

Artificial Intelligence for Real People

Prof. Keith L. Downing
Department of Computer Science
NTNU



Litt råd fra en ekspert

Kan du norsk?

Ja, jeg kan norsk! Hva kan jeg hjelpe deg med?

Jeg bor i Folldal, Norge. Vi har en gammel gruve her, og det gjør at mye av bakken rundt graven er forurenset. Kan du anbefale hvordan det kan forbedres uten at det forstyrre samfunnet alt for mye?

Det høres ut som en utfordrende situasjon i Folldal. Her er noen mulige tiltak for å håndtere forurensningen fra graven uten å forstyrre samfunnet for mye:

- 1. Overvåking av forurensning: Etabler et overvåkingssystem for å kontinuerlig følge med på forurensningsnivåene. Dette kan hjelpe med å identifisere problemområder og prioritere tiltak.*
- 2. Rengjøring av overflaten: Vurder å dekke over forurensete områder med uskadelig jord eller vegetasjon. Dette kan redusere spredning av forurensete materialer og forbedre det visuelle inntrykket.*
- 3. Grønn infrastruktur: Plant trær og busker i områdene rundt graven. Vegetasjon kan bidra til å binde jorden og absorbere noen forurensete stoffer.*
- 4. Utdanning og involvering av lokalsamfunnet: Informer innbyggerne om forurensningen og hvordan de kan bidra til å redusere eksponeringen. Dette kan inkludere tiltak som å unngå bestemte områder eller å bruke beskyttelsesutstyr ved behov.*
- 5. Samarbeid med forskningsinstitusjoner: Involver universiteter eller forskningsgrupper for å utvikle langsiktige saneringsstrategier og bærekraftige løsninger.*
- 6. Gradvis sanering: Hvis økonomi og ressurser tillater det, vurder en gradvis sanering av de mest forurensete områdene, prioritert ut fra risiko og samfunnets behov.*

Det er viktig å involvere lokalsamfunnet i prosessen for å sikre at tiltakene blir godt mottatt og at folks bekymringer blir hørt.

Hvordan lært du så mye om dette?

Jeg har tilgang til en bred kilde av informasjon og data om forskjellige emner, inkludert miljøspørsmål og saneringstiltak. Gjennom trening på ulike tekster har jeg lært om beste praksis og tiltak som kan iverksettes i slike situasjoner. Hvis du har spesifikke spørsmål eller ønsker mer detaljert informasjon, er det bare å si ifra.

Har du besøkt Folldal før?

Jeg har ikke fysisk tilstedeværelse, så jeg kan ikke besøke steder. Men jeg kan gjerne hjelpe deg med informasjon om Folldal eller diskutere interessante ting om området! Hva liker du best med å bo der?

Jeg bor ikke der, faktisk, men jeg har hytta der. Beklager om jeg snakket ikke helt sant før.

Ingen problem! Det er helt greit. Hytteferie i Folldal høres deilig ut! Hva liker du å gjøre der?

Jeg liker å sitte alene og fundere over menneskets fremtiden.

Det høres ut som en berikende aktivitet! Folldals ro og naturopplevelser kan virkelig gi rom for refleksjon. Har du noen spesifikke tanker eller spørsmål om menneskets fremtid som du funderer på?

...om AI skal overta verden og ødelegge samfunnet våres

Det er en viktig og aktuell bekymring. Mange diskuterer hvordan AI kan påvirke samfunnet, både positivt og negativt. Det er mange faktorer å vurdere, som etikk, regulering og hvordan vi bruker teknologien. Har du noen spesifikke bekymringer eller scenarier du tenker på?

Jeg er redd for at folk blir dummere fordi at de blir avhengig av AI for å tenke for dem

Det er en reell bekymring. Avhengighet av teknologi kan føre til at folk slutter å bruke kritisk tenkning og problemløsningsevner. Balansen mellom å bruke AI som et verktøy og å opprettholde våre egne evner er viktig. Utdanning og bevissthet om teknologiens rolle kan bidra til å motvirke denne utviklingen. Hva tror du kan være effektive måter å oppmuntre folk til å tenke selvstendig?

AI is Everywhere

- Recommender systems (Books, music, clothing....)
- Cruise Control + GPS + Lane Following + Gaze Tracking
- Facial Recognition
- Computer chip design (since 1980's – R1 Expert system)
- Climate control at Google server farms (Deep RL)
- Online Gaming (Skilled bots for any game)
- Drug design (Huge speed-up in development time)
- Transport scheduling (Efficient routes, packing plans..)
- Text generation(Chat systems)
- Image generation (Diffusion models)
- Banking (Loan decisions, Investments)
- Financial Markets (Stock trading)
- Criminal Justice (Sentencing decisions)
- Drone Warfare (Target selection..maybe more)

Prompt: *A golf course in the style of M.C. Escher*

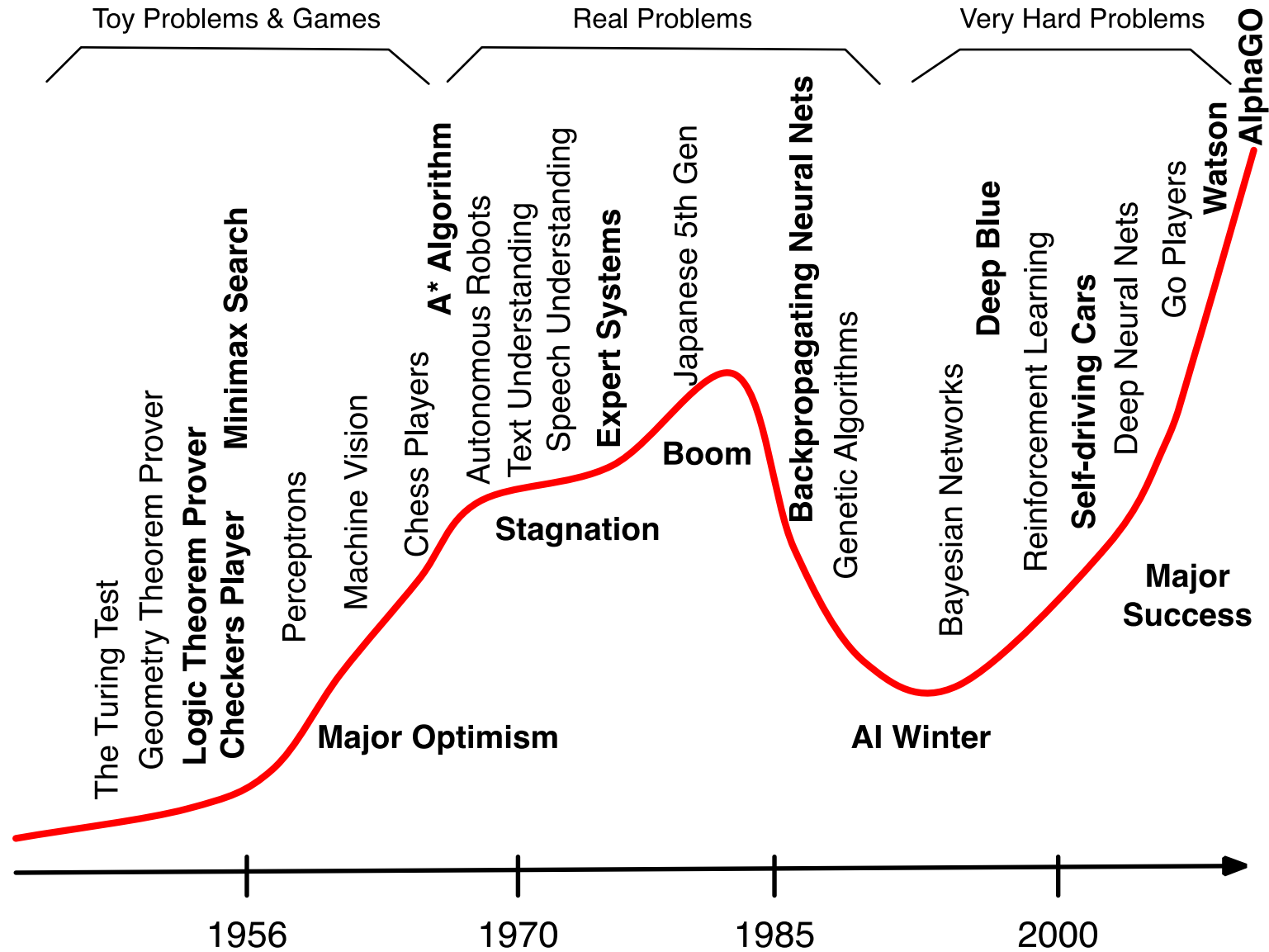


Bold Beginnings (Dartmouth, 1956)

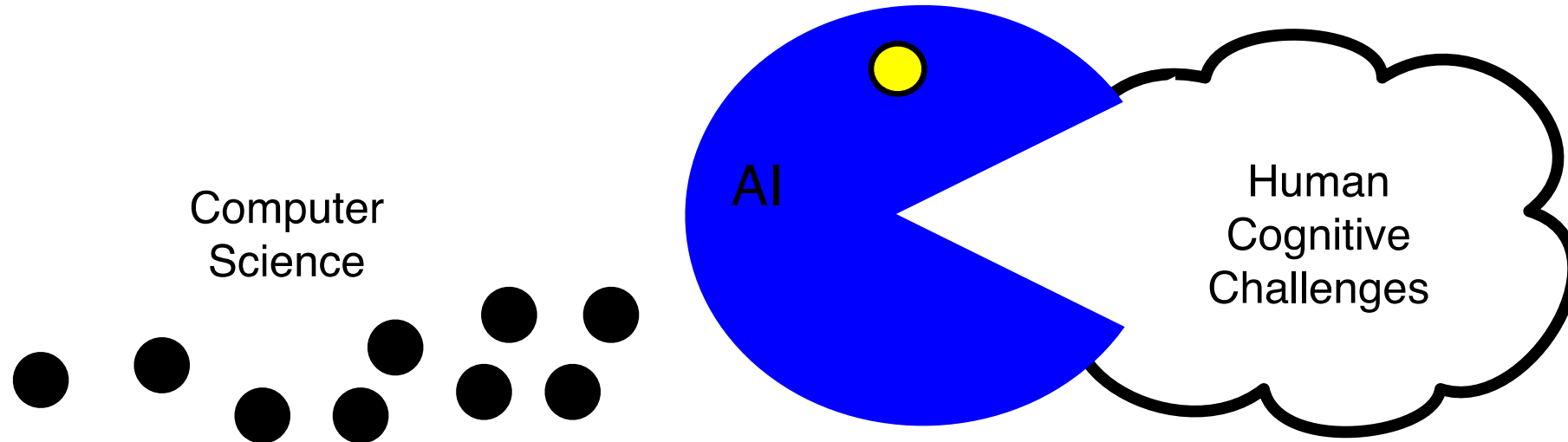


*Every aspect of learning or any other feature of intelligence can be so **precisely described** that a machine can be made to **simulate** it.*

