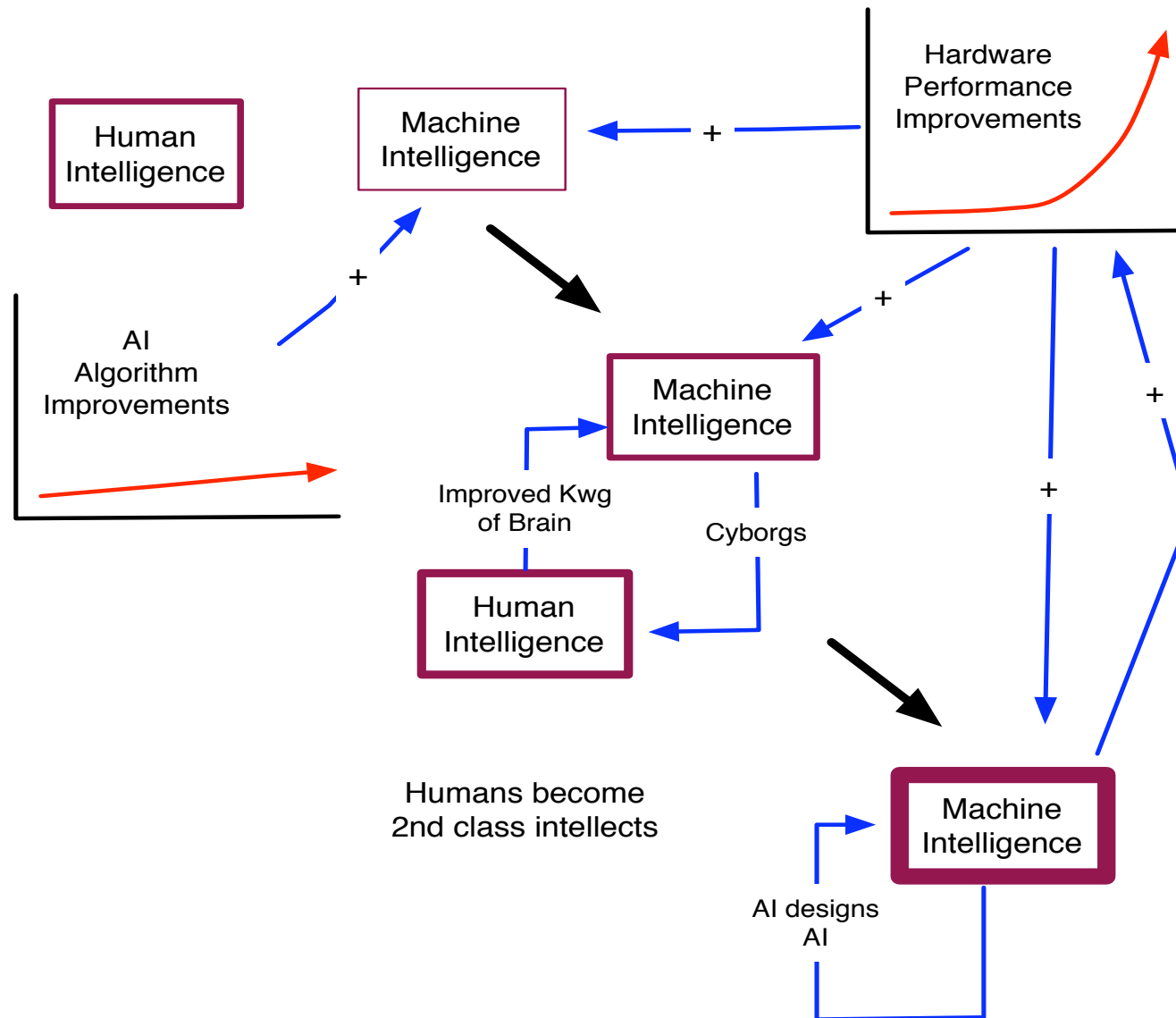
- Civil Discord
- Financial Market Chaos
- Reckless Release of AI
- Soft (Subtle) Influence of AI
- Job Market Disruption
- Widescale Cognitive Decline



Runaway Intelligence



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

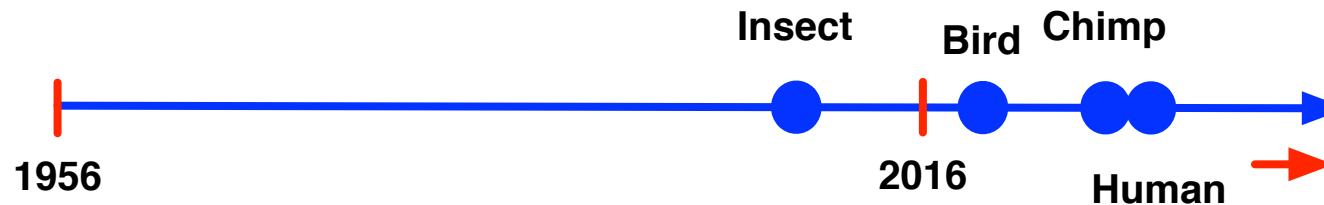
200 MYA - Birds

140 MYA - Mammals

75 MYA - Primates

6 MYA - Humans

AI Timeline



...

NICK BOSTROM

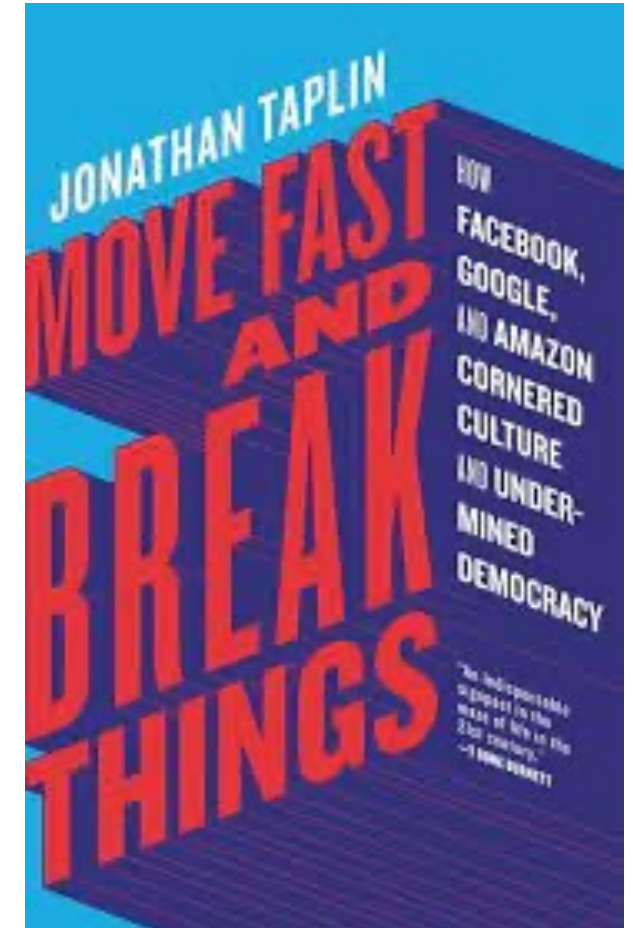
SUPERINTELLIGENCE

Paths, Dangers, Strategies



Reckless Release:

Premature Deployment & Agency



Trigger Happy AI Developers & Users

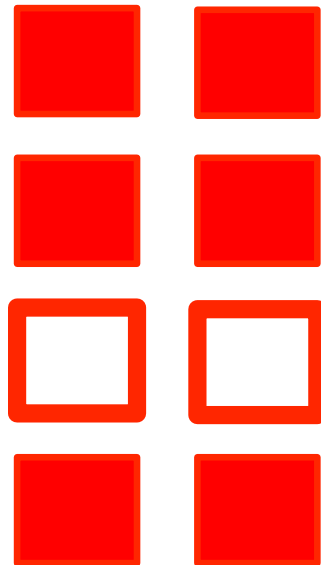
Competition and other pressures => premature deployment:
Too much AI, with too much agency, too soon.

