Artificial Intelligence –
Are We There Yet?

Lou Mazzucchelli
(with a massive assist from AvD and the CSCI 0015 STA team)

CSCI 0020 – October 2024

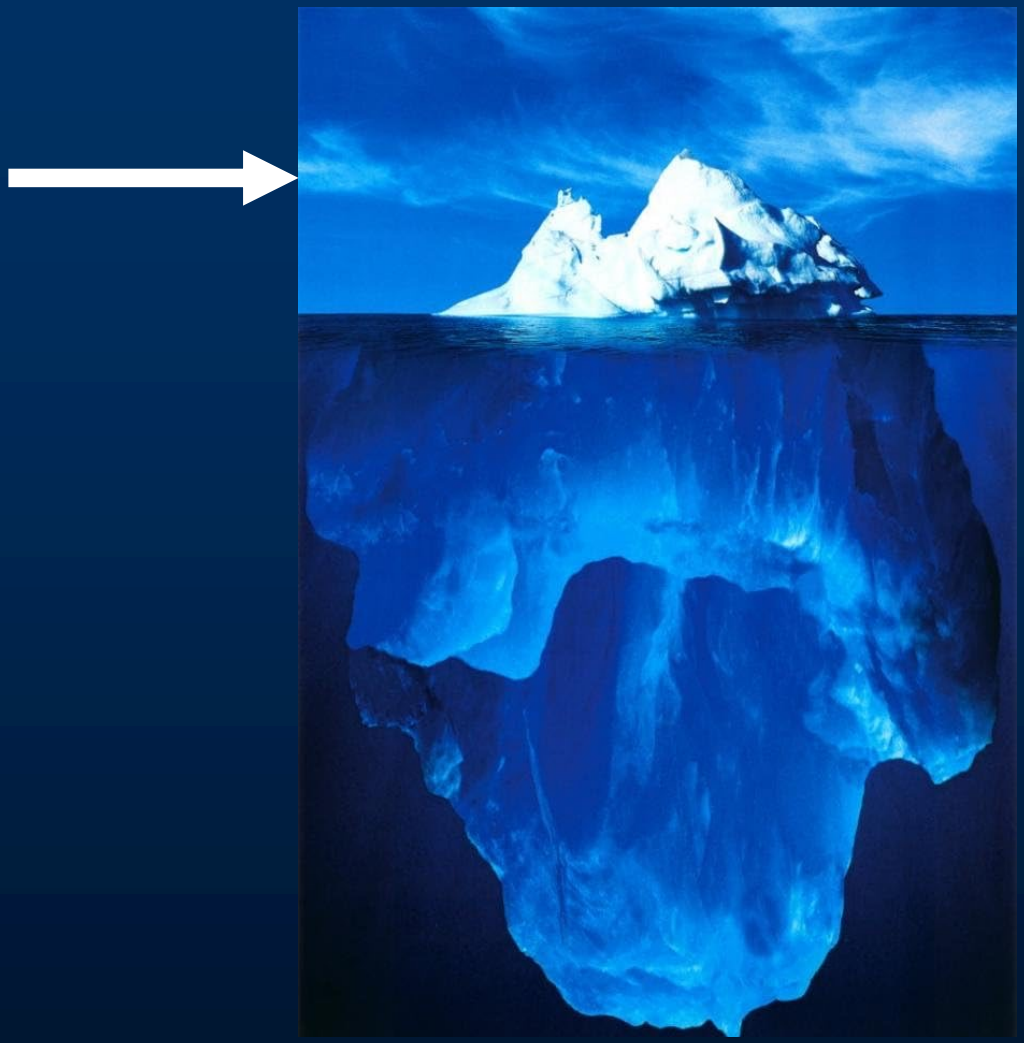
What is it?

Artificial intelligence (AI) is technology that enables computers and machines to *simulate* human learning, comprehension, problem solving, decision making, creativity and autonomy.

- IBM

- https://www.ibm.com/topics/artificial-intelligence

This Presentation

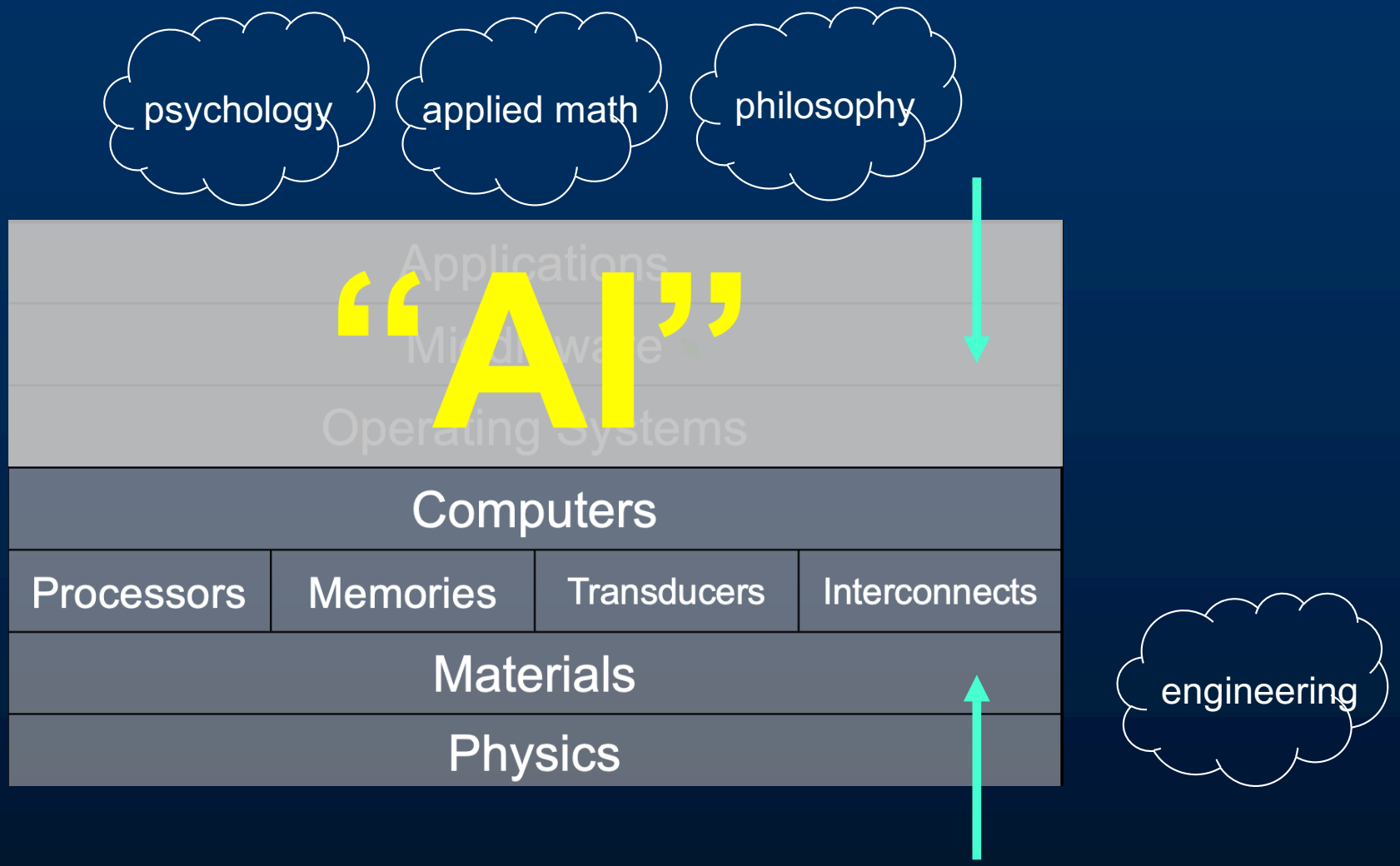


Origins - 1956

"We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

- John McCarthy

AI - then


psychology
applied math
philosophy
Applications
Middleware
Operating Systems
"AI"
Computers
Processors
Memories
Transducers
Interconnects
Materials
Physics
engineering

An inconvenient truth?

"I think, therefore I am."