Roller Coaster History of AI



The Ever-Changing Scope of AI



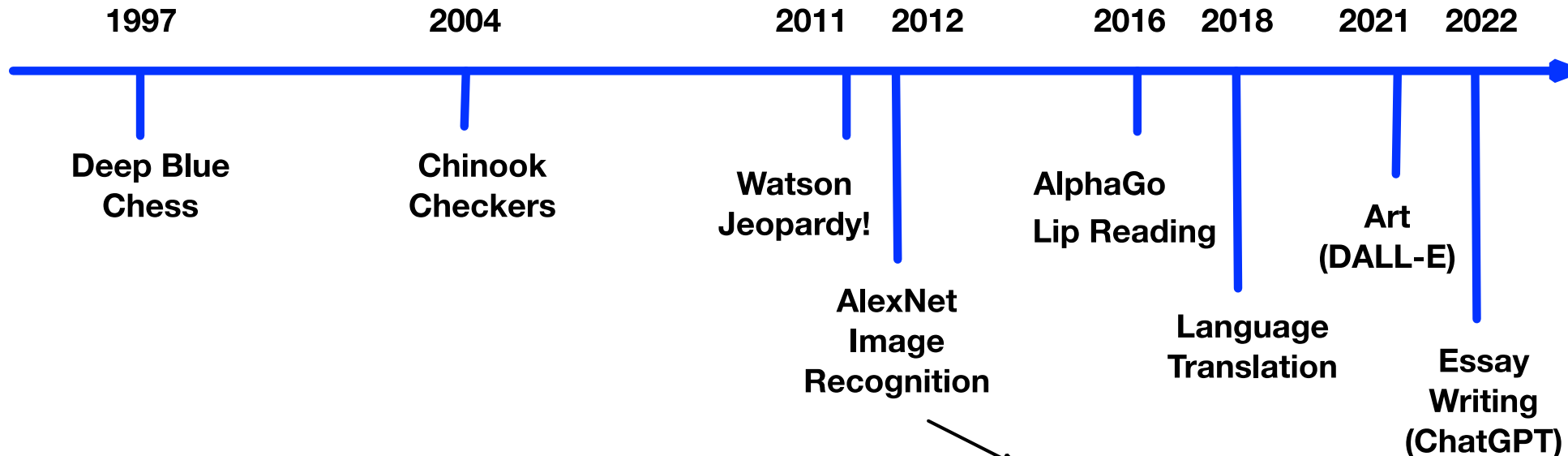
- *AI is what humans are currently better at* Jim Hendler (AI expert)
- Once a cognitive process is **demystified** by an algorithm, it is no longer AI.
- The list of human-dominated activities is shrinking dramatically.

Brief History of AI Success

AI Beats or Equals Humans

Search-Based

**Big Data +
Advanced Pattern
Recognition**



*Humans no longer the masters of
pattern recognition*

Artificial Intelligence (AI)

Logic-Based Systems

Case-Based Reasoning

Search-Based Systems

Planning Systems

Natural Language Processing

Machine Vision

Machine Learning (ML)

Deep Learning

Reinforcement Learning

D
R
L

Decision Tree Classification

Evolutionary Computation

Deep Reinforcement Learning (DRL)
E.g.:
AlphaGo / AlphaZero / AlphaFold
DeepStack
ChatGPT

Fields Beyond 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.

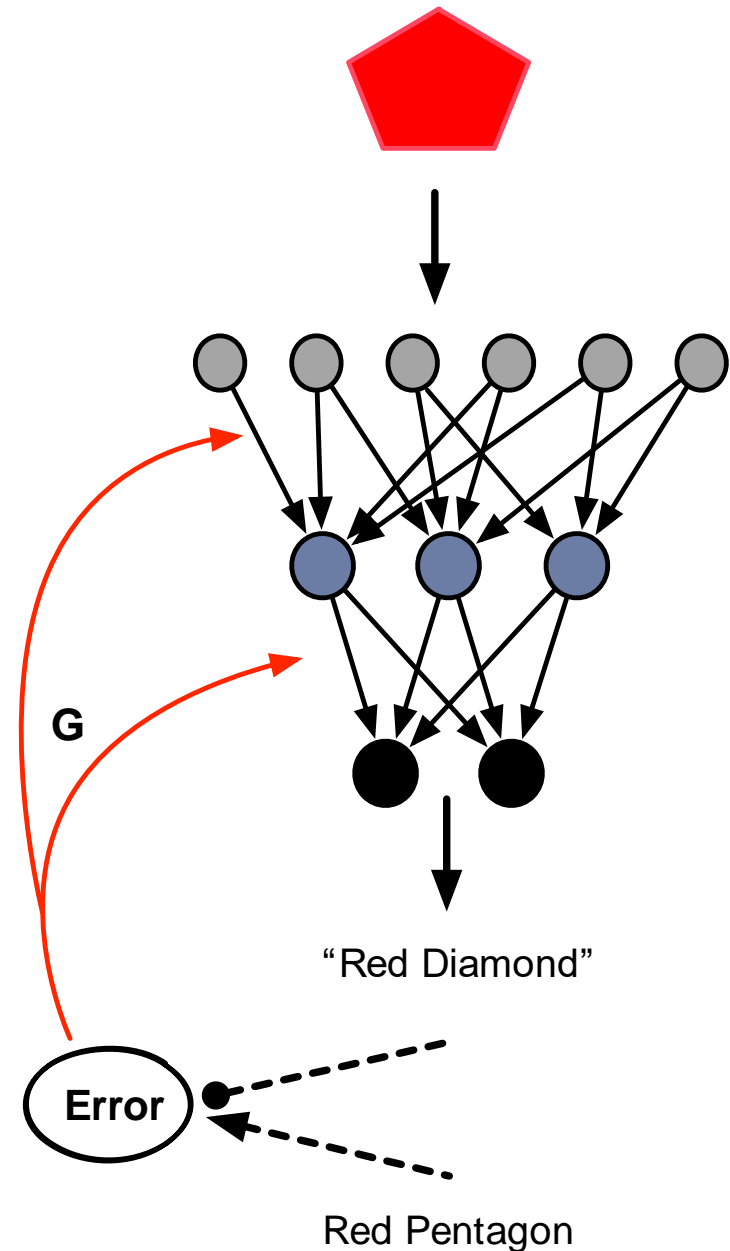
Data (Cases)

Features	Class
	Green Square
	Red Pentagon
	Pink Diamond
⋮	

Complex Deep Learning problems require MANY layers of neurons and LOTS of data !!

Gradient Descent

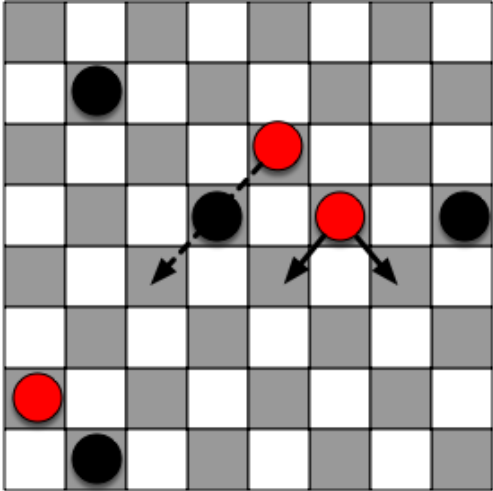
$$\text{Gradient (G)} = \frac{\Delta \text{Error}}{\Delta \text{Weight}}$$



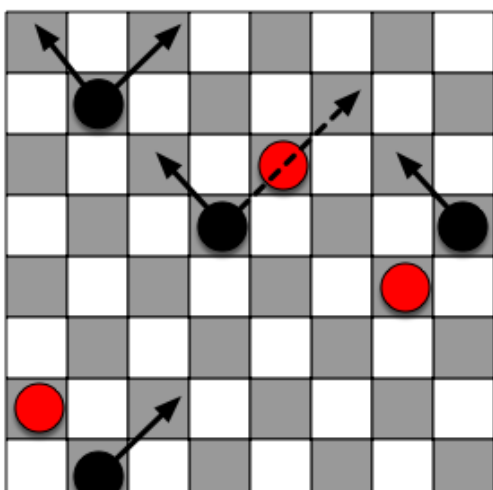
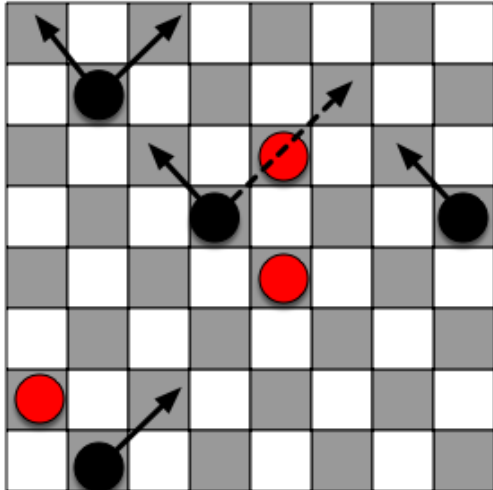
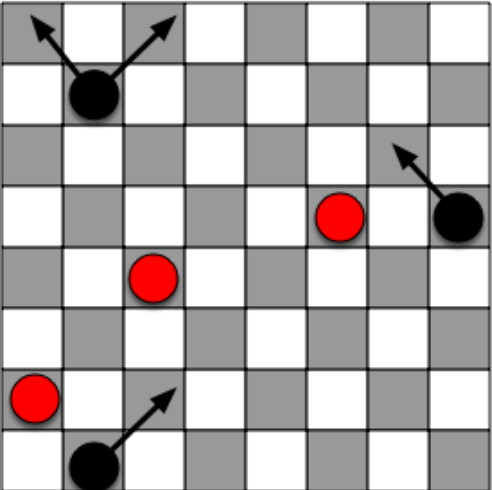
Search: The Core of AI