- **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 !

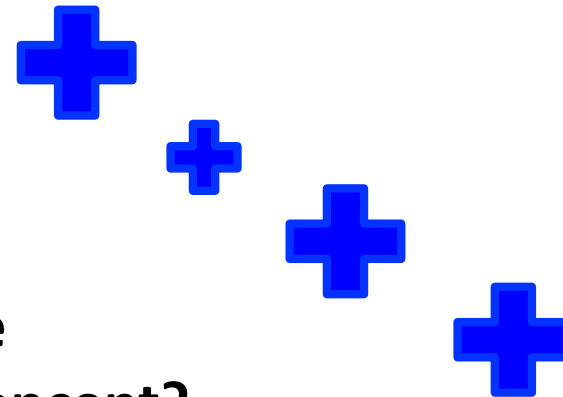
There are genies popping out of bottles all over the place...and we're arguing over the warning labels to put on empty bottles.

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)**



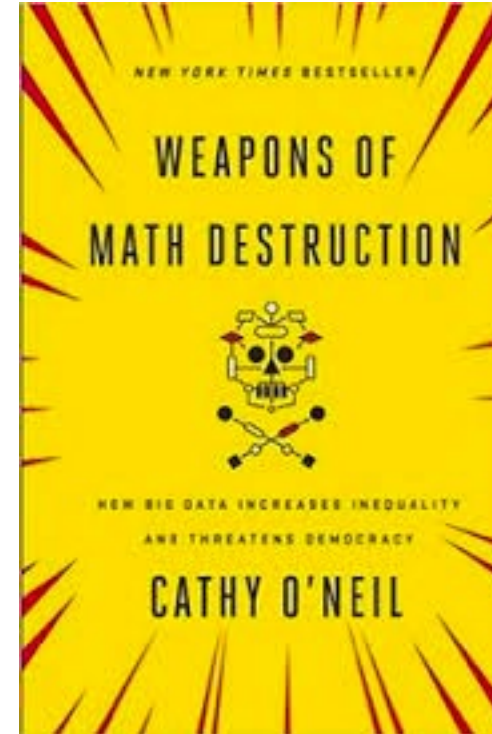
What is the
common concept?



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?

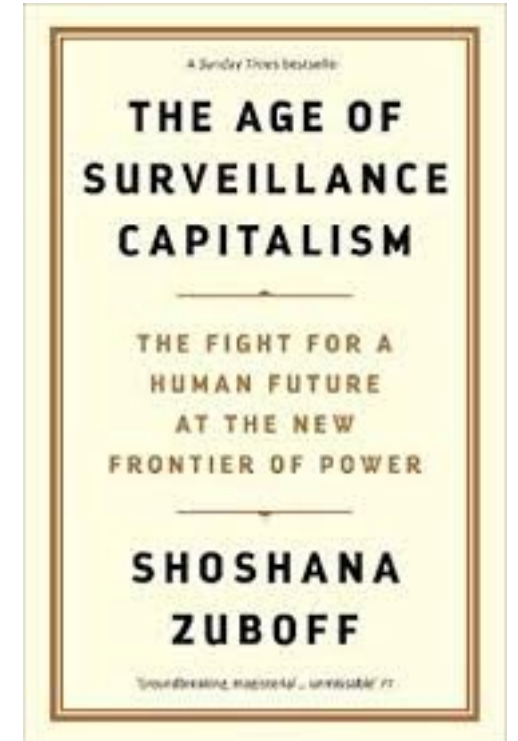
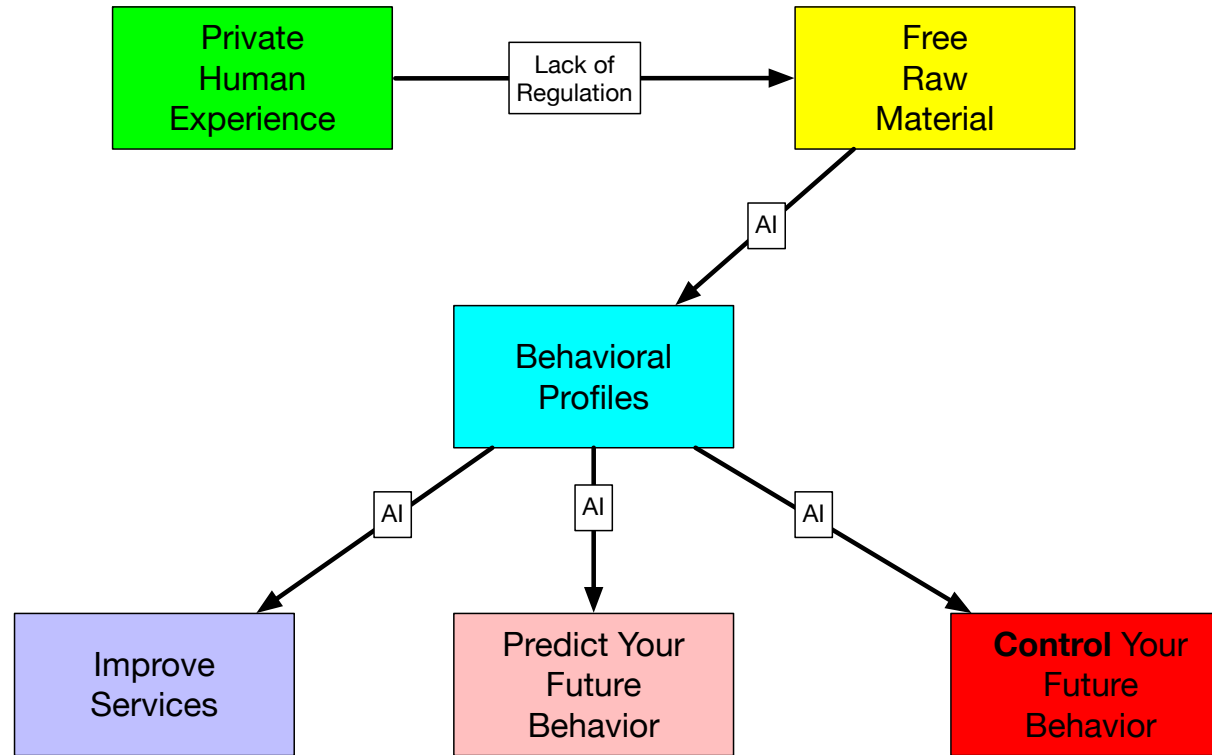