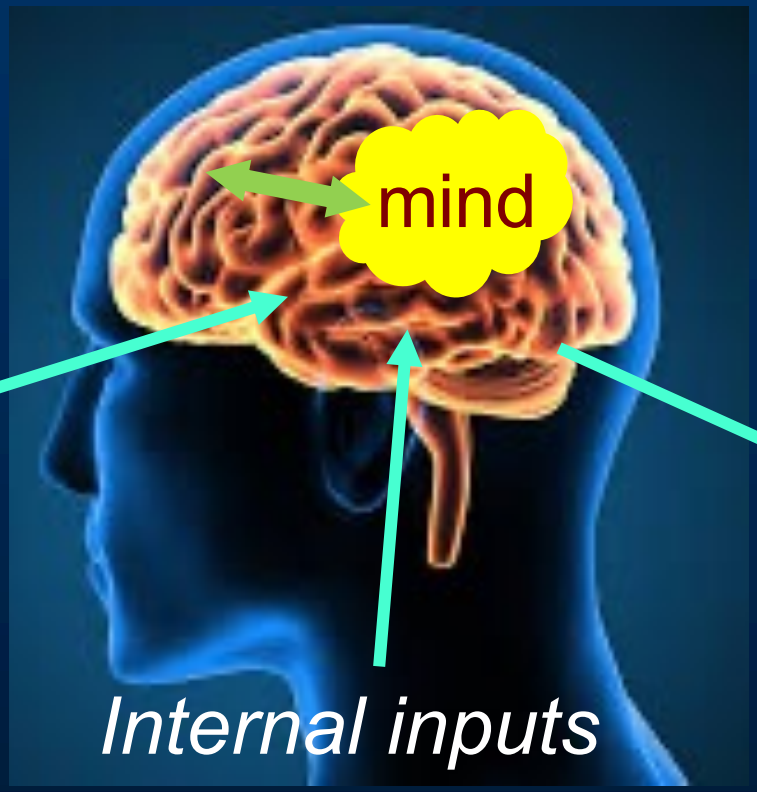
The mind-body issue

- You have a body

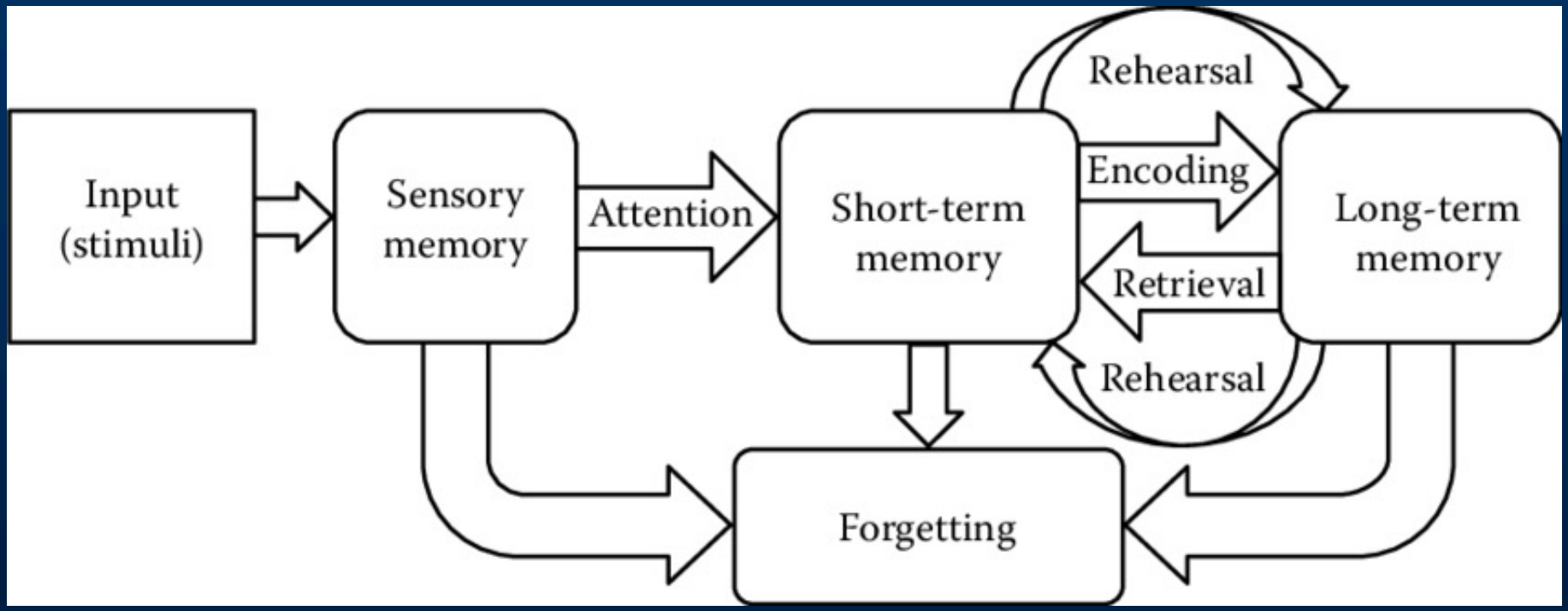
- You have a mind

How do you know?
How are body and mind related?