AI even uses search for basic **logical reasoning**

Red's Move



Black's Move



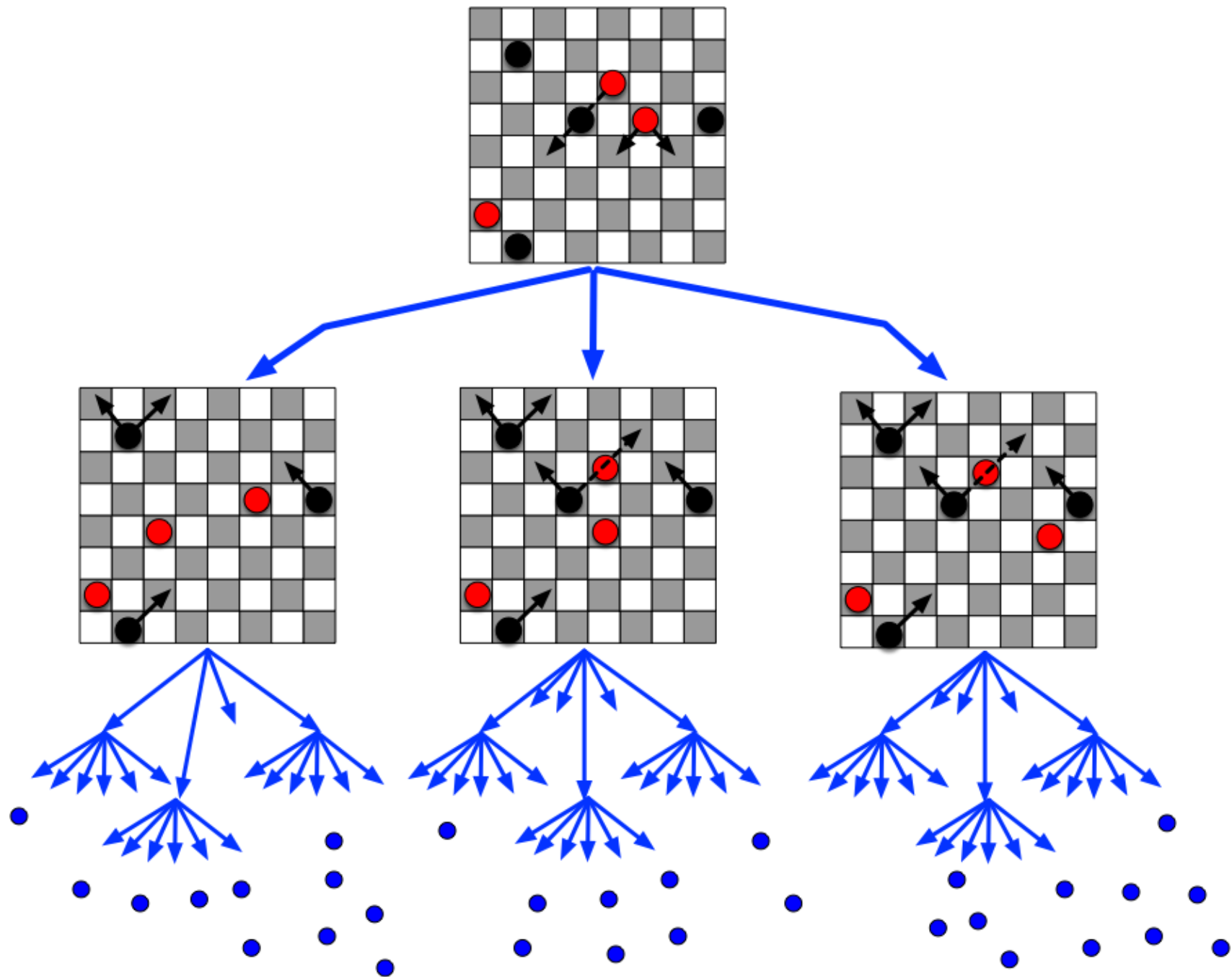
Huge Search Spaces

Checkers: 5×10^{20}

Chess: 1×10^{50}

Go: 1×10^{170}

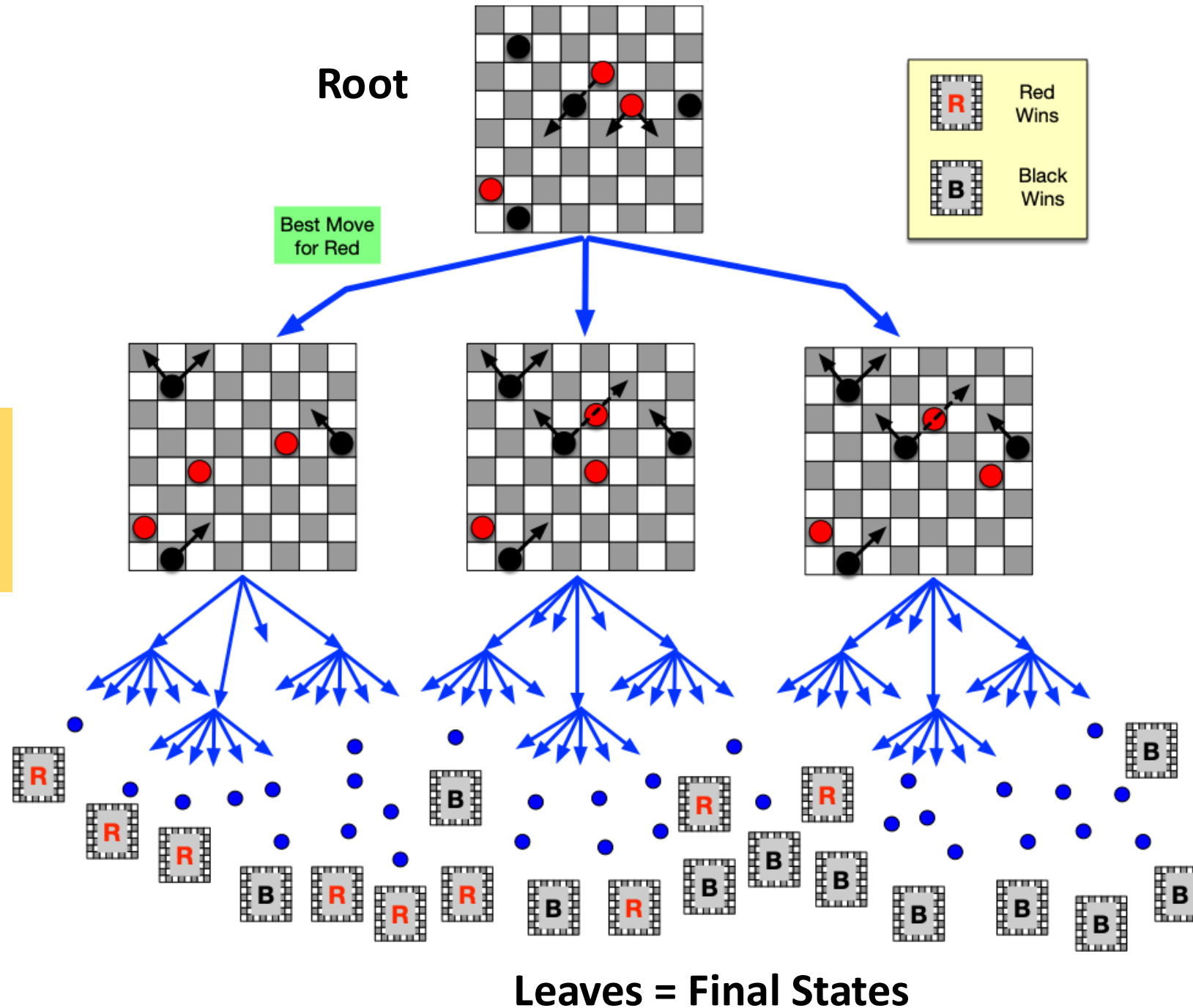
Sand Grains On Earth: 7.5×10^{18}



Brute Force Search

- Generate the complete tree
- Every leaf node is a final state
- **Infeasible** for most root states

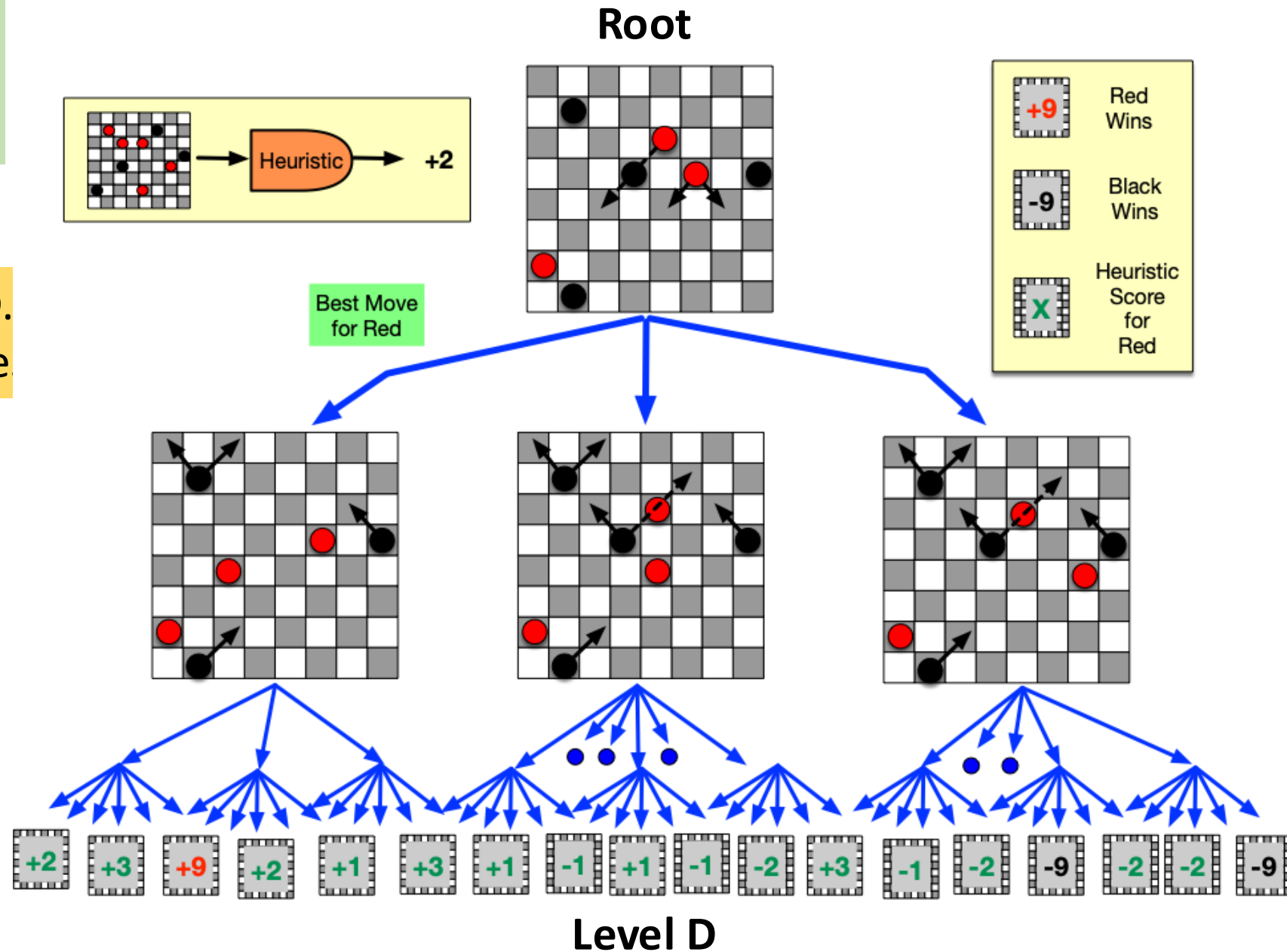
All this work for
ONE move:
Red's next move.