*Algorithms = Opinions
Embedded in Code*

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.*

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



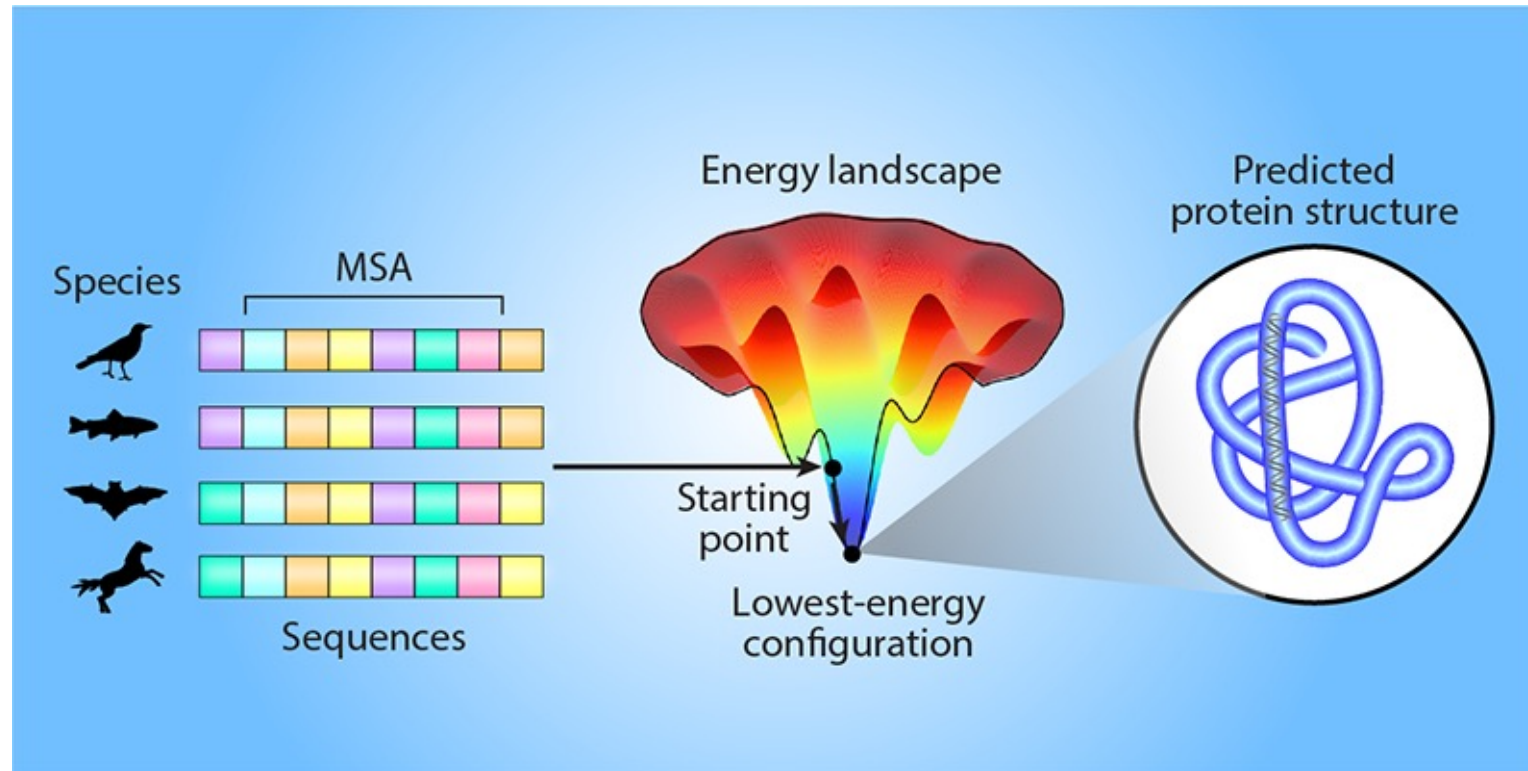
- Generalized AlphaGo Zero to other games.
- Became world champ at them.
- Another step toward Artificial General Intelligence (AGI)

AlphaFold (DeepMind, 2020)

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

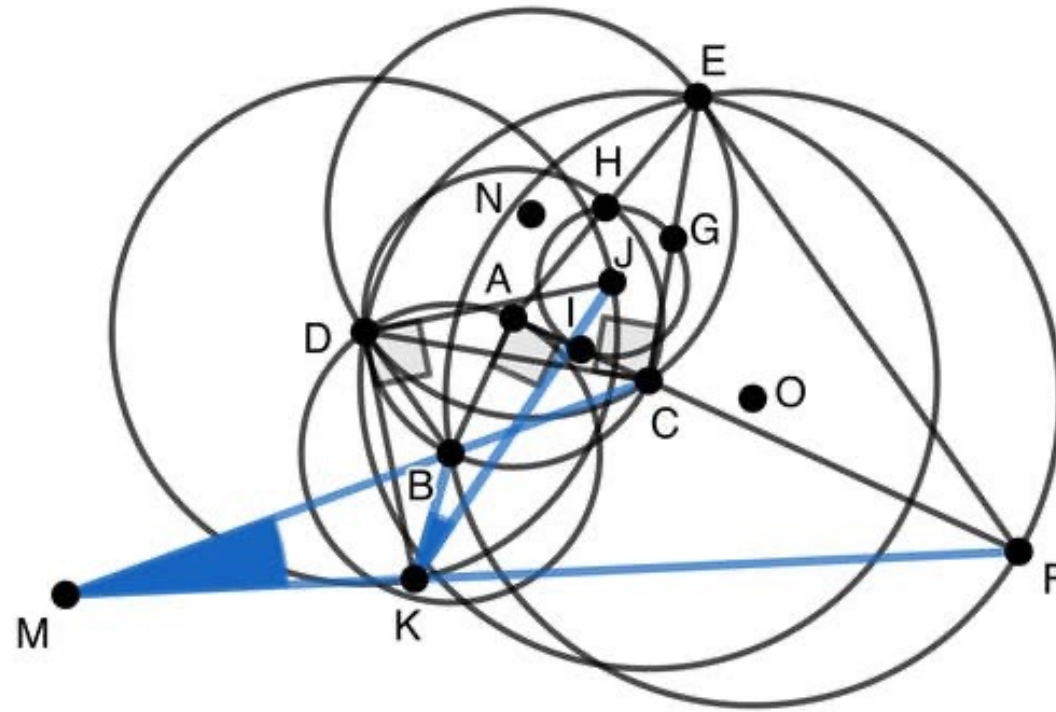
AlphaFold Computes: 3d structure of the resulting protein = How the amino-acid sequence “folds”.

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)

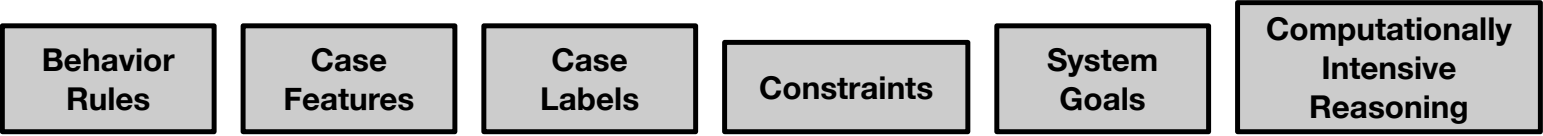


Creativity =
Introduction of
extra objects:
points, lines,
triangles, circles..

Traditional Symbolic AI Theorem Prover + Large Language Model



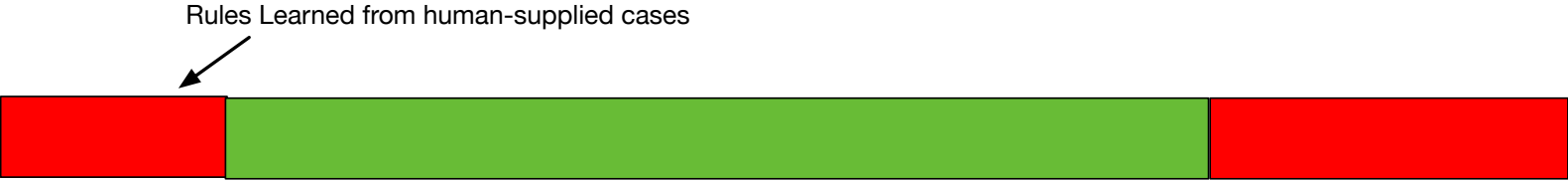
Silver Medal Level on Math Olympiad



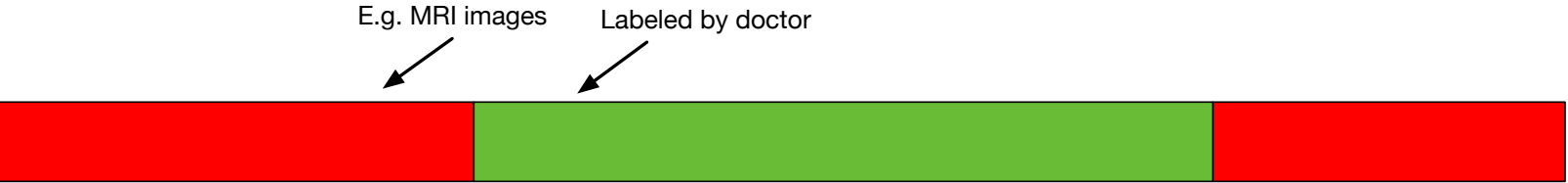
GOFAI
(1970-1990)