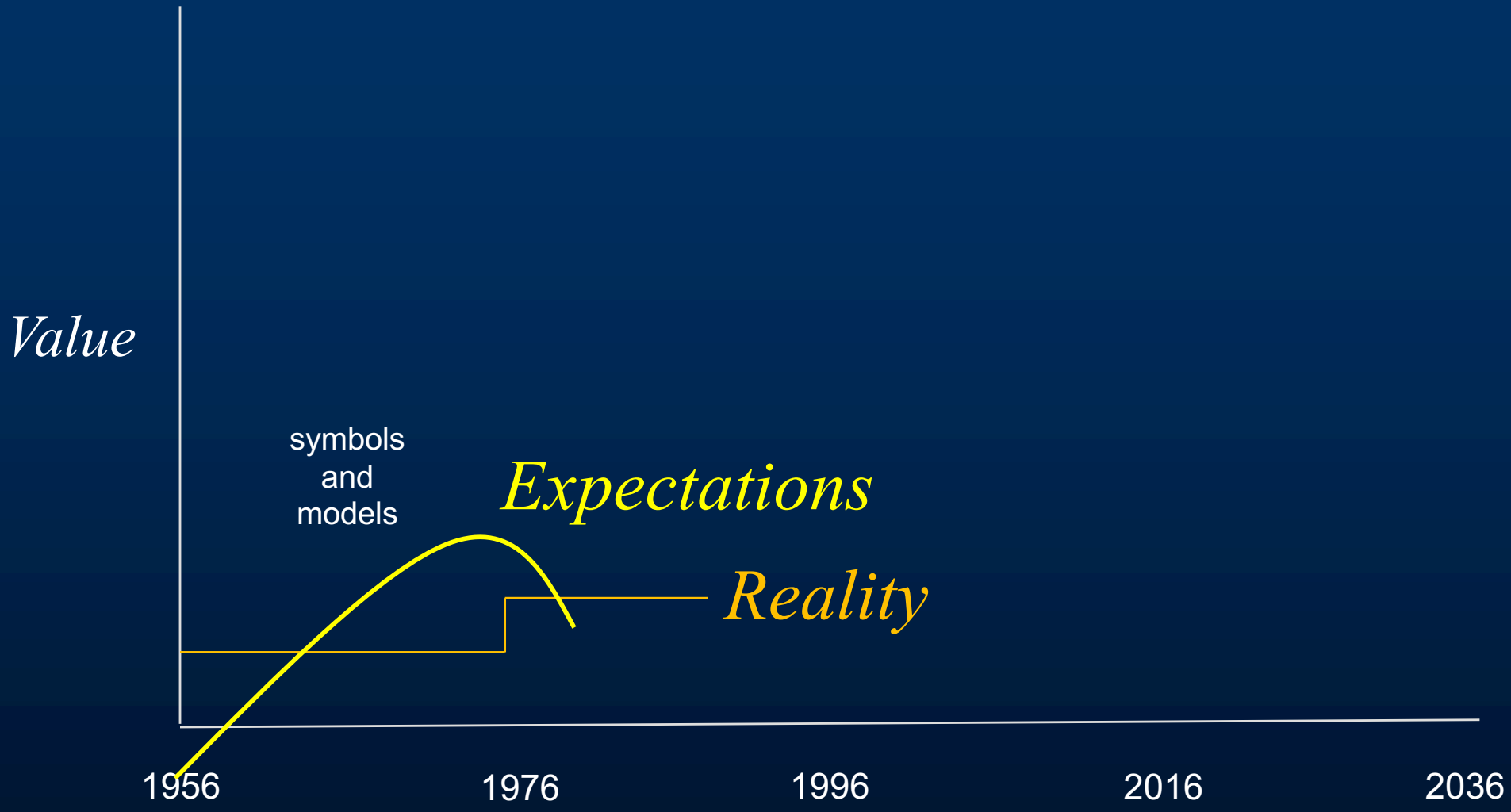
Living in the physical / mental world



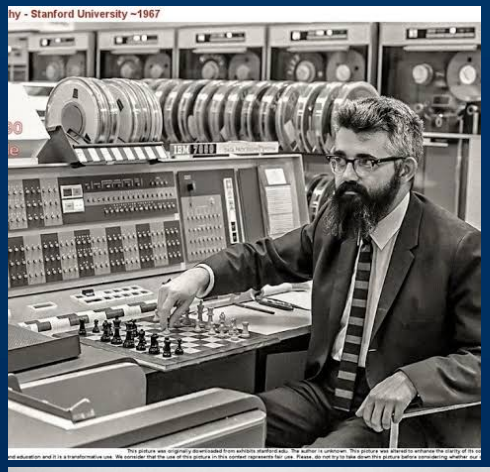
Human Memory Model



A Brief Graphic History of AI



John McCarthy (LISP), 1958



Unimate, 1961

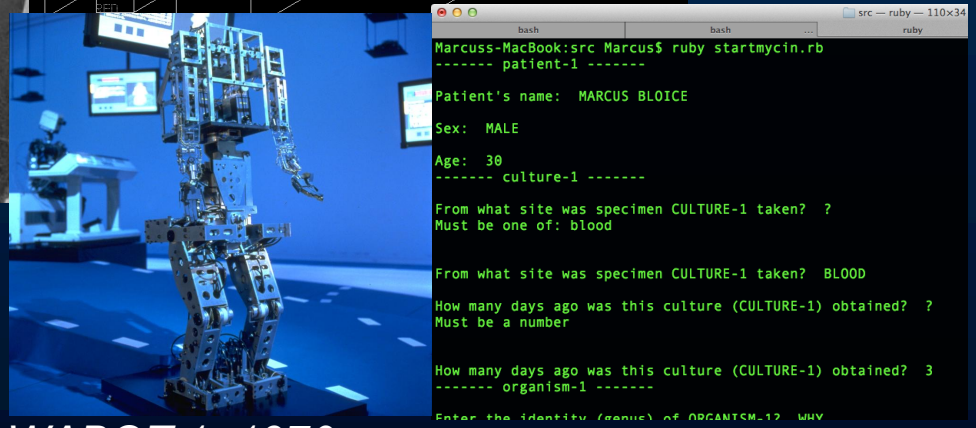
Joseph Weisenbaum, 1965

Feigenbaum et al, DENDRAL, 1965

Shakey, 1966

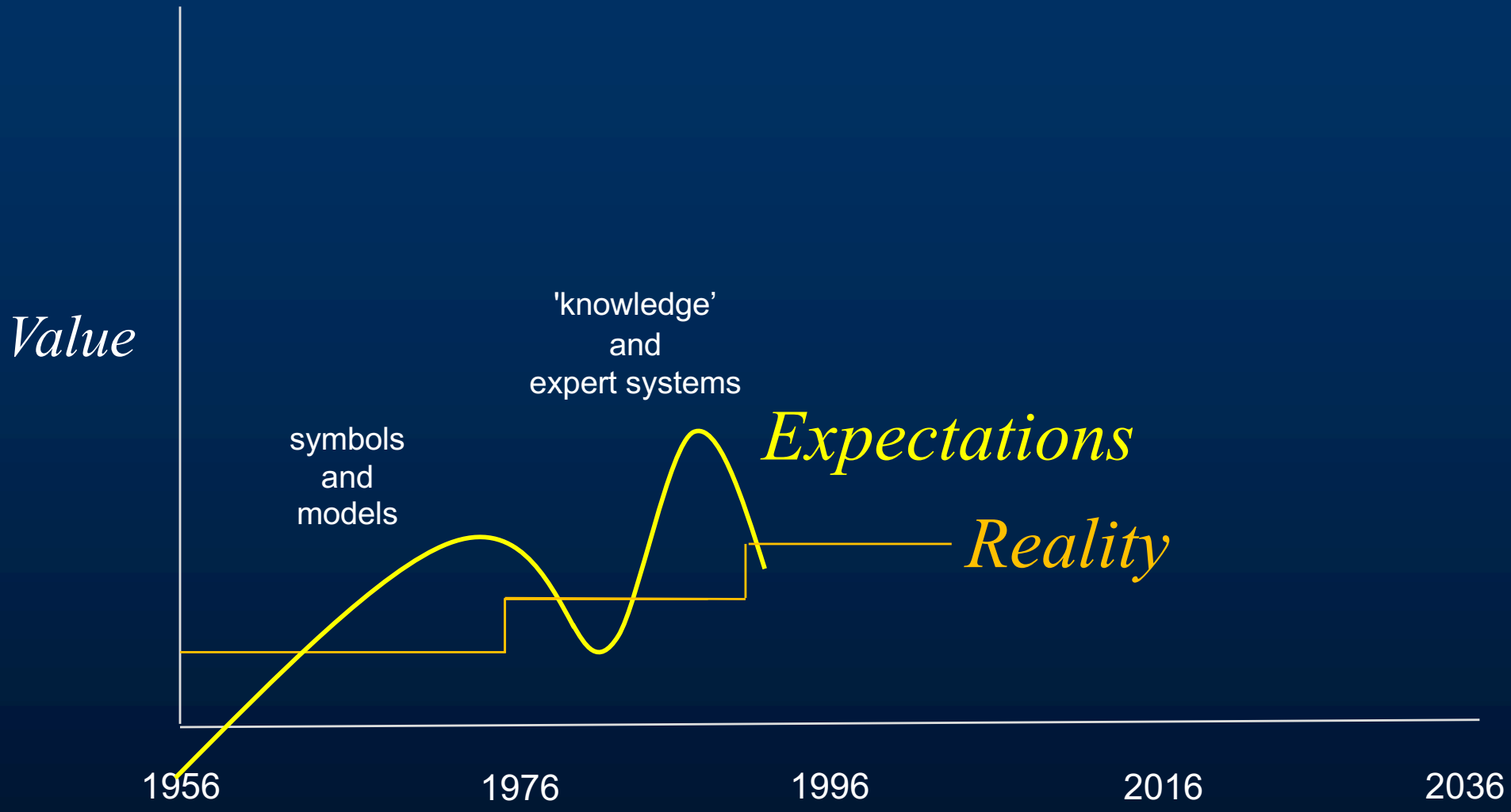
Wingrad, SHRDLU, 1968

MYCIN, 1972

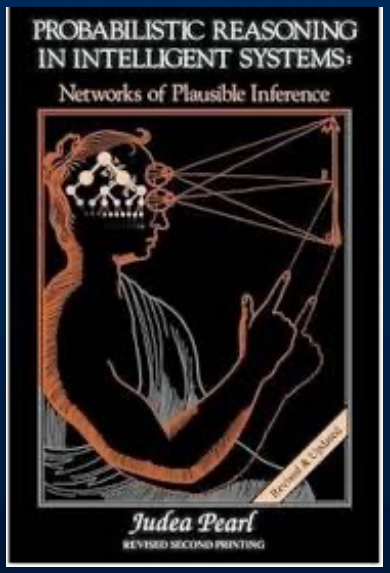


WABOT-1, 1970

A Brief Graphic History of AI

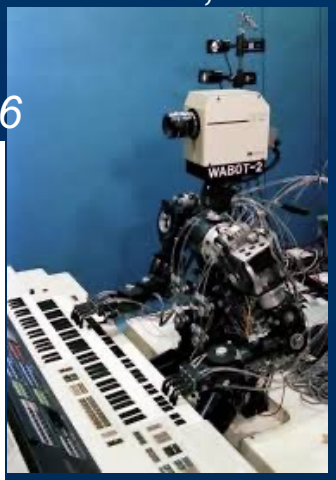


LMI / Symbolics, 1979-1996

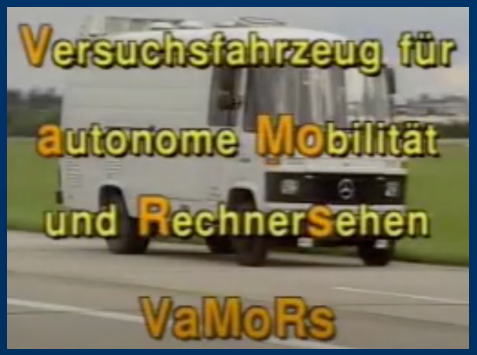


Pearl, 1986

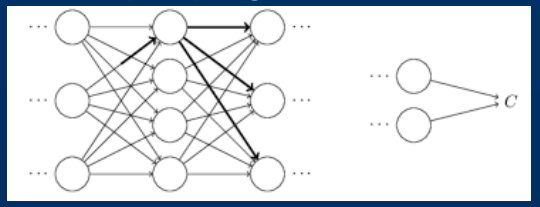
WABOT-2, 1980



Dickmanns, 1986



Rumelhart et al, Back-propagation, 1986



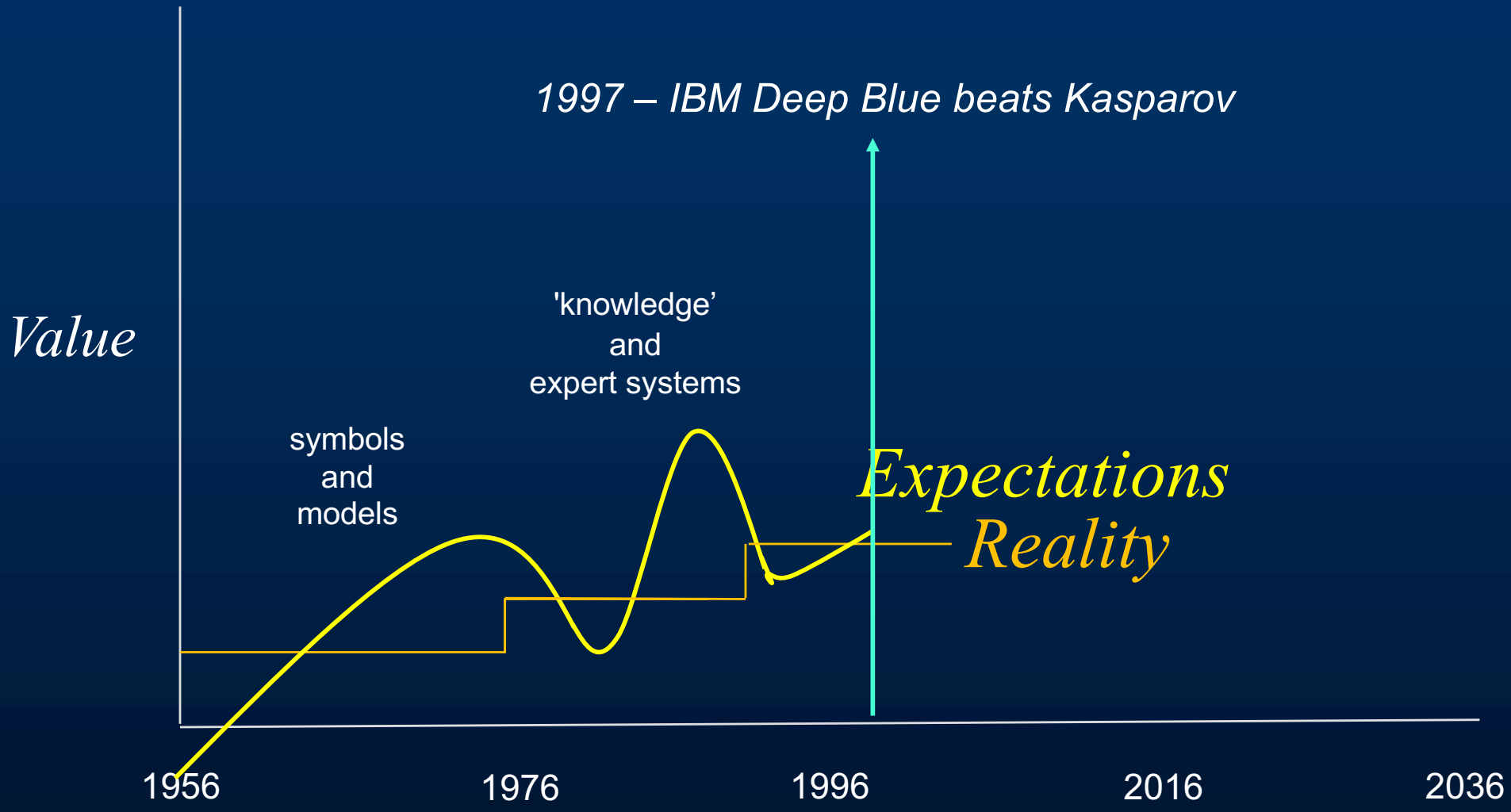
Vinge, Singularity, 1993

IBM Deep Blue, 1997

Brooks, “Elephants Don’t Play Chess”, 1990

Charniak, “Statistical Techniques for Natural Language Parsing”, 1990

A Brief Graphic History of AI

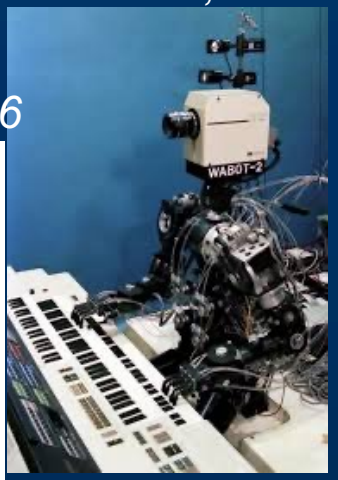


Deep Blue – the last of its kind?

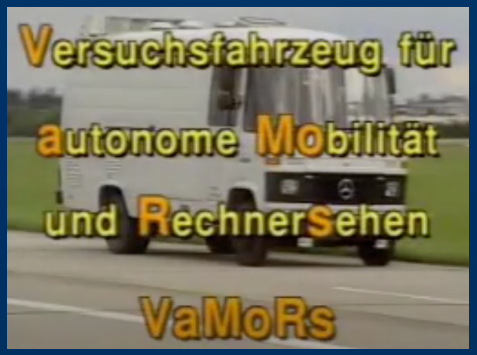
- *Evaluation function designed for chess*
- *Evaluation function designed for chess*
- *30 CPUs*
- *480 custom 'chess chips' to speed search*
- *18.38 Gflops*
- *200M chess positions/sec*