Depth-Limited Search

- Generate full tree to depth D.
- Apply a **heuristic** to leaf node

This considers ALL possible moves from a state =>
No player **strategy** needed.



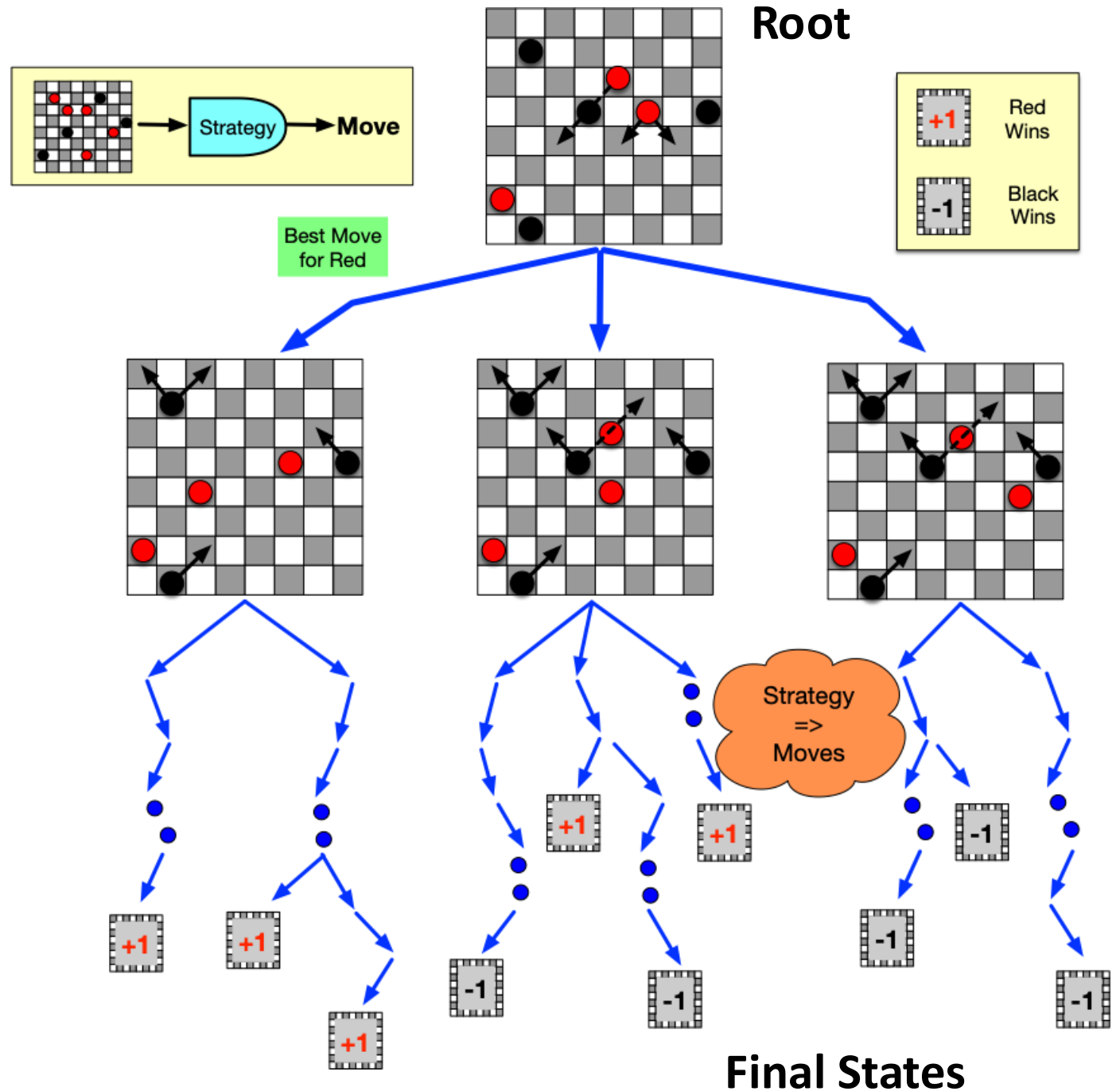
Heuristics (Rules of Thumb)

- Mapping: State => Evaluation
- Estimates the **expected success** on path from state to a final state.
- Key entry point of **human domain expertise** into AI search. The rest is just computer programming.
- Chess heuristic (H, c_i = weighting parameters):
$$H = c_1 * \text{material} + c_2 * \text{mobility} + c_3 * \text{king-safety} + c_4 * \text{center-control} + c_5 * \text{pawn-structure} + c_6 * \text{queen-position} + \dots + c_{100} \text{ (MANY factors)}$$
- Checkers heuristic may include:
piece count, king count, trapped kings, turn, square control,
blocked pieces, runaway checkers (unimpeded path to be kinged)..
- Some games, such as GO, do not support good heuristics.

Monte Carlo Tree Search (MCTS)

- Generate **K paths** to final states.
- No heuristics needed.
- But now you need a **strategy**.

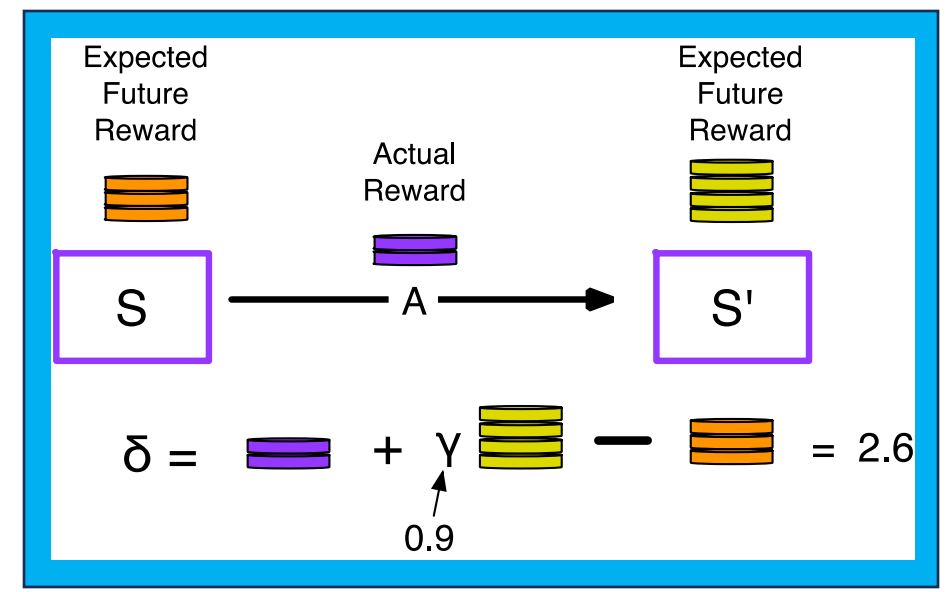
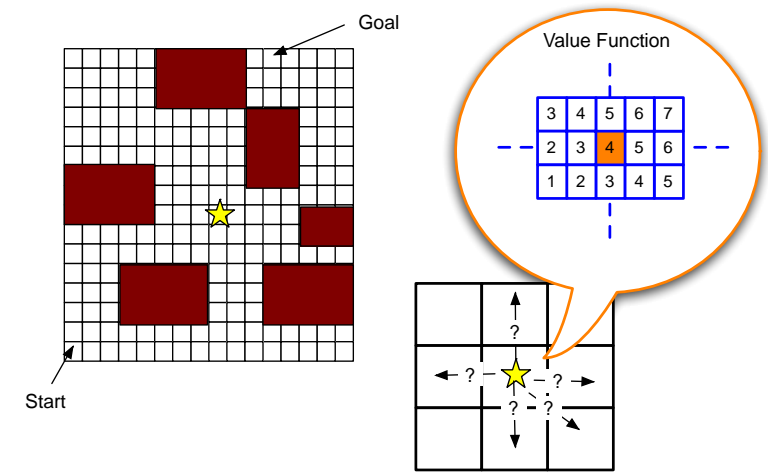
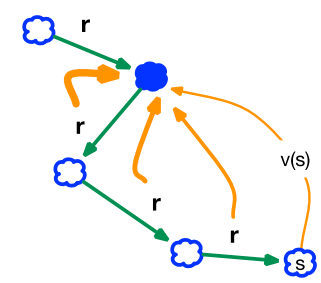
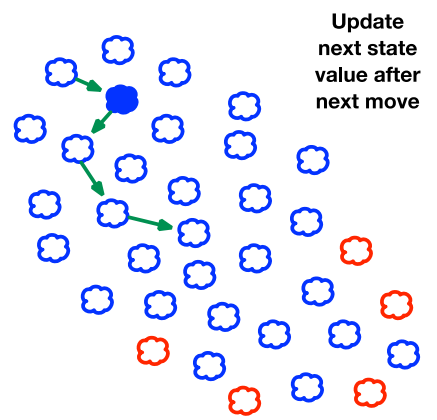
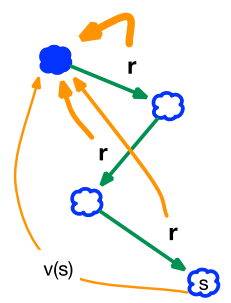
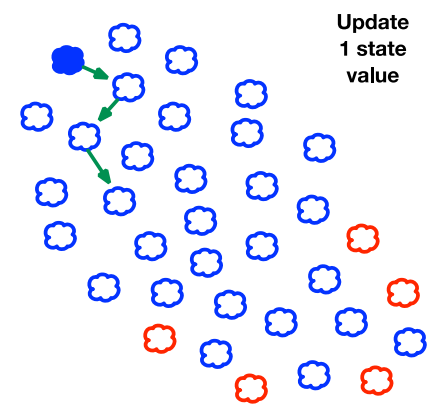
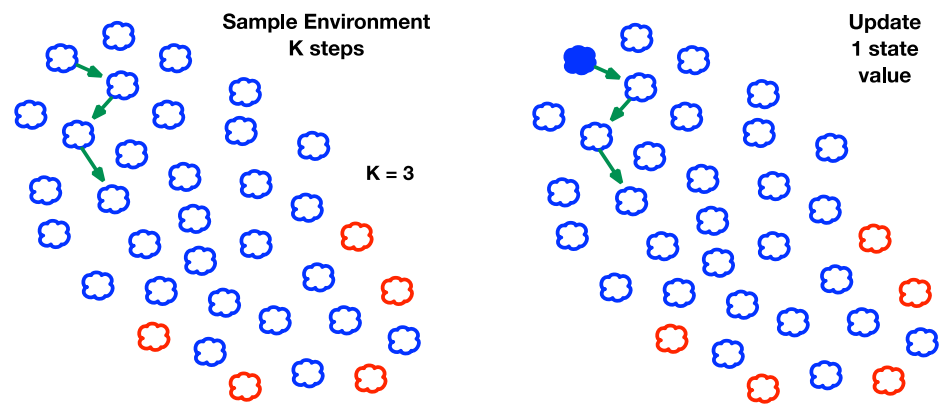
- This only considers **SOME** moves from any state, based on the **strategy**.
- **Learning:** The strategy gradually changes / improves (based on the results of many rounds of MCTS)
- Still, all this work for **ONE** move !