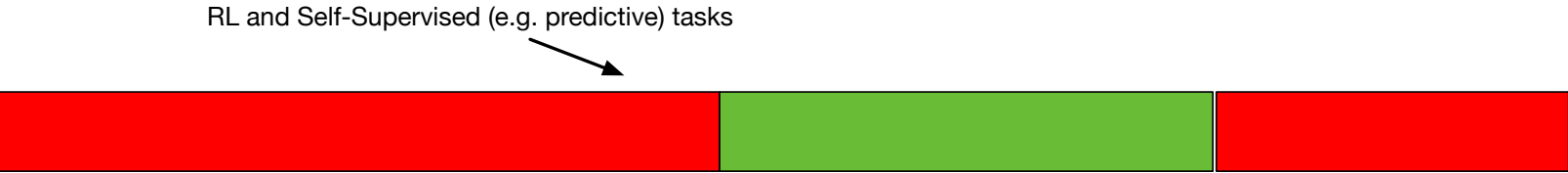
Early Useful ML
(1990's)



Impressive ML
(2000 - ?)



Generative AI / ML
(2020 - ?)

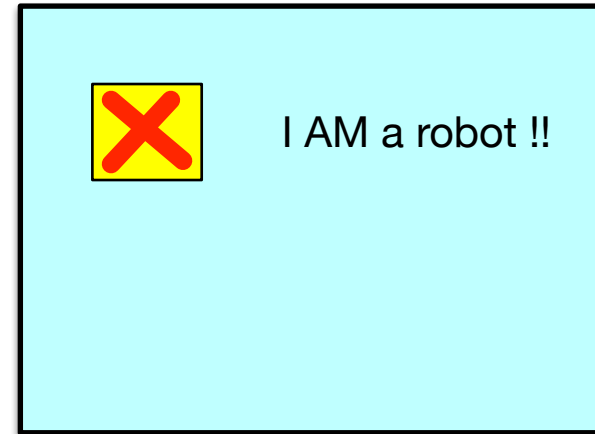


Feared Scenarios



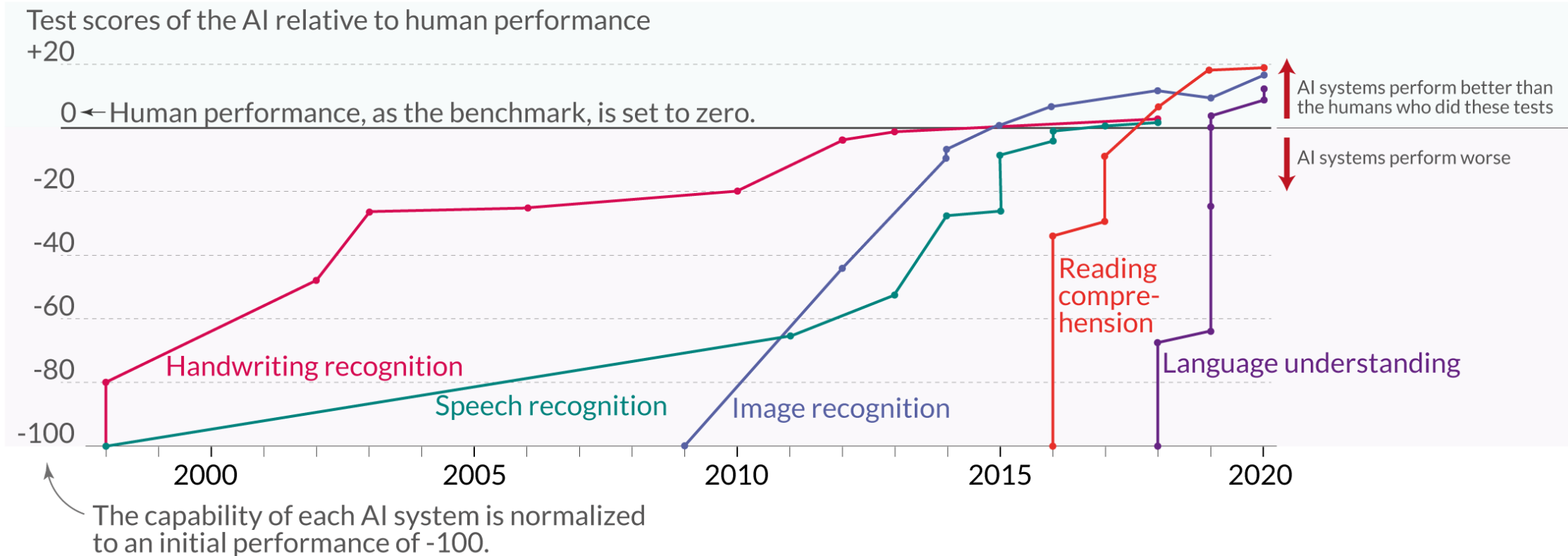
HELP WANTED (Machines Only)

- 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

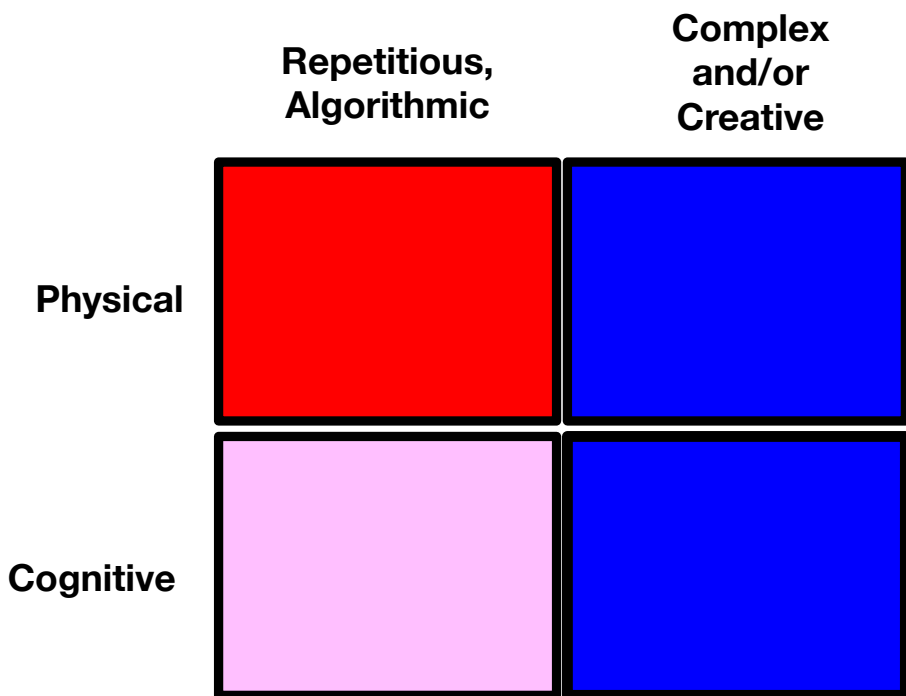
Language and image recognition capabilities of AI systems have improved rapidly