LMI / Symbolics, 1979-1996

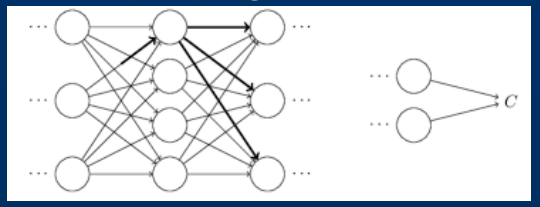
WABOT-2, 1980



Dickmanns, 1986



Rumelhart et al, Back-propagation, 1986

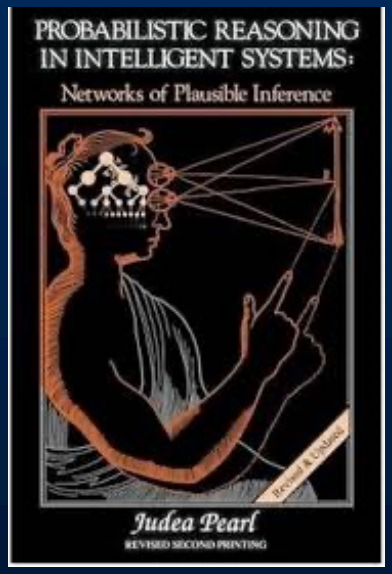


Vinge, Singularity, 1993

IBM Deep Blue, 1997

Brooks, “Elephants Don’t Play Chess”, 1990

Charniak, “Statistical Techniques for Natural Language Parsing”, 1990

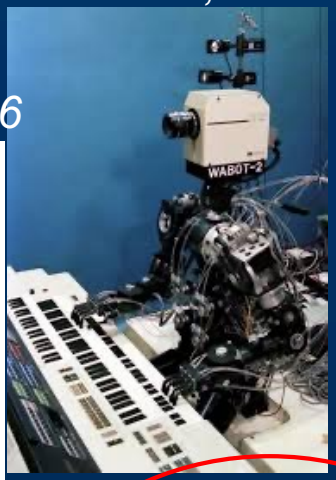


Pearl, 1986

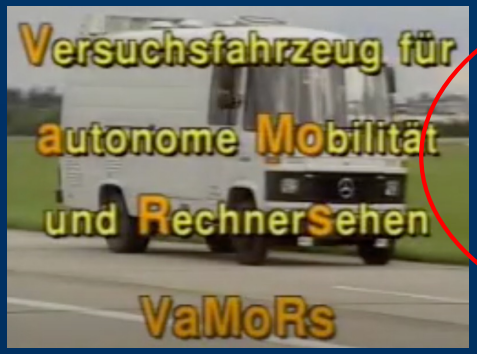
Hampton & Chung, 1998

LMI / Symbolics, 1979-1996

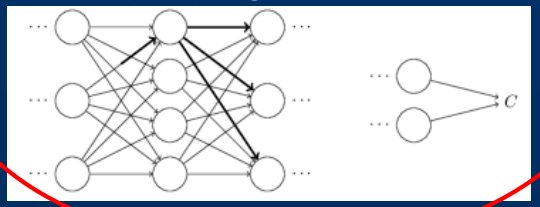
WABOT-2, 1980



Dickmanns, 1986

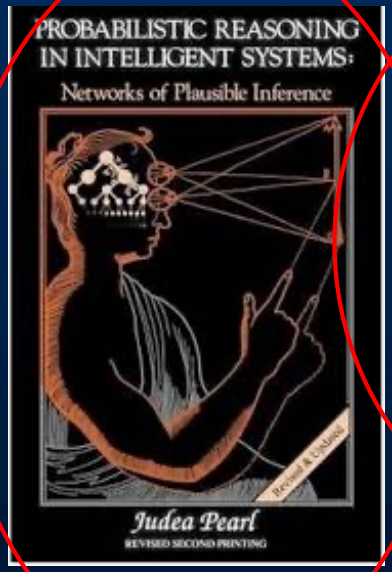


Rumelhart et al, Back-propagation, 1986



Vinge, Singularity, 1993

IBM Deep Blue, 1997



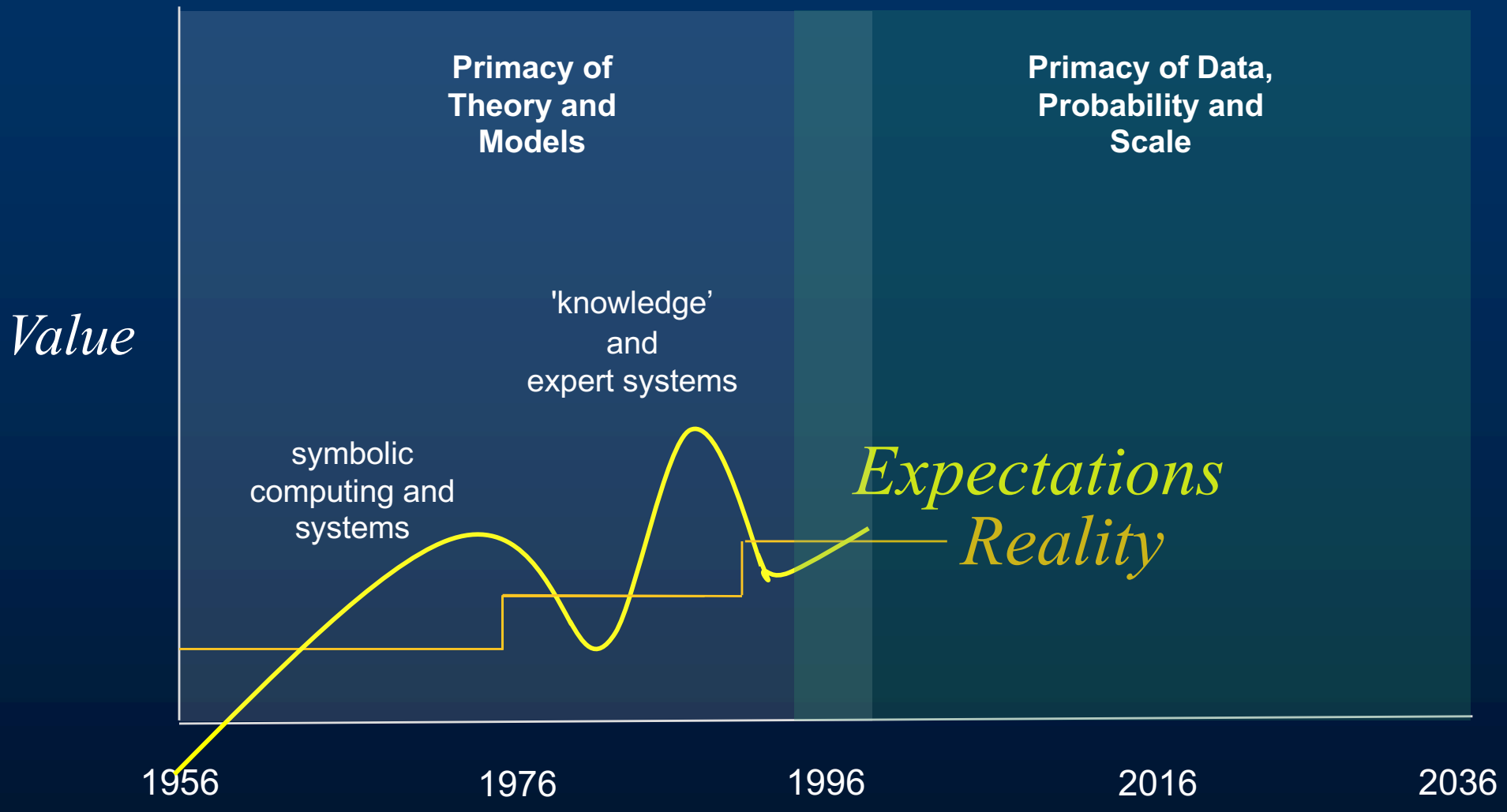
Pearl, 1986

Brooks, "Elephants Don't Play Chess", 1990

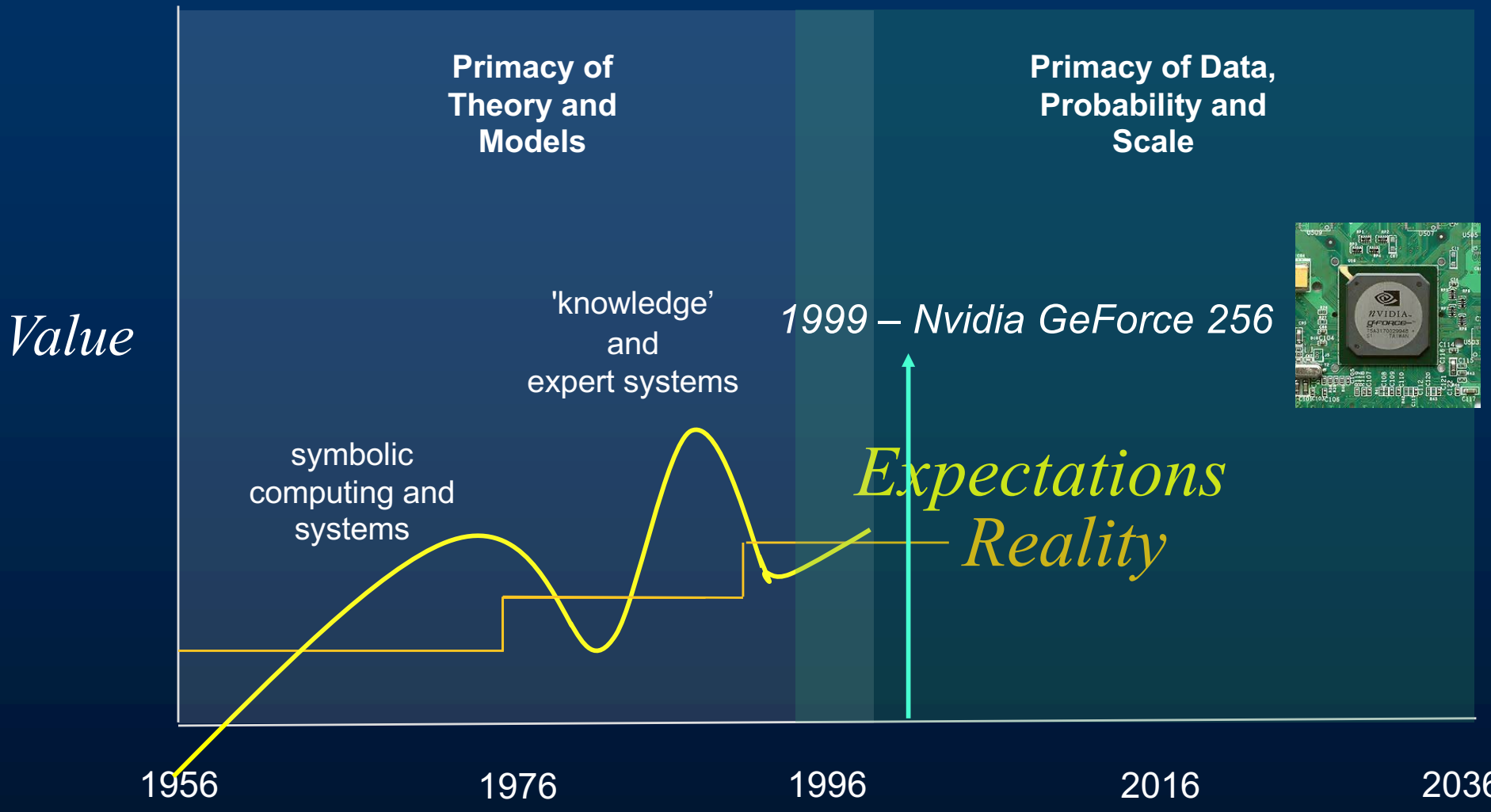
Charniak, "Statistical Techniques for Natural Language Parsing", 1990