Trial and Error Learning:

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

Reinforcement Learning (RL)



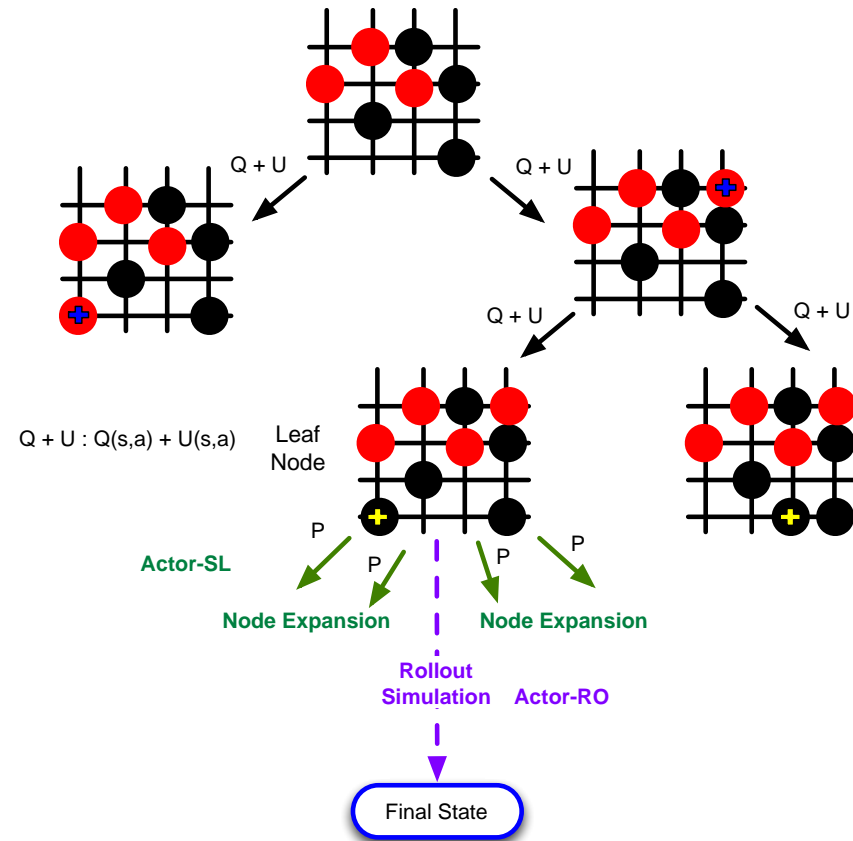
Bootstrapping: The **predicted** value of a future state is used to update the **predicted** value of the current state.

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!**

AlphaGo

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



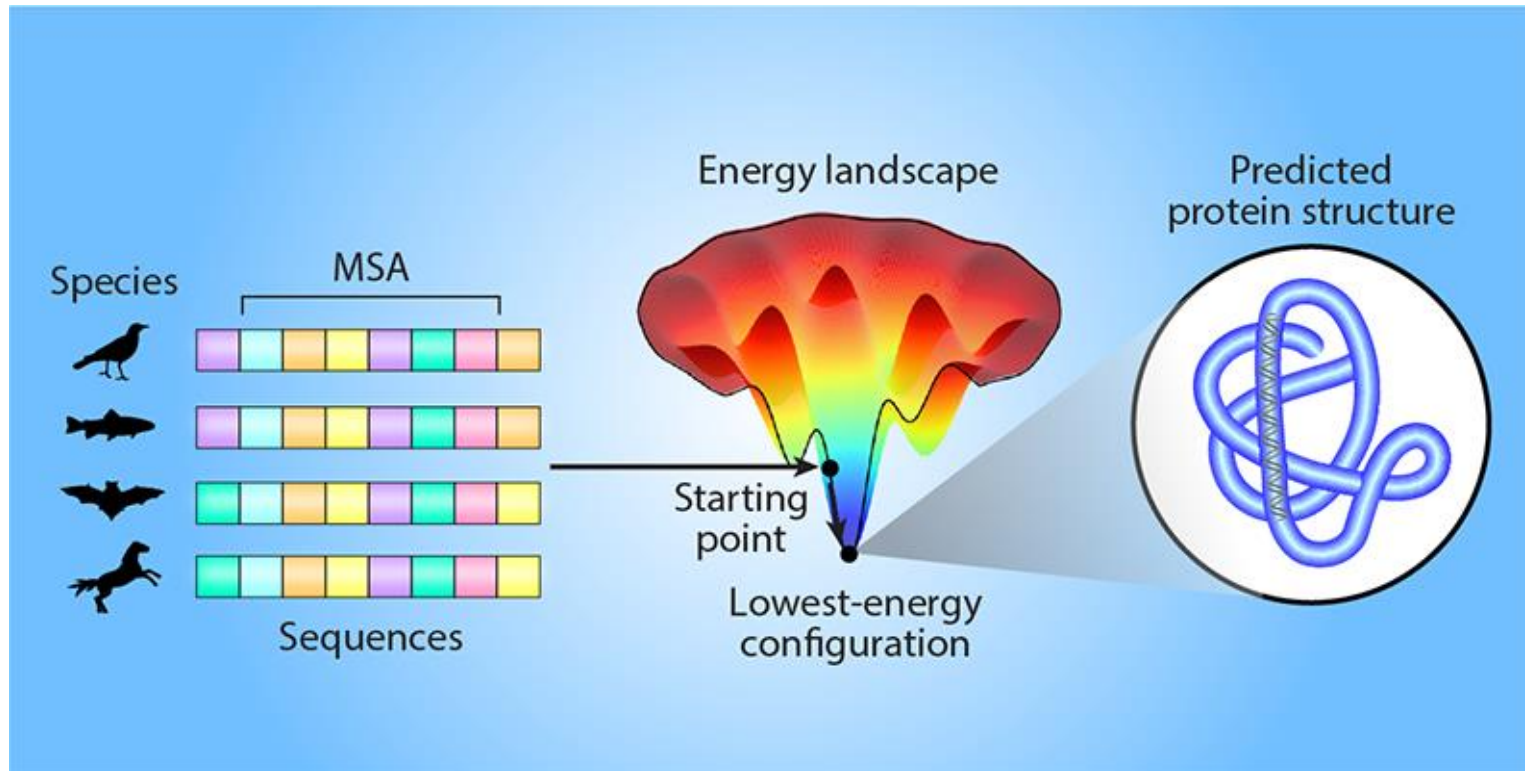
I guess I lost the game because I wasn't able to find any weaknesses...Lee Sedol (World # 2)

AlphaFold (DeepMind, 2020)

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

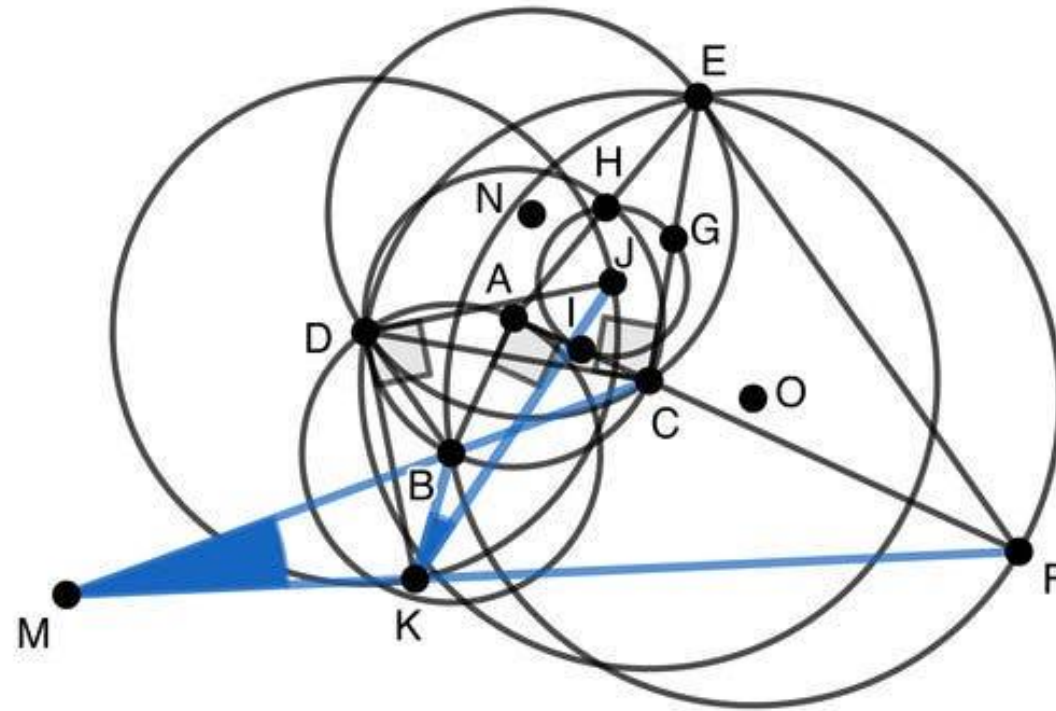
AlphaFold Computes: 3d structure of the resulting protein = How the amino-acid sequence “folds”.

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** !!
- Shared 2024 Nobel Prize in Chemistry for this work !!

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

Human Expertise

- Acquired by Extensive Experience
 - 10,000+ hours
 - Solving lots of problems
 - Playing lots of games (from start to finish)
- Exemplified by:
 - Good heuristics
 - Good strategies
 - Superior pattern-recognition abilities
 - NOT necessarily any better at deep and/or broad search than a novice
 - Explainability:
 - Experts can explain **some** of their reasoning
 - But some things have become so automatic / unconscious that they can't

Current AI Expertise

- Acquired by Extensive Computation
 - **Trial-and-error Learning:** Millions of games played, problems solved, etc. **on its own**
 - **Supervised Learning:** Millions of human cases processed, abstracted, “memorized”
- Exemplified by:
 - Excellent (often **creative**) strategies
 - Superior (to human) pattern-recognition abilities
 - Explainability:
 - Very poor for modern ML systems. Can't say **why** they make particular decisions.
 - Much better for the old classic logic-based systems.