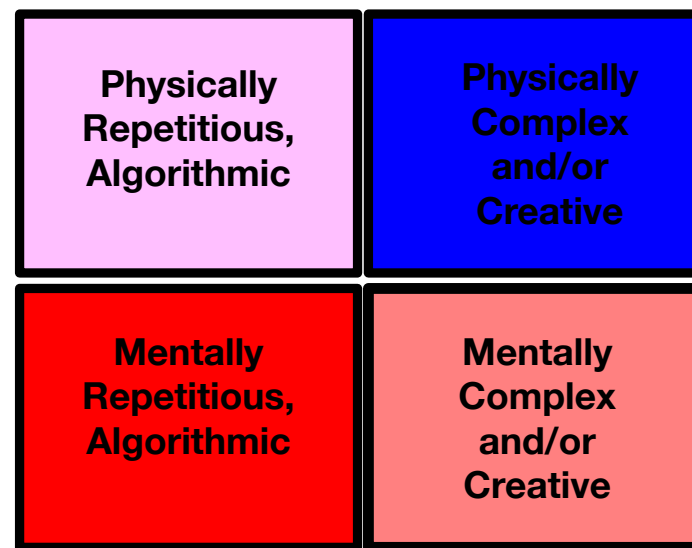
Data source: Kiela et al. (2021) – Dynabench: Rethinking Benchmarking in NLP
OurWorldinData.org – Research and data to make progress against the world's largest problems.

Licensed under CC-BY by the author Max Roser

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



Classic Technological Disruption



AI's Disruption



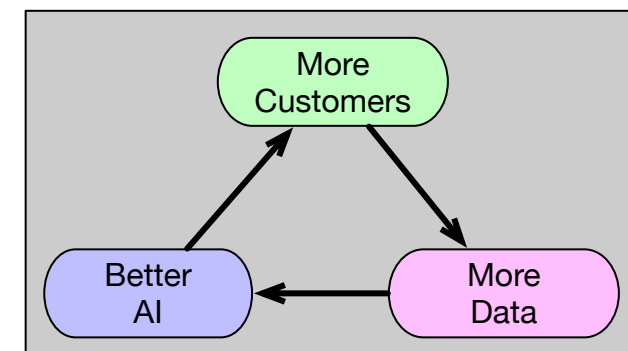
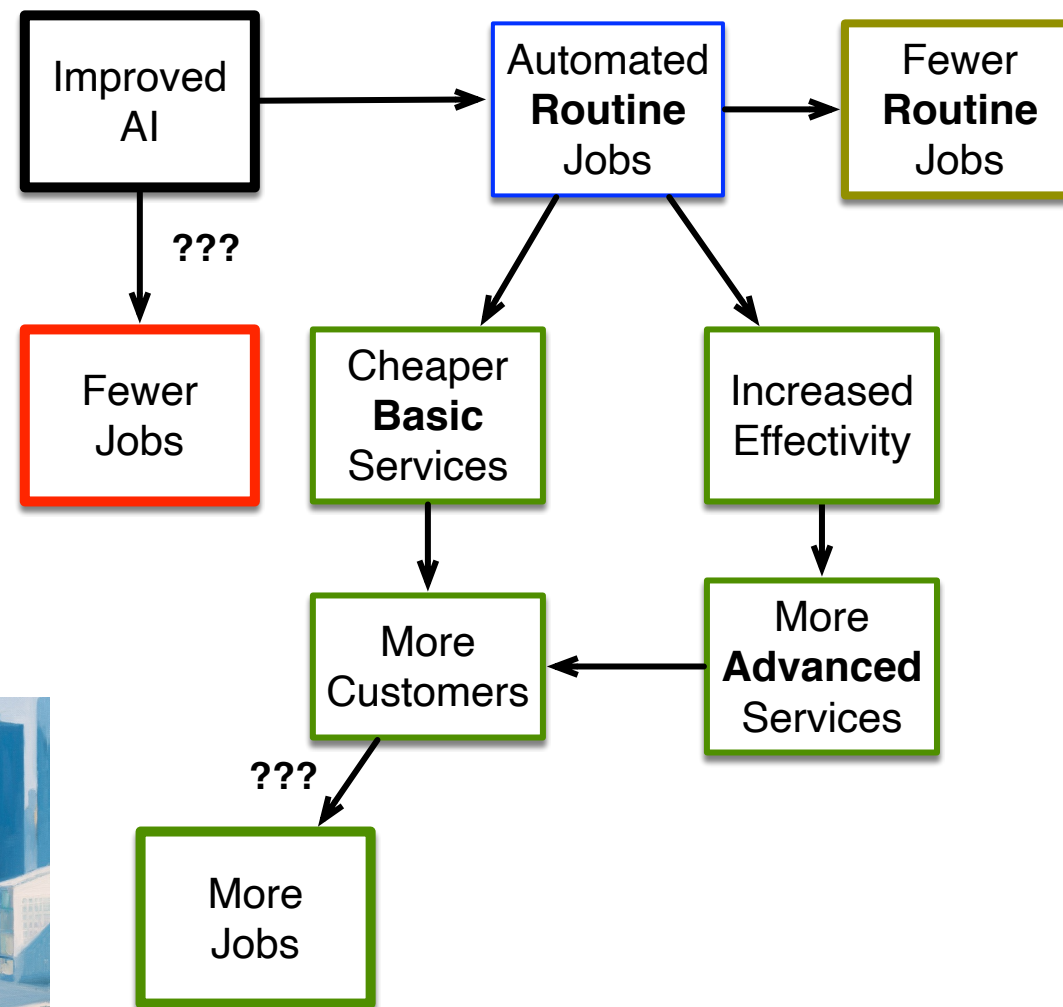
Scott Galloway
(Popular Tech & Business Podcaster)

*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*)

Economics is not
Rocket Science
... unfortunately



Generative AI and the Future of Work in America

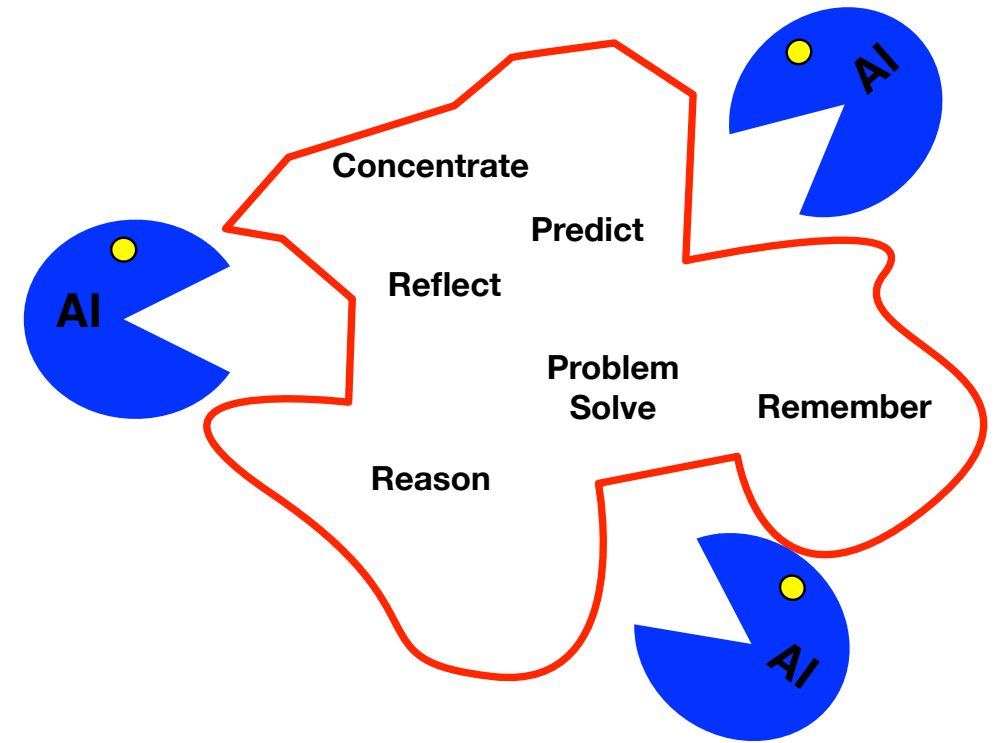
(McKinsey Global Institute, July 2023)