Hampton & Chung, 1998

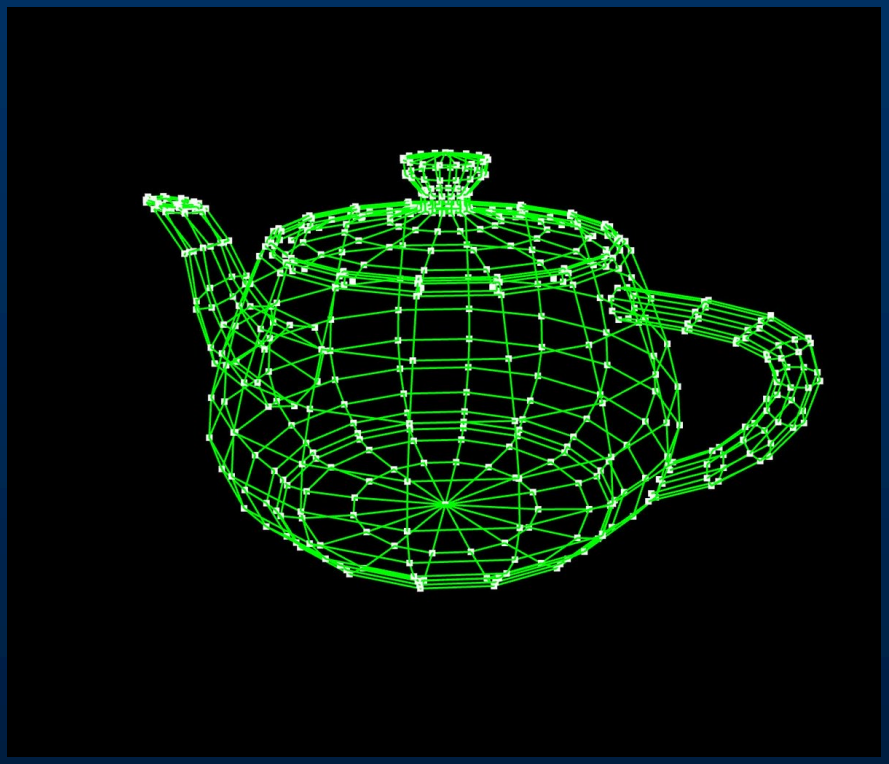
A Brief Graphic History of AI



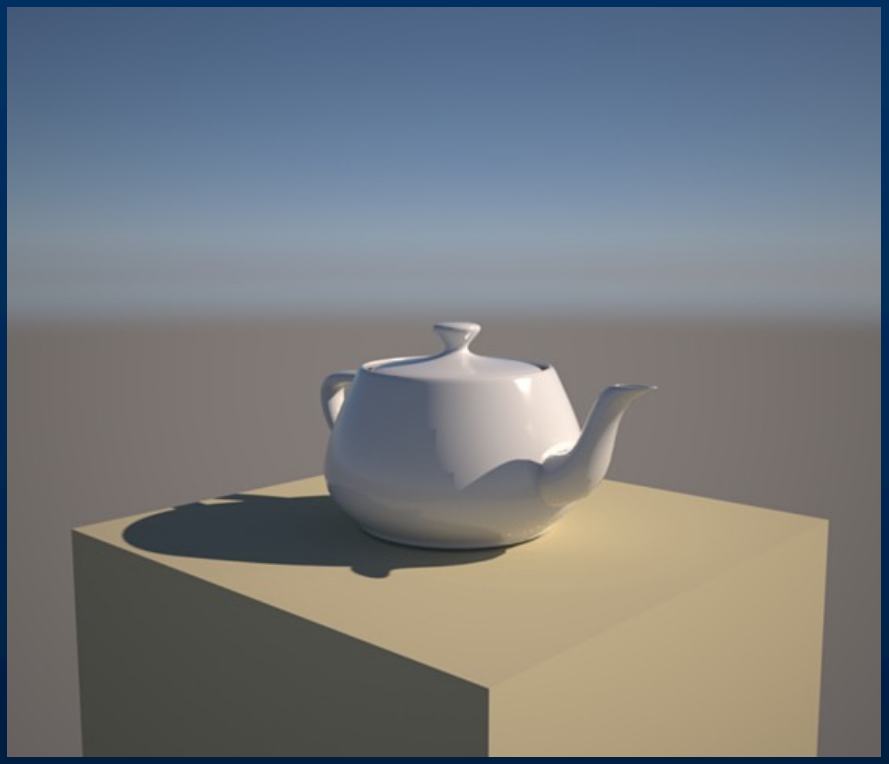
A Brief Graphic History of AI



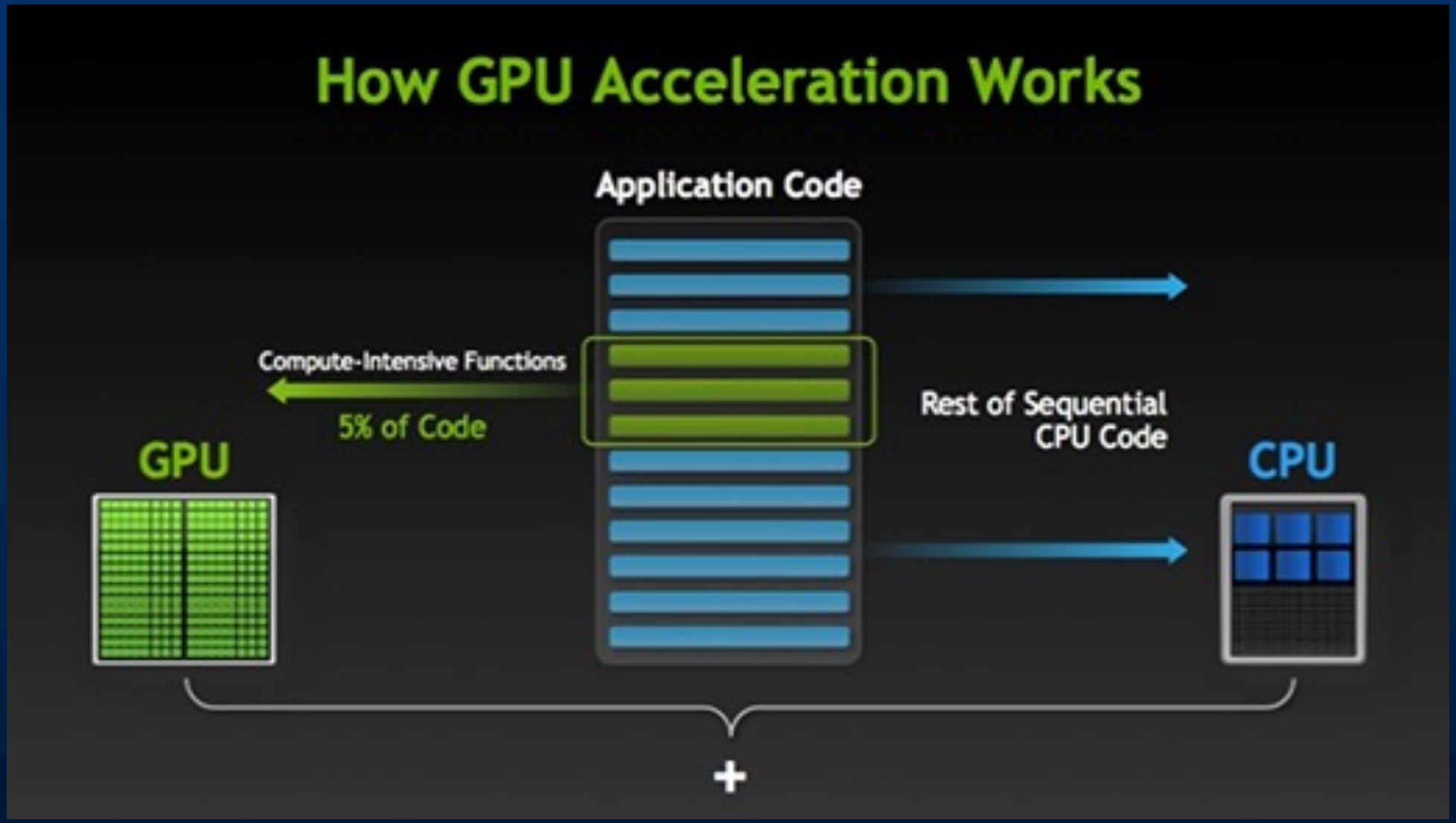
What Hath the GPU Wrought?



to

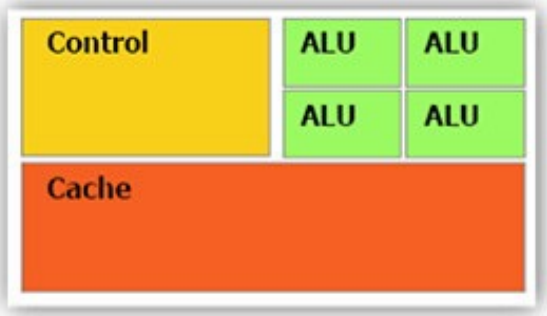


…at very low cost



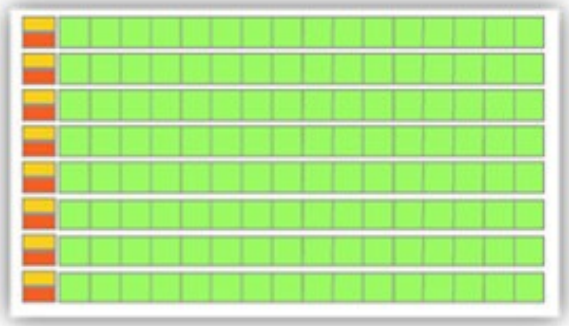
- https://nyu-cds.github.io/python-gpu/01-introduction/

CPU



* Low compute density
* Complex control logic
* Large caches (L1$/L2$, etc.)
* Optimized for serial operations
  - Fewer execution units (ALUs)
  - Higher clock speeds
* Shallow pipelines (<30 stages)
* Low Latency Tolerance
* Newer CPUs have more parallelism