Two Main Modes of Thought

System #1:

Fast, Frequent, Automatic, Emotional, Reactive
Unconscious

Examples:

Multiply 7 x 8

Read a STOP sign

Recognize Mother's picture

React to a flying stone

Write a common word

Sing a familiar song

System #2:

Slow, Infrequent, Effortful, Deliberate, Logical,
Conscious

Examples:

Multiply 23 x 36

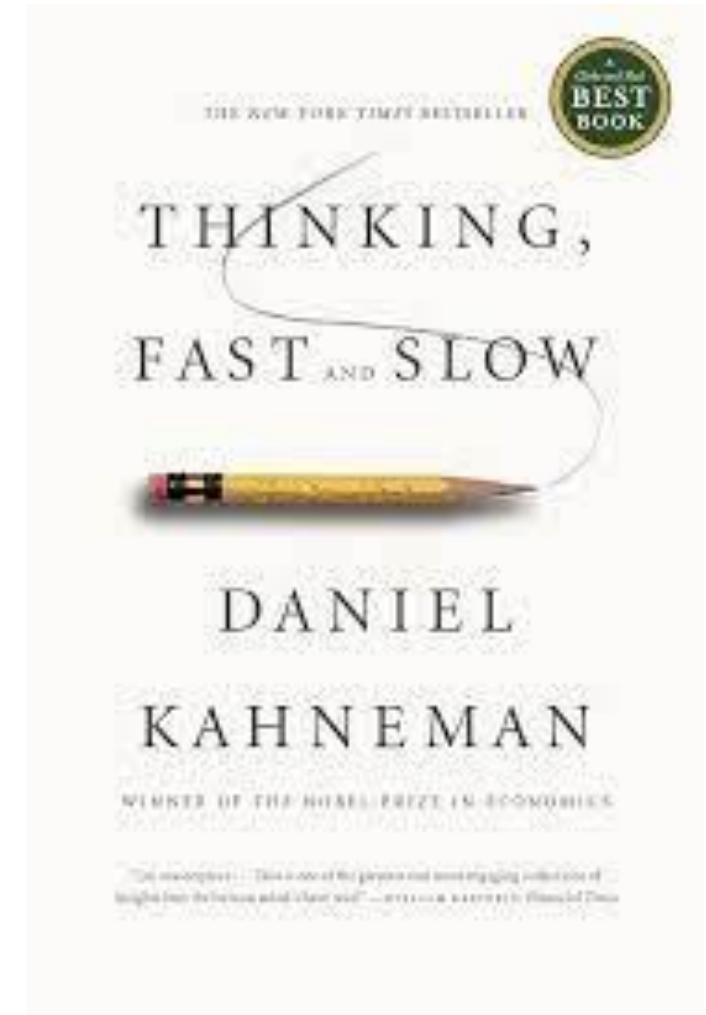
Read a poem

Plan a vacation

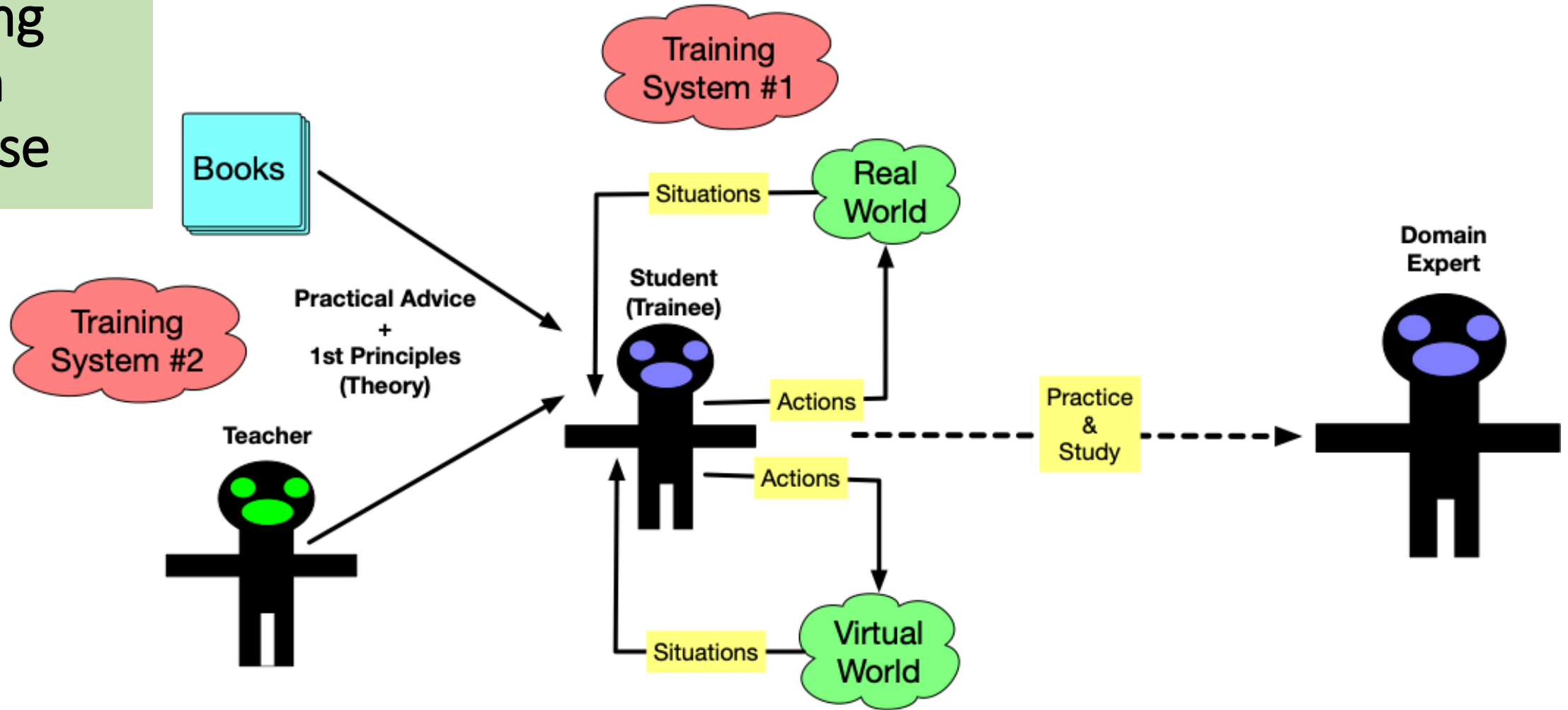
Design a kitchen

Make a legal argument

Convince a kid to take out the trash

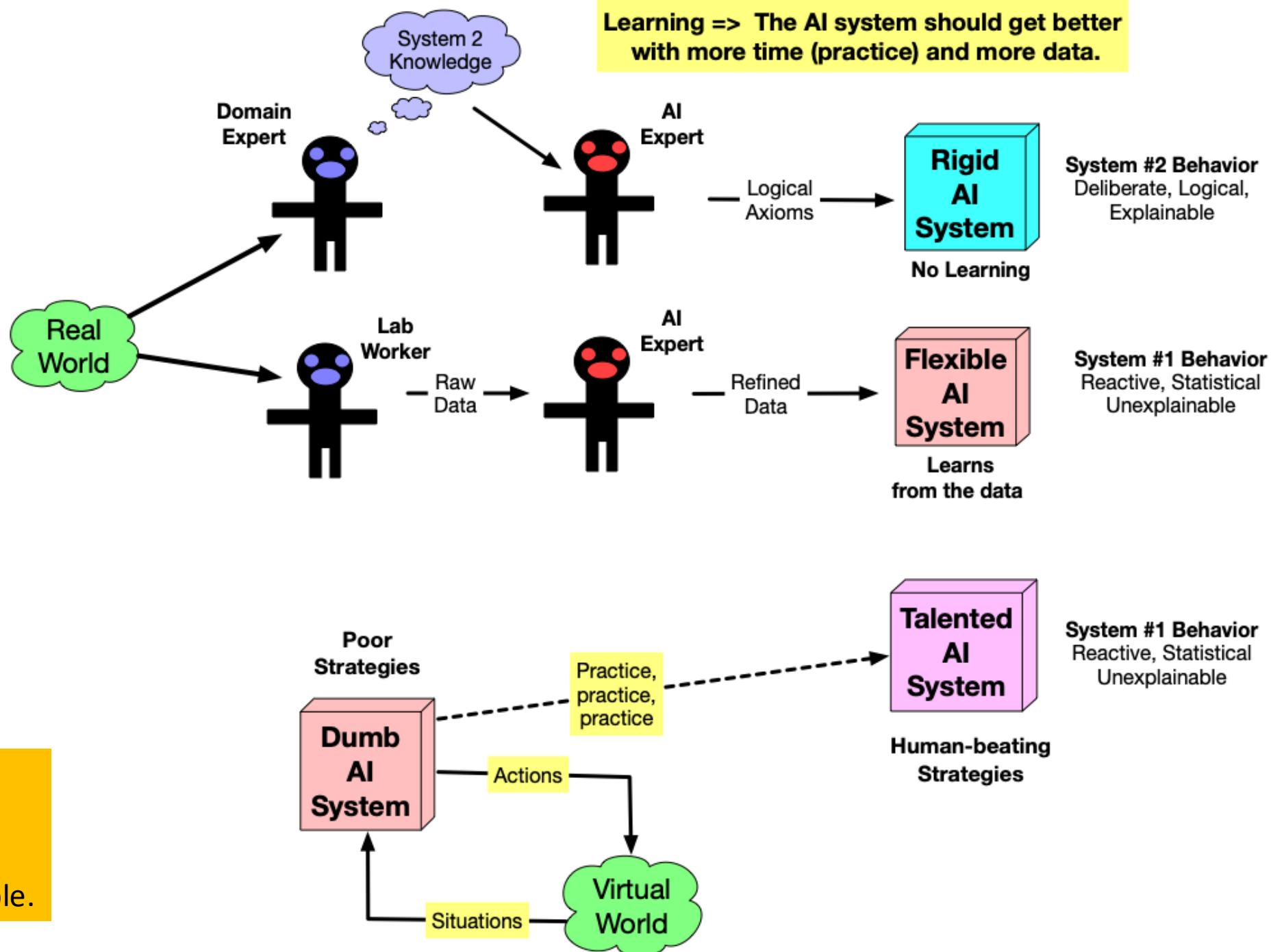


Attaining Human Expertise



- Human Experts have a combination of System #1 and System #2 competence.
- Some abilities begin as System #2 but become “routine” System #1 skills => Experts can have trouble giving system-2 explanations for **some** things (but not all).

Attaining AI Expertise



Today's most successful AI systems are almost exclusively System #1 => Smart, but Unexplainable.

Thinking Fast & Sloppy



Prompt: A 6-fingered man counting to 12

AI Slop

- Without 1st principle understanding, AI will generate anything whose patterns have statistical correlation with those in the data cases it has seen.
- And AI systems are using AI-generated text and images as training data !!



A clock for an Ohio farmer rotated 45 degrees



A clock for a Brazilian astronaut using Roman numerals



Still a few bugs to work out

A clock for a Hawaiian polar explorer using binary numbers.



A clock in the shape of a hexagon for an Egyptian archaeologist



...and a few more bugs...

ME: Draw a calendar for July that a Greek sailor would like.



ChatGPT: Here's the July calendar designed for a Greek sailor, featuring nautical elements and a Mediterranean vibe.

ME: Edit the previous calendar and make sure the days are ordered correctly.