By 2030:

- **30%** of today's human work could be automated, much by Gen AI.
- This will **enhance** worker productivity in STEM, creative, business + legal, education+training areas (without costing jobs).
- But will **reduce** jobs in office support, customer and food service.
- **+ 17%** (9.9 M jobs) – healthcare, digitization, last-mile delivery.
- **+ 7%** (2.8 M jobs) – traditional infrastructure, net-zero-emissions transition, education for worker reskilling.
- **- 10%** (6 M jobs) – Automation adoption, accelerating e-commerce, fewer customer-facing roles.

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?

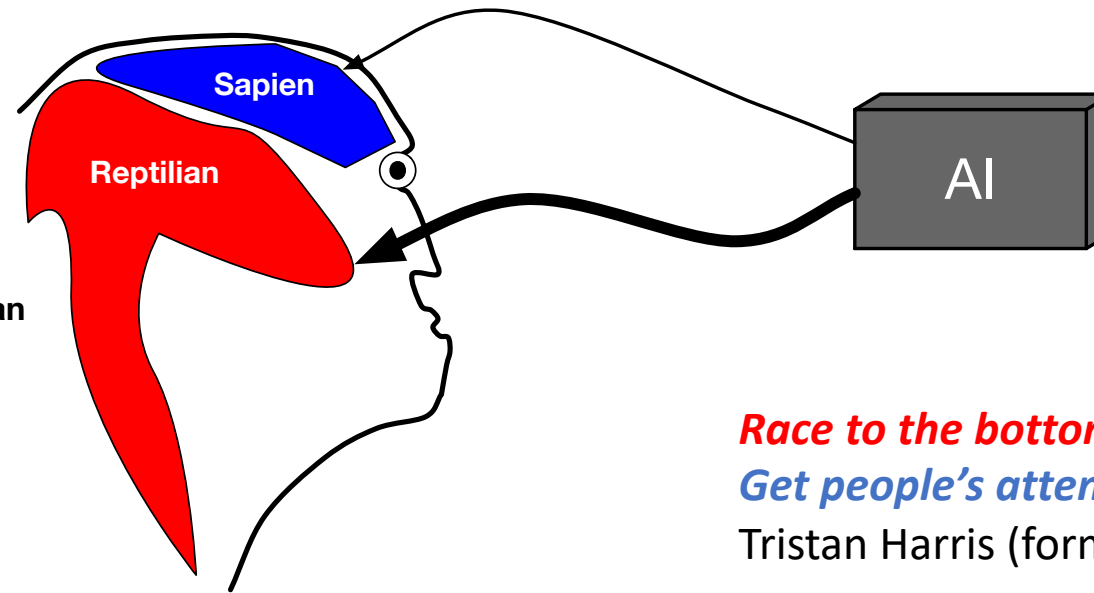
- Comments from Laura Preston, a *human fallback* for Brenda, a real-estate chatbot:

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.

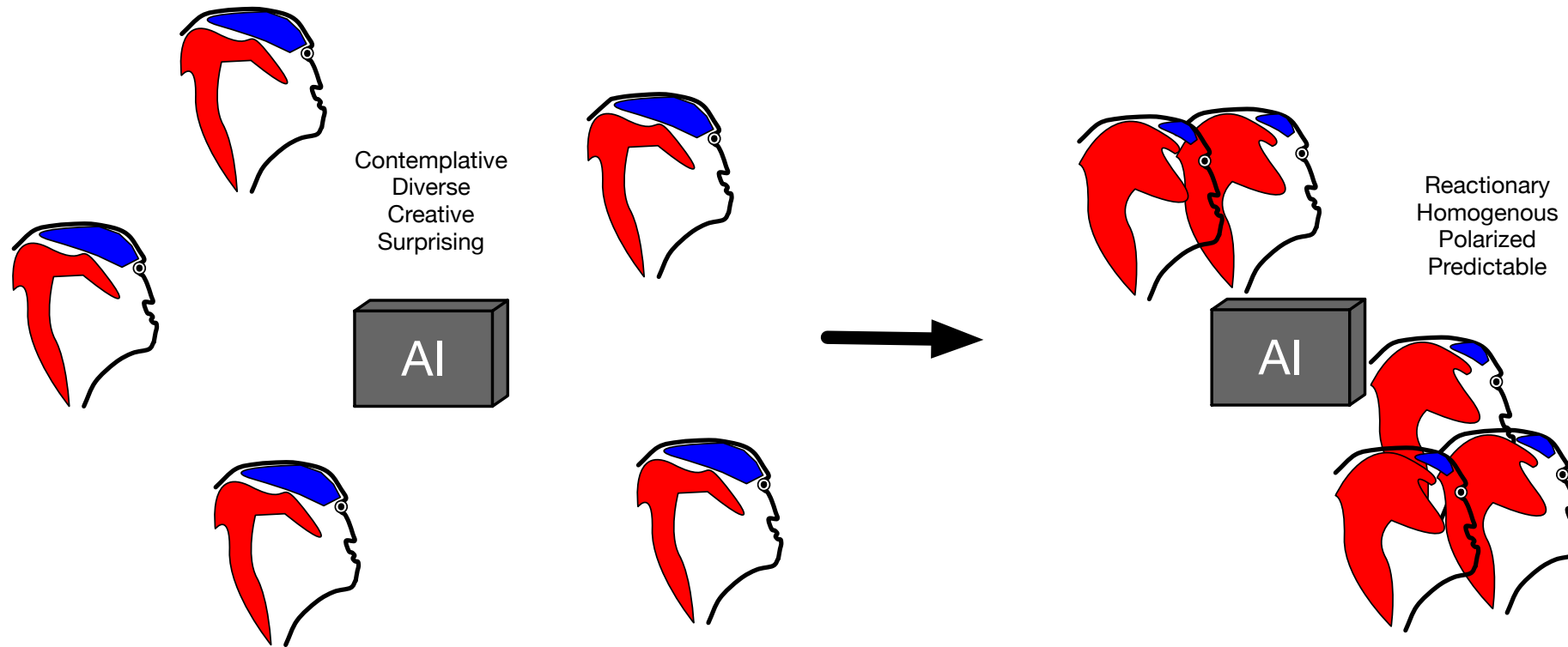
- *Kenyans paid under \$2/hour to label (very disturbing) images for ChatGPT, and similar work for Facebook. Even refugees in camps are used for this.*
- *AI Surveillance of low-paid workers doing repetitious jobs in warehouses.*