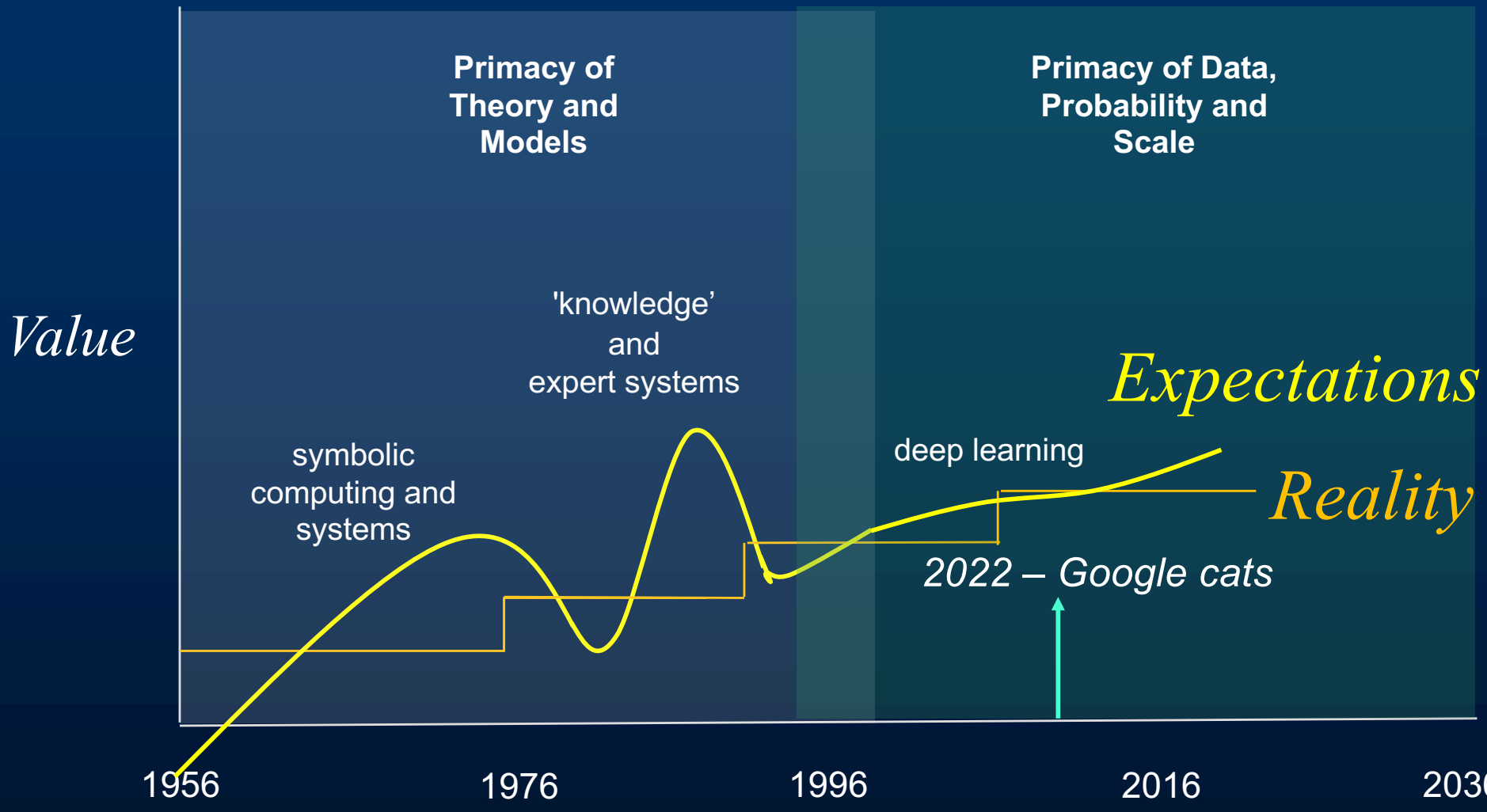
GPU



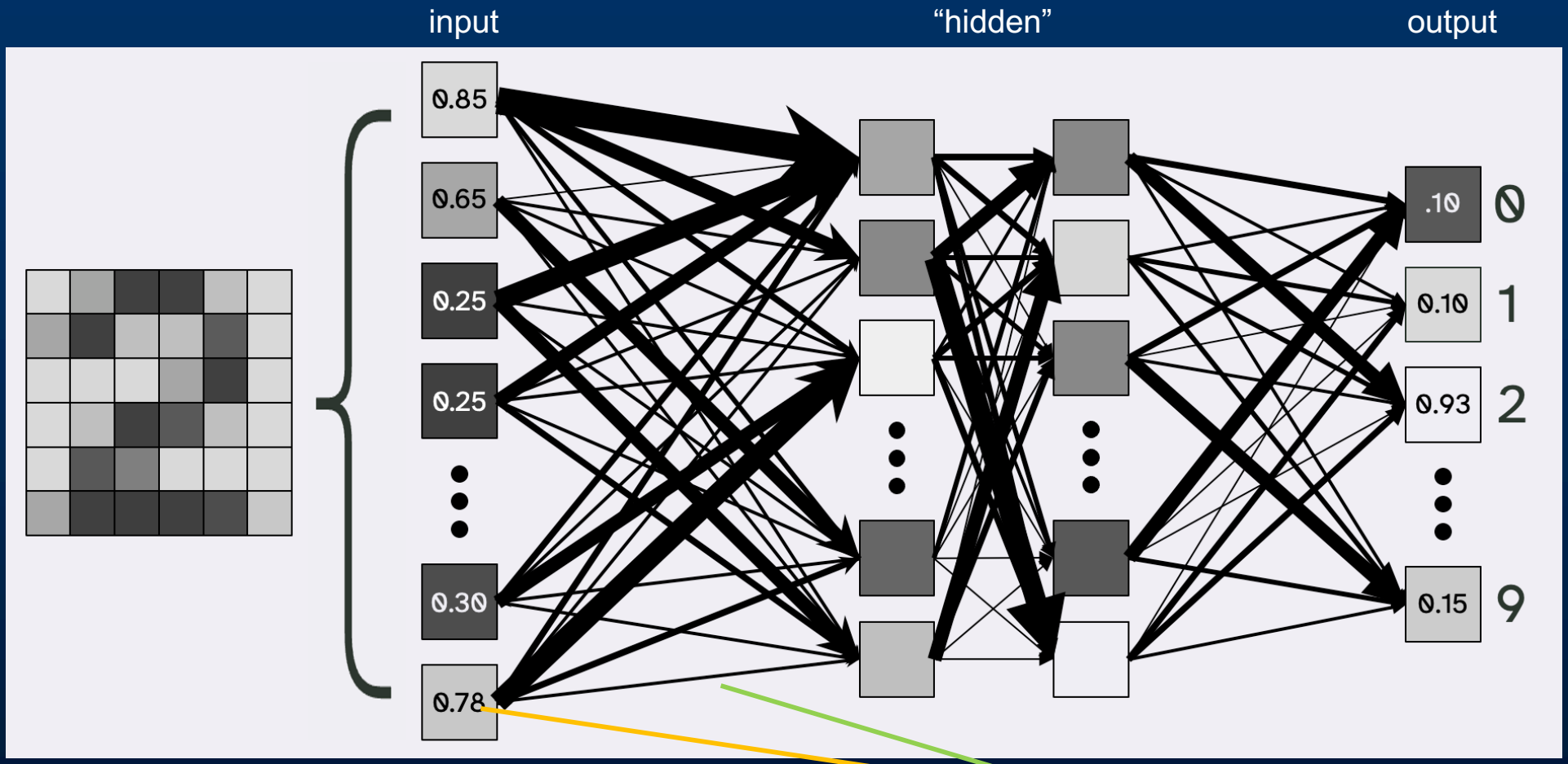
* High compute density
* High Computations per Memory Access
* Built for parallel operations
  - Many parallel execution units (ALUs)
  - Graphics is the best known case of parallelism
* Deep pipelines (hundreds of stages)
* High Throughput
* High Latency Tolerance
* Newer GPUs:
  - Better flow control logic (becoming more CPU-like)
  - Scatter/Gather Memory Access
  - Don't have one-way pipelines anymore

- https://embeddedcomputing.com/technology/processing/understand-the-mobile-graphics-processing-unit

A Brief Graphic History of AI



Neural Networks

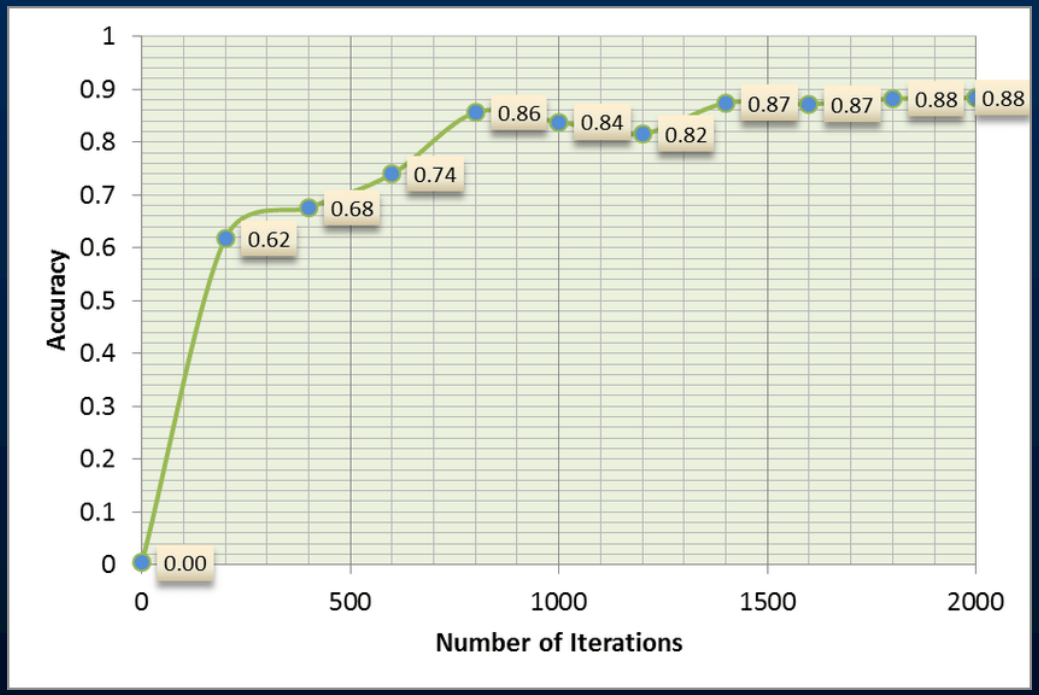


$$\text{activation value}_{l,j} = \varphi\left(w_0 i_0 + w_1 i_1 + \ldots + w_{35} i_{35} \right)$$

How does a network learn?

Training process:

- Label each item of training data with the correct expected output
- Process training data through the network to generate predictions
- Compare predictions with correct labels to determine error
- Adjust weights to reduce error*
- Iterate until desired accuracy is achieved



$$\text{activation value}_{l,i} = \varphi\left(w_0 i_0 + w_1 i_1 + \ldots + w_{35} i_{35}\right)$$

* Oh, and that guy...

Large networks have to calculate billions of weights.

If only we had a way to go faster…

8 MINUTE READ

How AI is contributing to the GPU shortage

Emmanuel Ohiri
Jan 2, 2024 at 2:10 AM

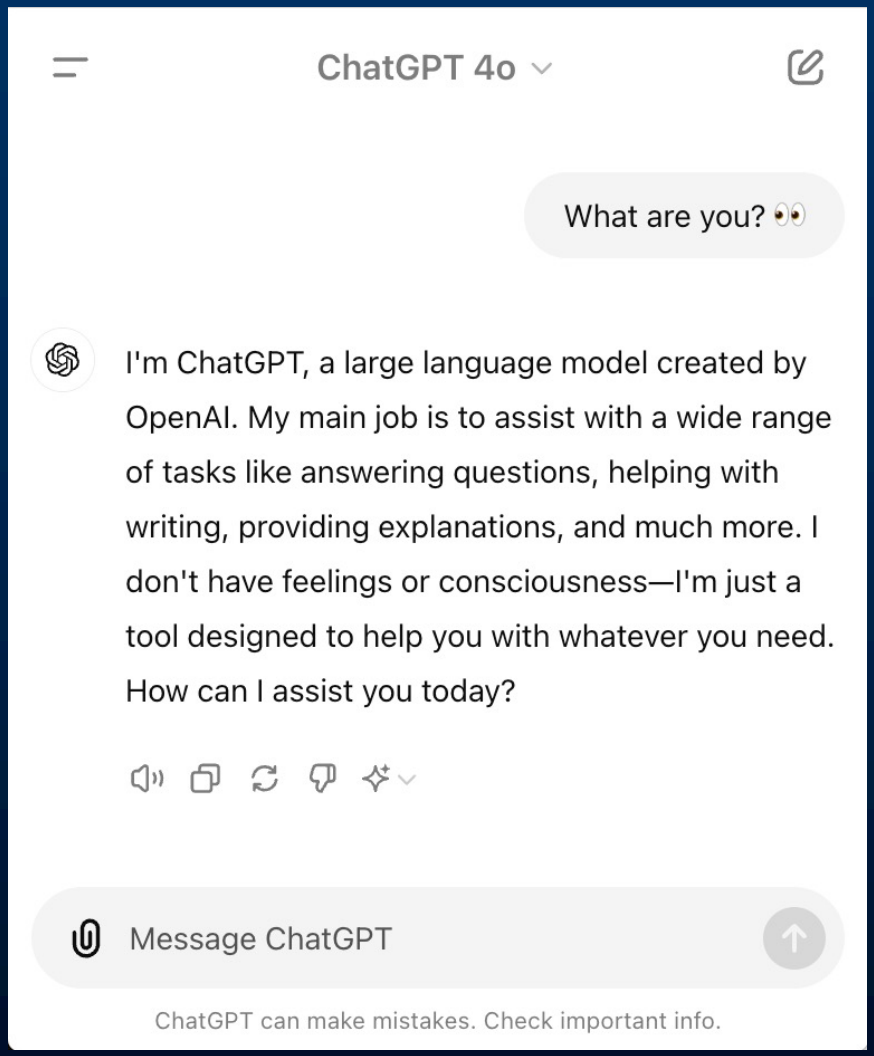
NVDA market capitalization10/30/24:
$3.44 Trillion