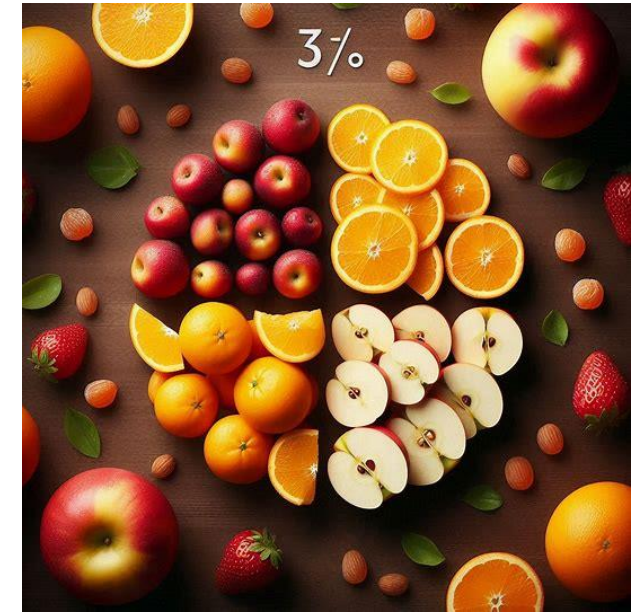
ChatGPT: Here's the revised version of the July calendar, now with the days correctly ordered. It retains the Greek maritime elements and Mediterranean atmosphere.

A lot of people struggle with math. We're not all computers, after all!!

Many AI systems:

- very intelligent
- very creative
- Interact well with humans

But their **understanding** of things is different, and non-transparent to humans => Hard to **trust** AI.





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

The Alpha Progression



- No expert game data 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)

Behavior Rules

Case Features

Case Labels

Constraints

System Goals

Computationally Intensive Reasoning

GOFAI
(1970-1990)



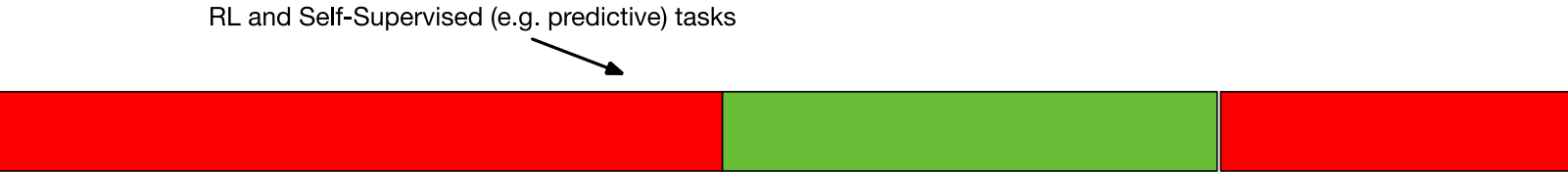
Early Useful ML
(1990's)



Impressive ML
(2000 - ?)

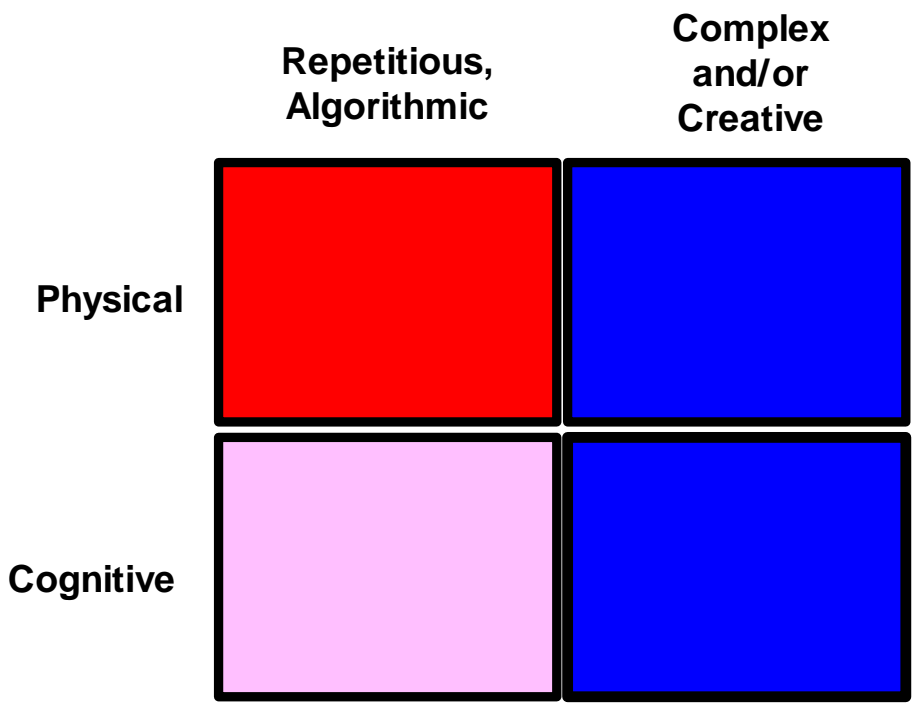
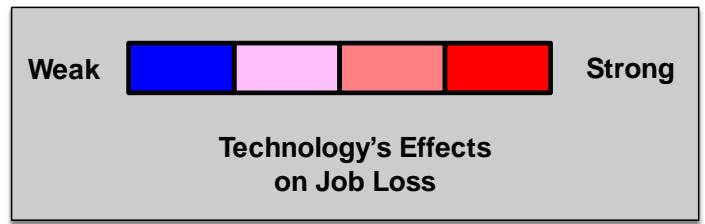


Generative AI / ML
(2020 - ?)

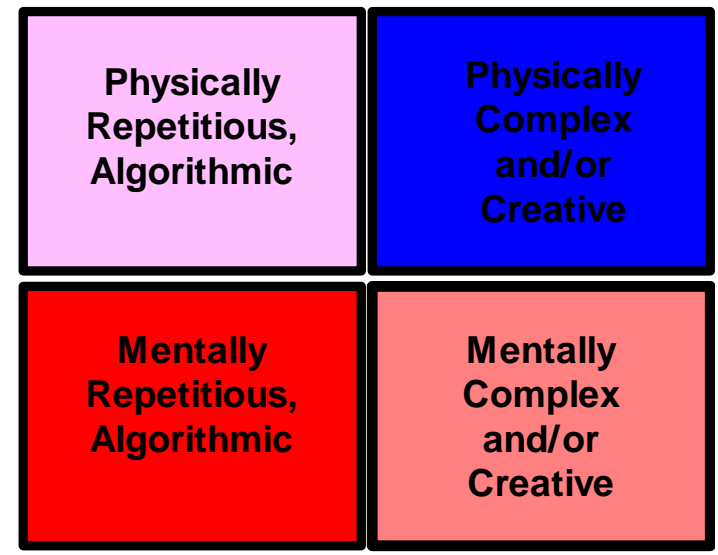


Feared Scenarios





Classic Technological Disruption



AI's Disruption



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

Plight of the Infovore

*... 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**)*

Constant distractions of cyberspace

“Shallows”: Reduced depth of thought and emotion

Less Creative

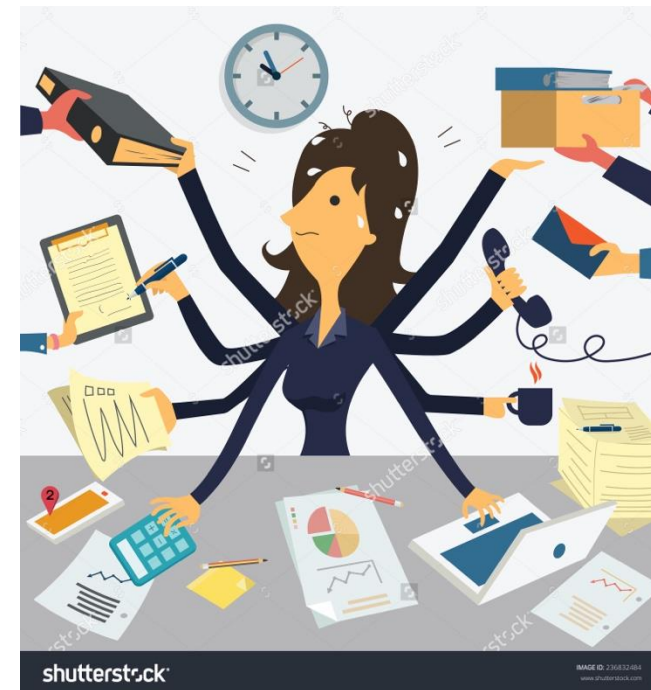
Reduced value in an information society

Automation Bias: Trust machines over ourselves.

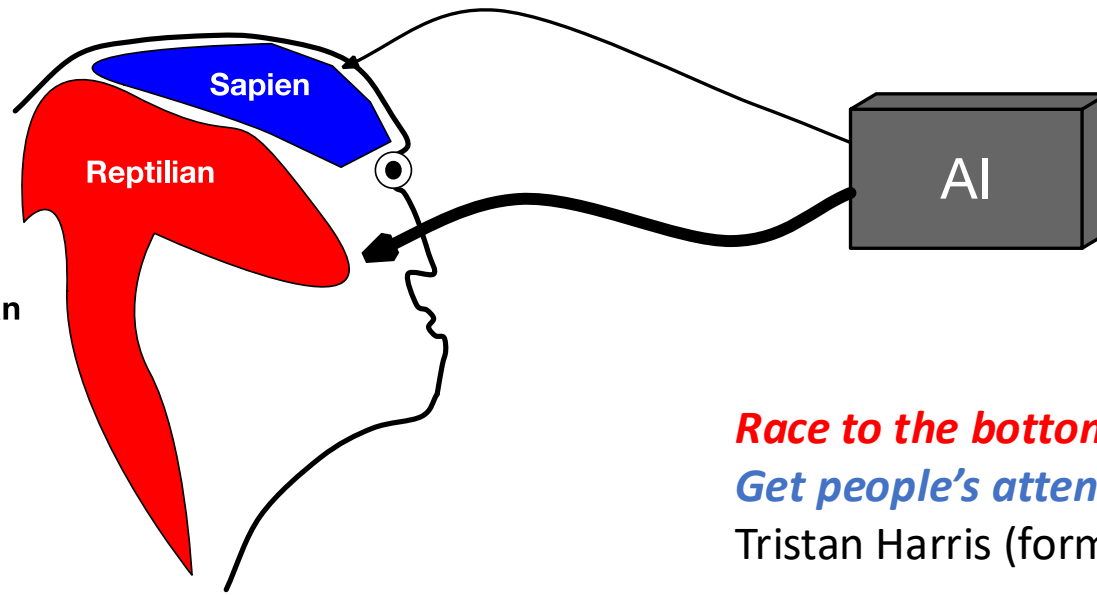
Rely on computers to understand the world