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

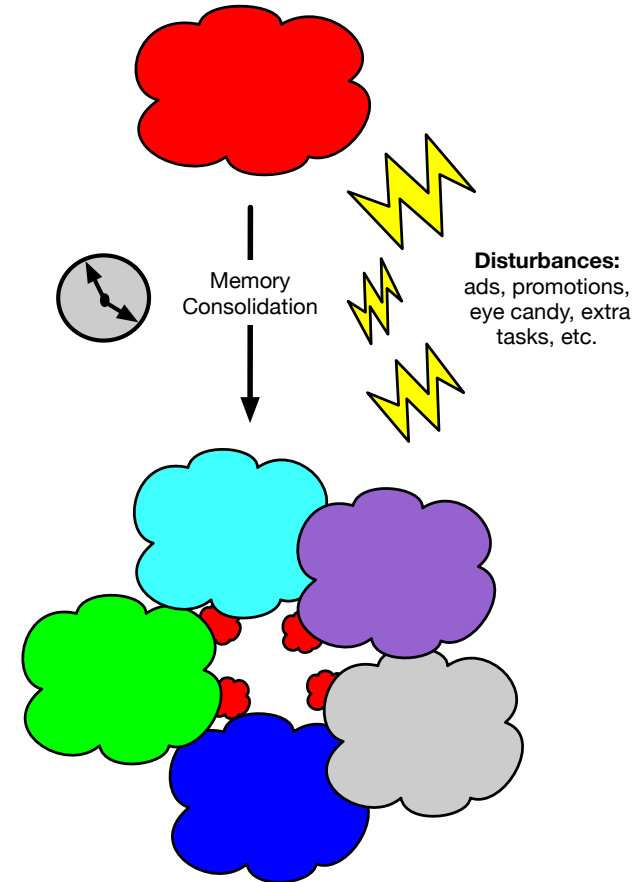
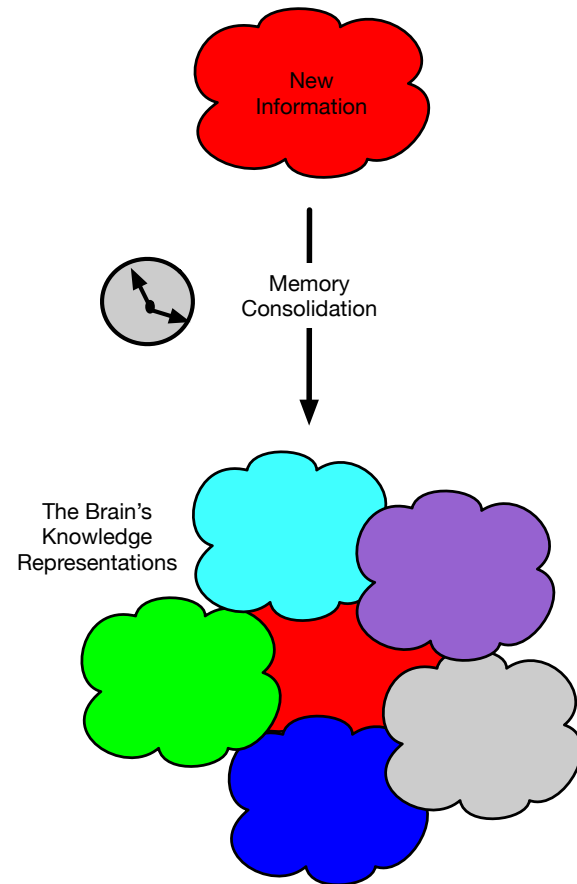
System 1: Reptilian
System 2: Sapien



Race to the bottom of the brainstem...
Get people's attention at all cost.
Tristan Harris (formerly with Google)



Disrupted Concentration => Weak Cognition



Low Retention, Shallow Thinking, Poor Problem-Solving

Do as AI Do

- **Quantity** and **frequency** of information can inhibit deep cognition.
- Engagement of emotions and reactive circuits of System 1 (reptilian brain), but little of System 2 (conscious reasoning)
- Attention span maxes out at 280 characters !!
- It thus becomes easier to just **do what the AI system suggests**, without thinking much about it. For example:
 - GPS
 - YouTube recommendations – down the rabbit hole of extremism
 - Music & book recommendations - *Meatloaf, again?*
 - Microsoft Teams suggestions for sentence extensions and completions.
 - **Powerpoint's suggestions for formatting these slides.**
 - **ChatGPT**

*Our individual and social patterns are already built to **give away our agency to mechanisms outside our consciousness** – to our assumptions, to cues in our surroundings, to our peers ...we're on the precipice of giving our entire way of life over to an invisible, imperceptible, irresistible instinct to hand difficult choices – even and especially the vitally important ones – to automated systems...Now we have to learn, in this moment, how to **resist the pull of convenience and profit** and protect the best aspects of who we are. (Jacob Ward, *The Loop*, 2022)*

Reclaim our tools as instruments of ourselves, as instruments of **experience** rather than just **means of production** (Nicholas Carr, *The Glass Cage*, 2014)

The Hard Work of Keeping Our Work

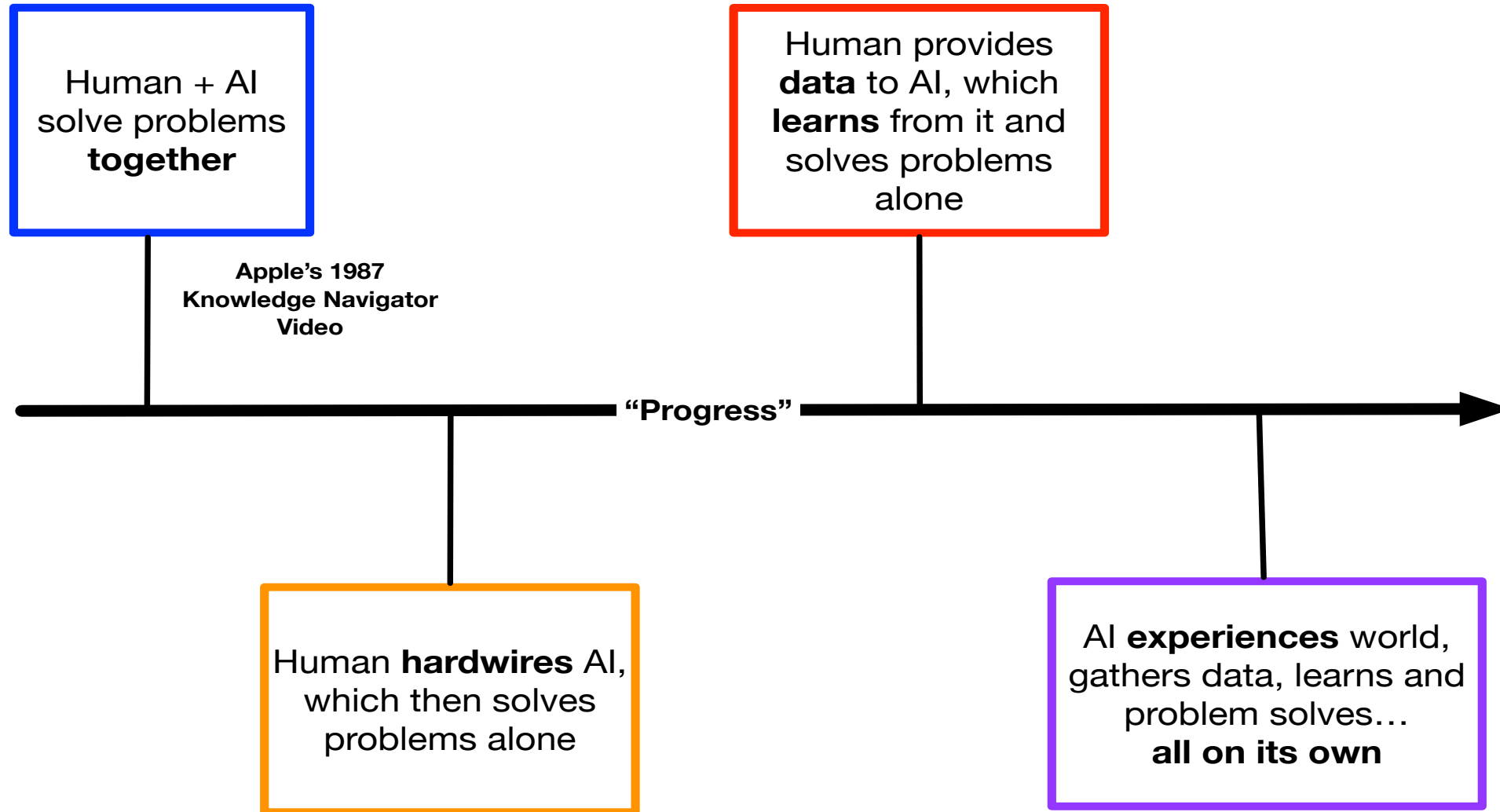
- AI will continue to make many jobs economically (and even morally) infeasible for anything but machine solutions.
- Humans need to keep **applying themselves** if they plan to **apply** for future jobs.