Large Language Models (LLMs)

- A (deep) network-based learning model for understanding *and generating* human language

- Many hidden layers and extensive dataset training

- Examp[le uses:
  - Question answering
  - Translation
  - Text generation

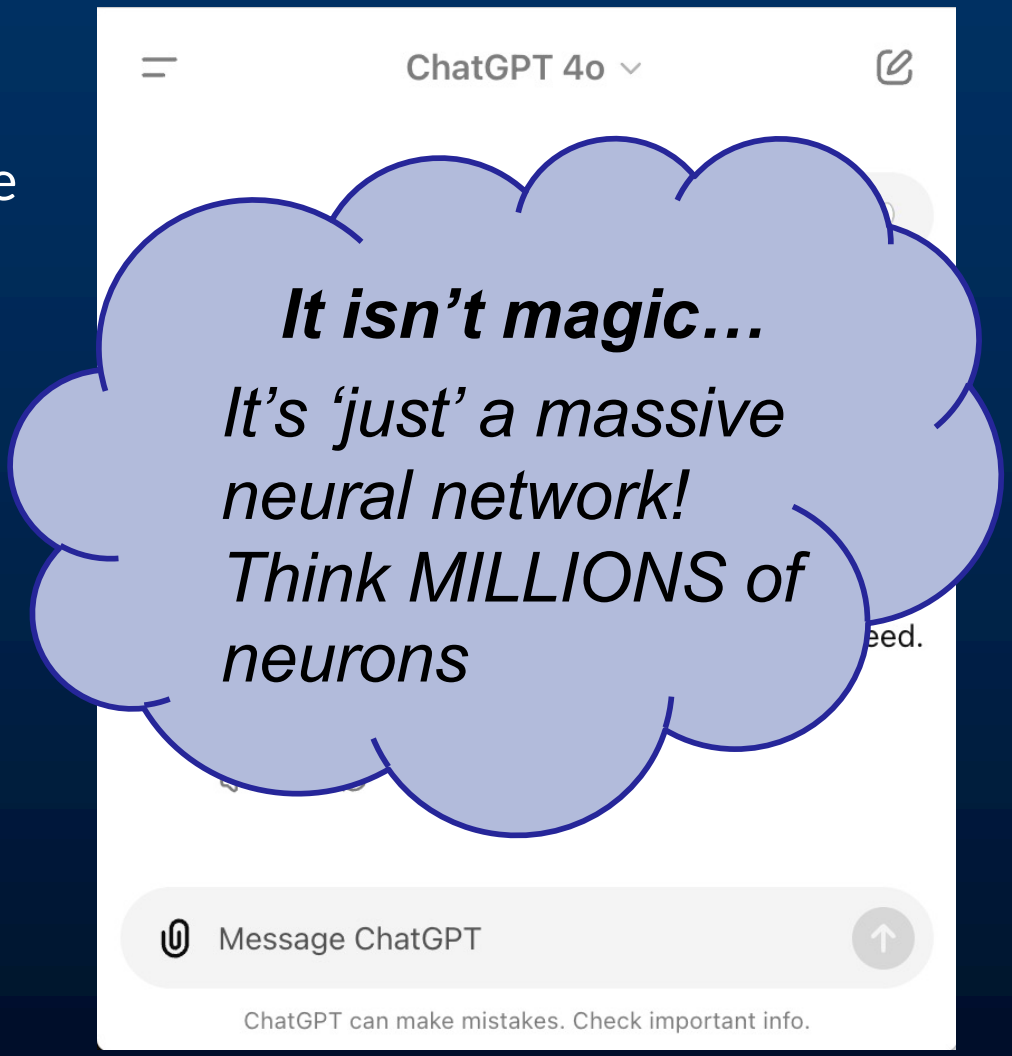


Large Language Models (LLMs)

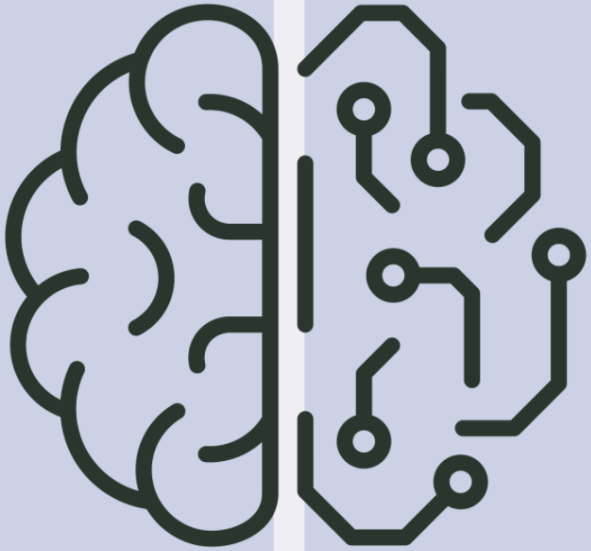
- A (deep) network-based learning model for understanding *and generating* human language

- Many hidden layers and extensive dataset training

- Example uses:
  - Question answering
  - Translation
  - Text generation



Large Language Models (LLMs)

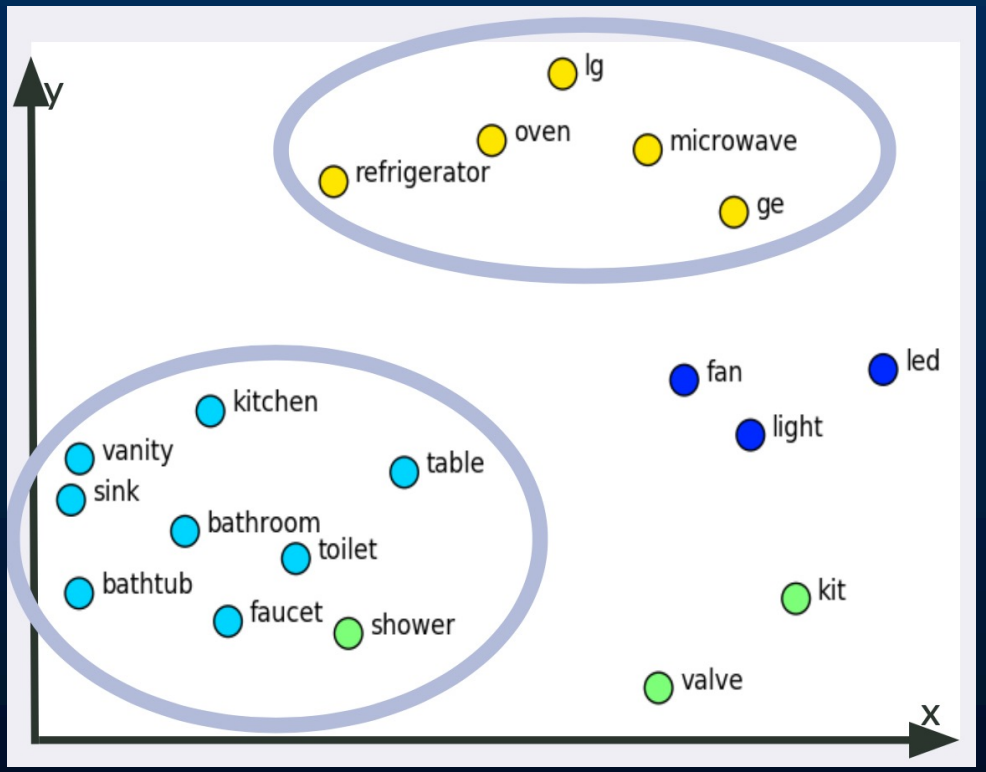
The human brain		GPT-4
~86 billion neurons[c]		~1-2 million[c] neurons
> 100 trillion synapses		~176 billion parameters

LLM Word Embeddings

- A technique to capture a word's semantic meaning
- Words are converted to large vectors (eg, 1024-d space)
- Vectors are adjusted during training -
  Similar words will have similar vectors

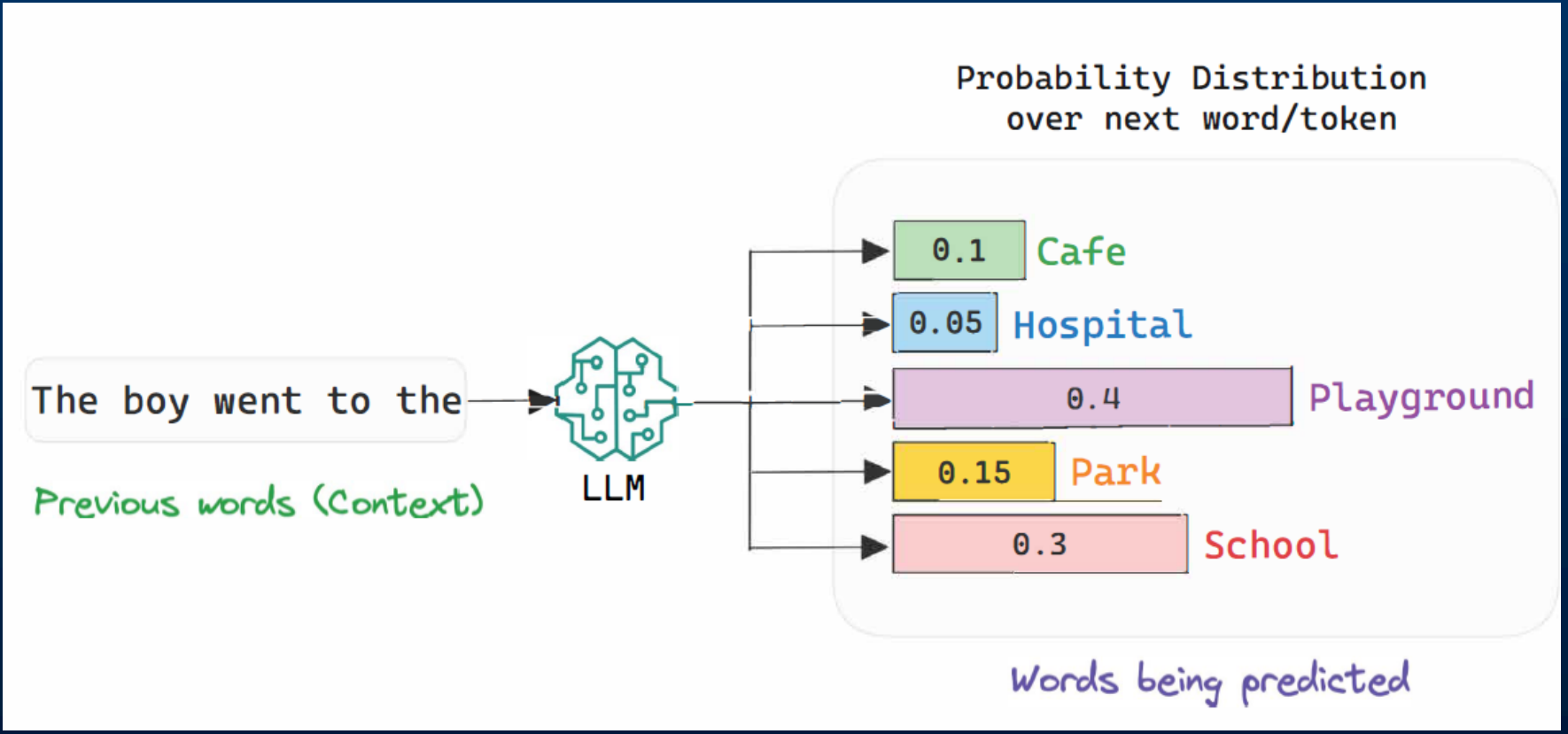
Simple Human (2d) view

- Each word is a point in space
- Nearness indicates similarity
- Clusters suggest related words