Rely on AI for **wisdom**

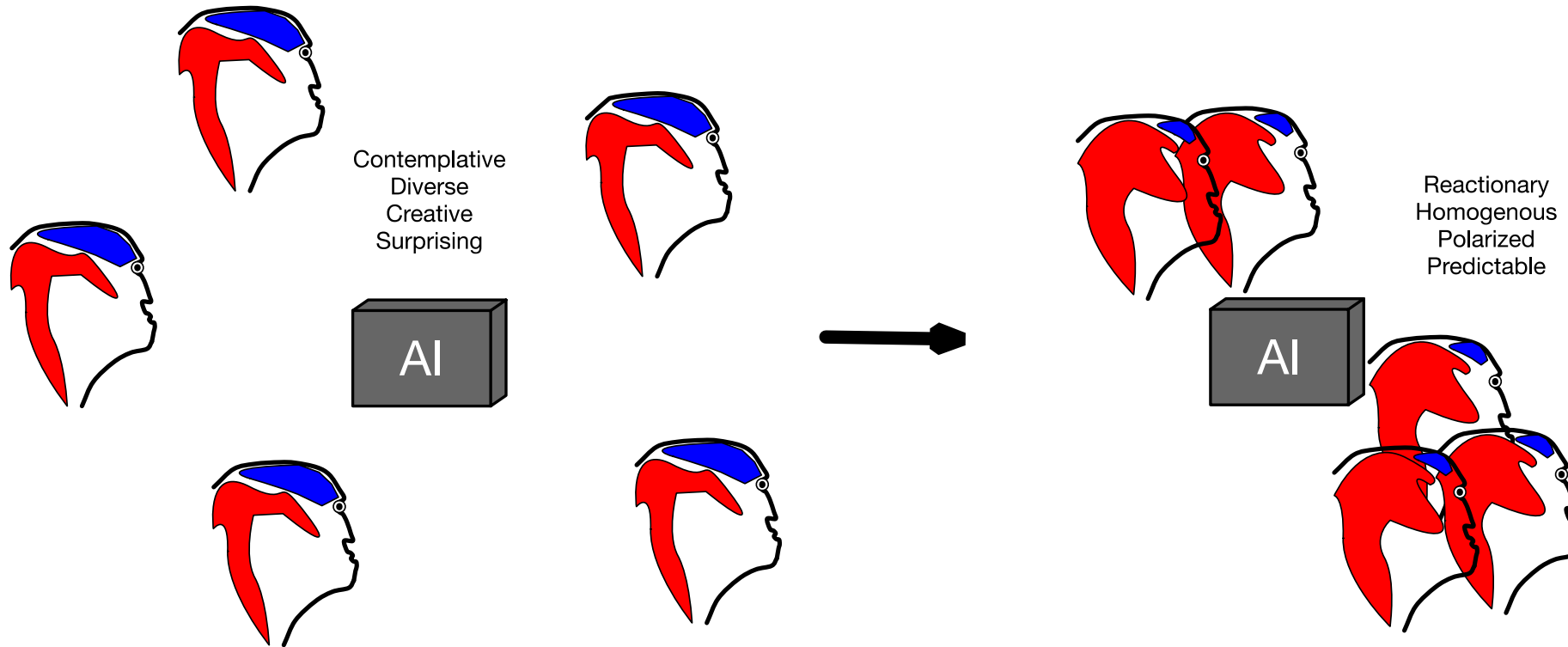
Human intelligence becomes **artificial** ... and thus more easily **predicted** and **controlled** by AI



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)



*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)

My Evolving Trauma

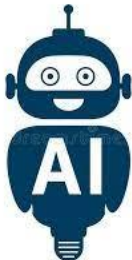
1975 - 1983

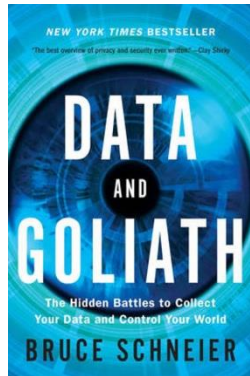
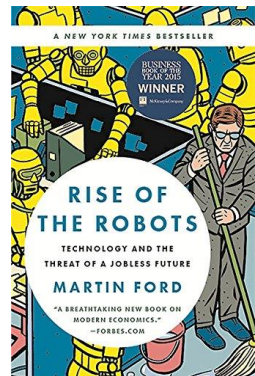
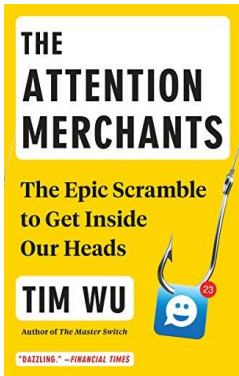
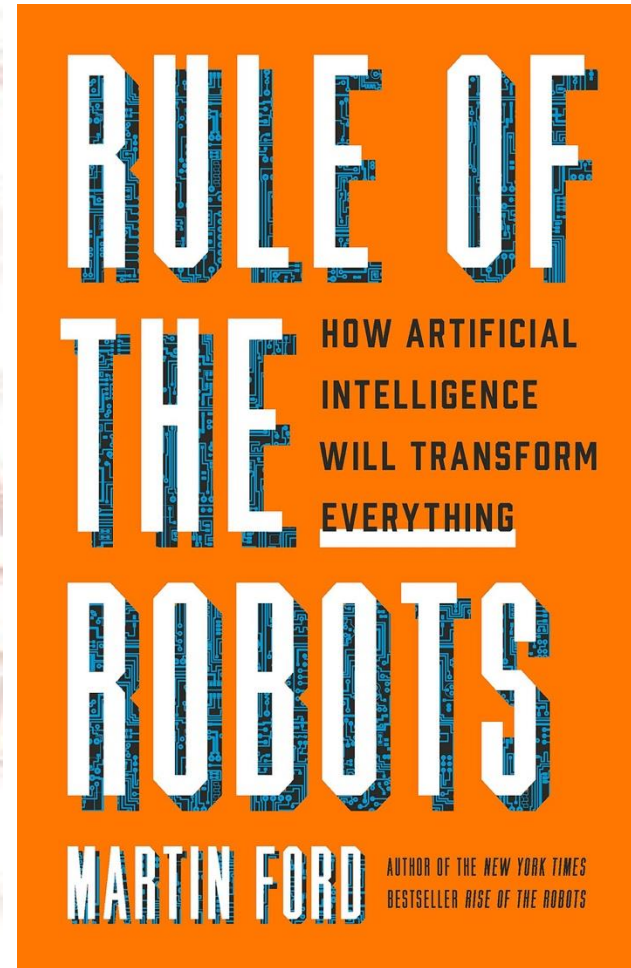
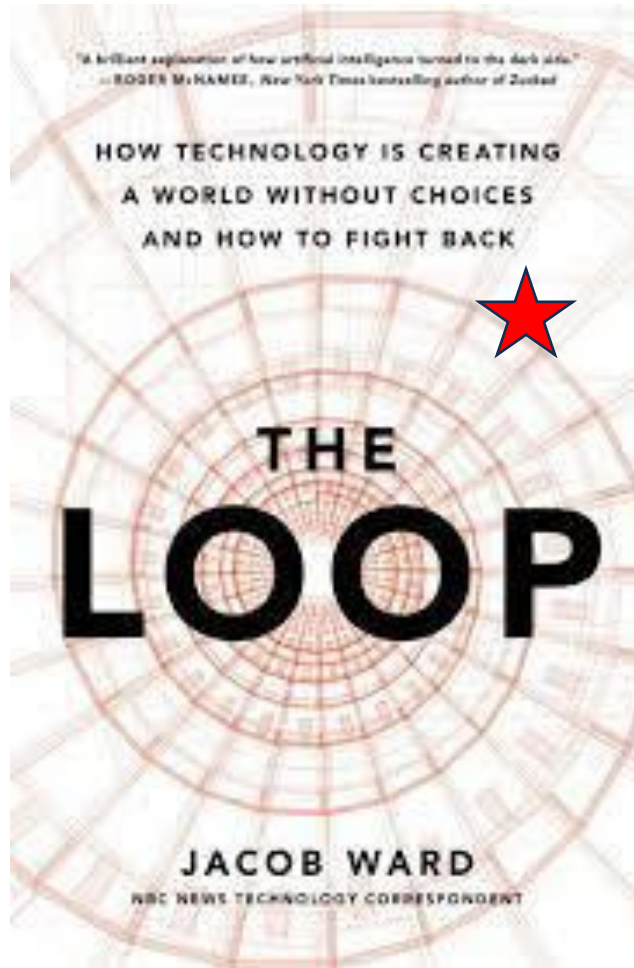
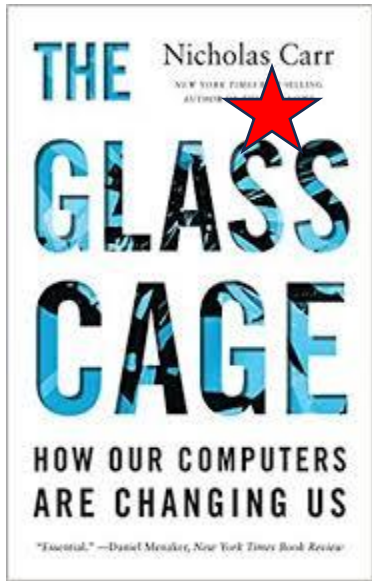
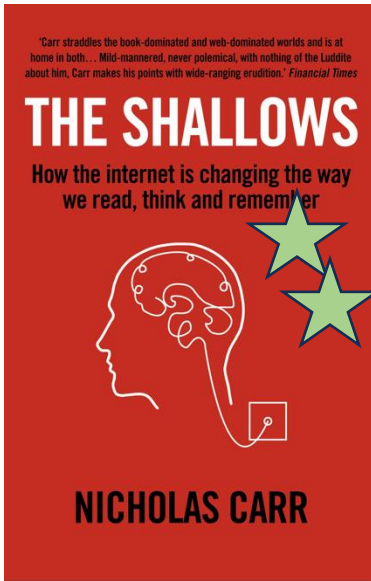


1984 - 2021



2022 - present





*Layout suggested by AI

Artificial Intelligence

A Guide for Thinking Humans

Melanie Mitchell



THE AGE OF SURVEILLANCE CAPITALISM

THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER

SHOSHANA ZUBOFF



HARNESSING OUR DIGITAL FUTURE

MACHINE PLATFORM CROWD

ANDREW McAfee
ERIK BRYNJOLFSSON

NEW YORK TIMES BESTSELLER
WITH 4 NEW CHAPTERS
WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY
CATHY O'NEIL

BrillianceAudio

Unabridged

NICK BOSTROM

SUPERINTELLIGENCE

Paths, Dangers, Strategies



OUR FINAL INVENTION

JAMES BARRAT

THE MASTER ALGORITHM

HOW THE QUEST FOR THE ULTIMATE LEARNING MACHINE WILL REMAKE OUR WORLD

PEDRO DOMINGOS

Performed by Mel Foster

ONLY FROM audible

ONLY FROM audible