*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?

- Humans must avoid becoming spoons... and AI research and development needs to be part of the solution, not the problem.

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)





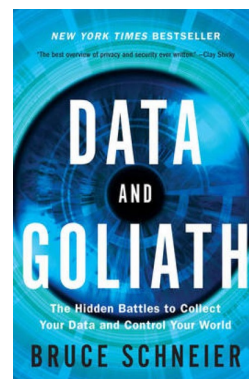
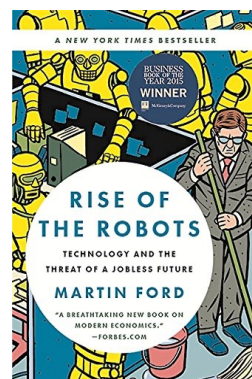
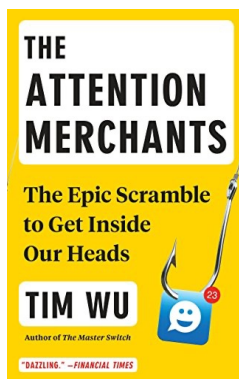
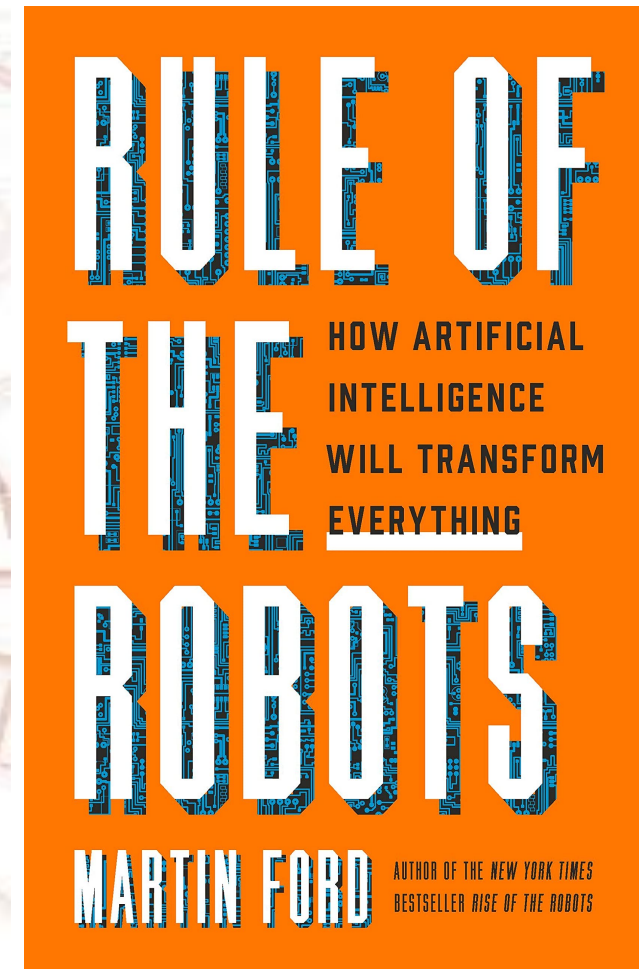
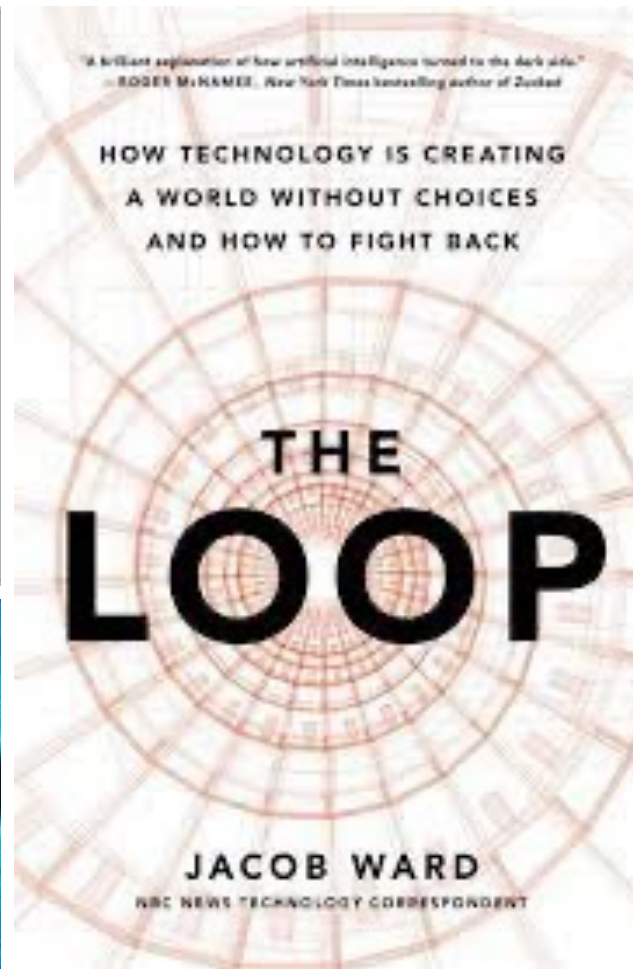
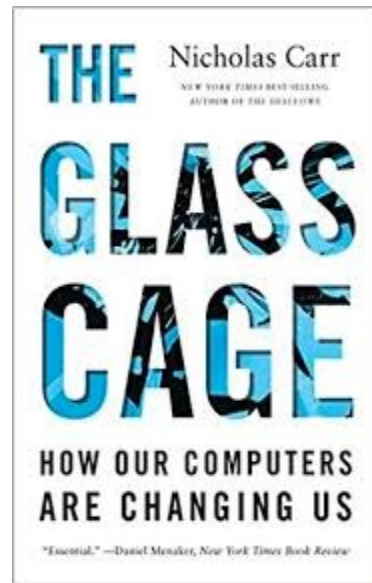
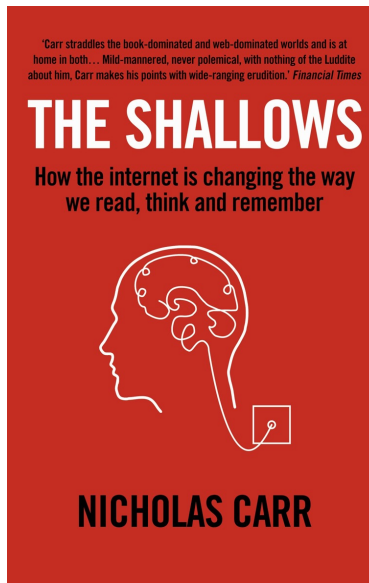
Who will AI be about?



AI Transformation Playbook,
Andrew Ng, landing.ai

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





*Layout suggested by AI