LLM Learning

An LLM learns by identifying patterns and structures from its unlabeled training data to **predict the next most likely word**…



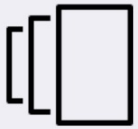
GPT-4's training data contained a whopping **300 billion words** of content!

- BUT No understanding of the world or of truth - it's been called a "stochastic parrot"[1]

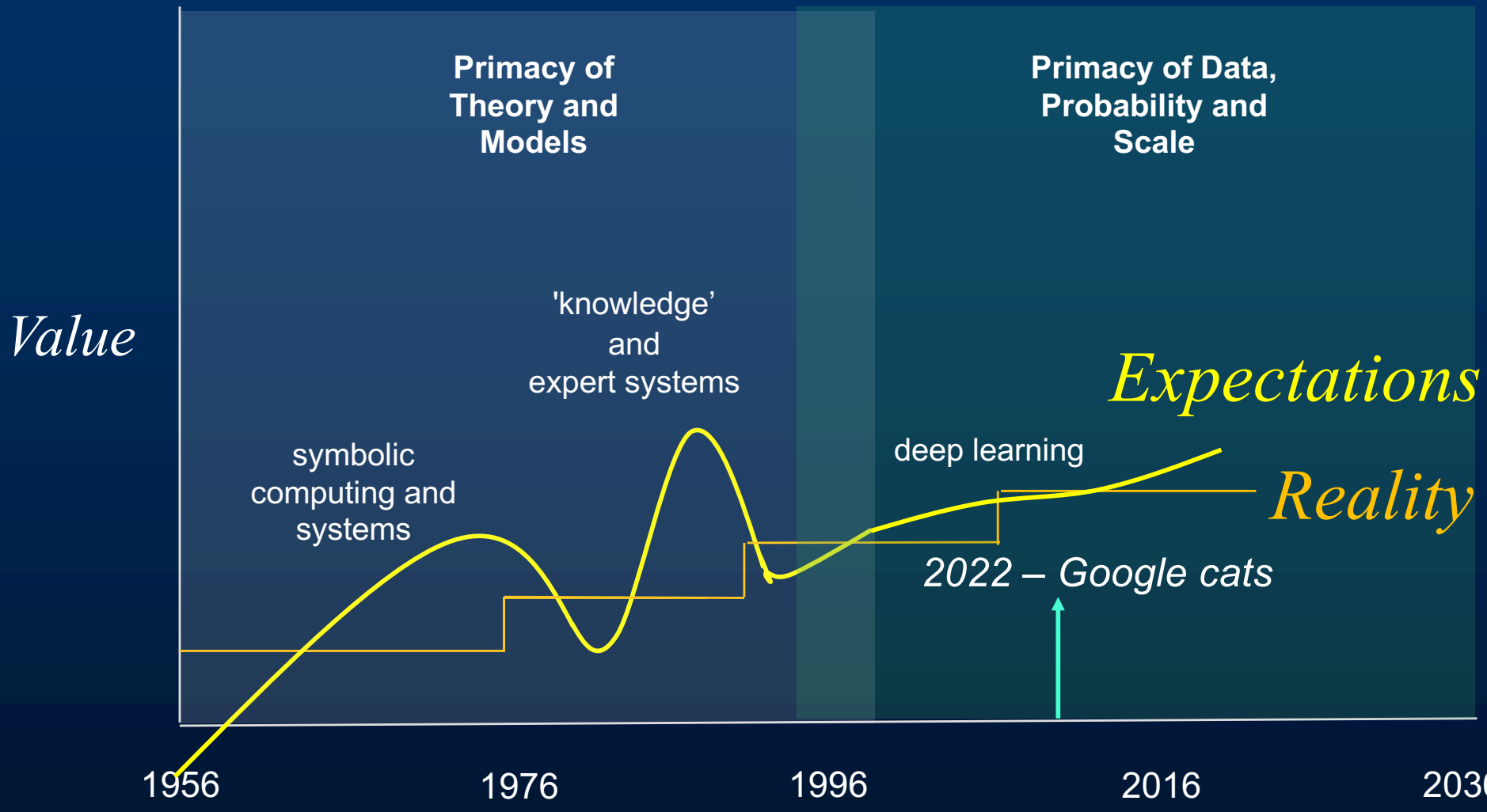
[1]Emily M Bender "*On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?"*

Fine-Tuning LLMs

Reinforcement
Learning
From
Human **F**eedback

Prompt is selected from prompt database	The model will generate multiple possible responses	Human evaluators rank these responses	The rankings are used to fine-tune the LLM using RLHF
			
"Write a polite response meant for an upset customer."	• *"I'm sorry you're upset..."* • *"Your concern is important to us..."* • *"Why are you so mad?..."*	Evaluator ranks responses based on criteria like politeness, clarity, and helpfulness.	LLM learns to prioritize higher ranked responses, adjusting weights to generate outputs aligned with human preferences

A Brief Graphic History of AI



Unsupervised Learning

future tense

In Artificial Intelligence Breakthrough, Google Computers Teach Themselves To Spot Cats on YouTube

BY WILL OREMUS

JUNE 27, 2012 • 9:34 AM

Unsupervised Learning

The Unreasonable Effectiveness of Data*

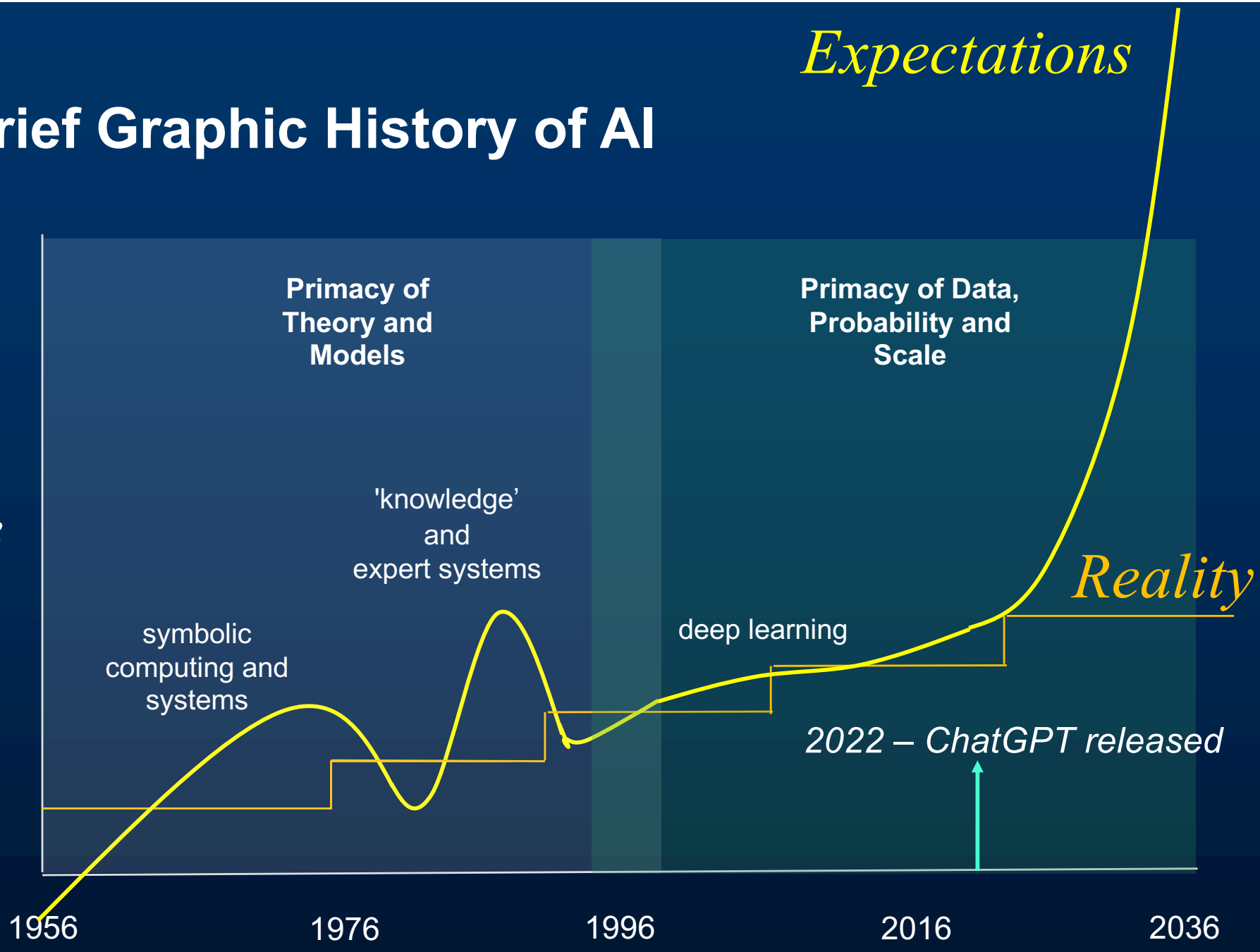
"Google's machines did home in on human faces as one of the more relevant features in the data set. They also developed the concepts of cat faces and human bodies—not because they were instructed to, but merely because the arrangement of pixels in image after image suggested that those features might be in some way important.

"Google's computers … fell far short of humans in several respects. After unsupervised learning followed by a period of supervised training, they picked out human faces with 82 percent accuracy. But their accuracy on a broad range of features that humans consider relevant was a far more humble 16.7 percent."

"…the computers "learned" a slew of concepts that have little meaning to humans. For instance, they became intrigued by "tool-like objects oriented at 30 degrees," including spatulas and needle-nose pliers."

* Halevy, Norvig and Periera, *The Unreasonable Effectiveness of Data*, IEEE Intelligent Systems, 2009

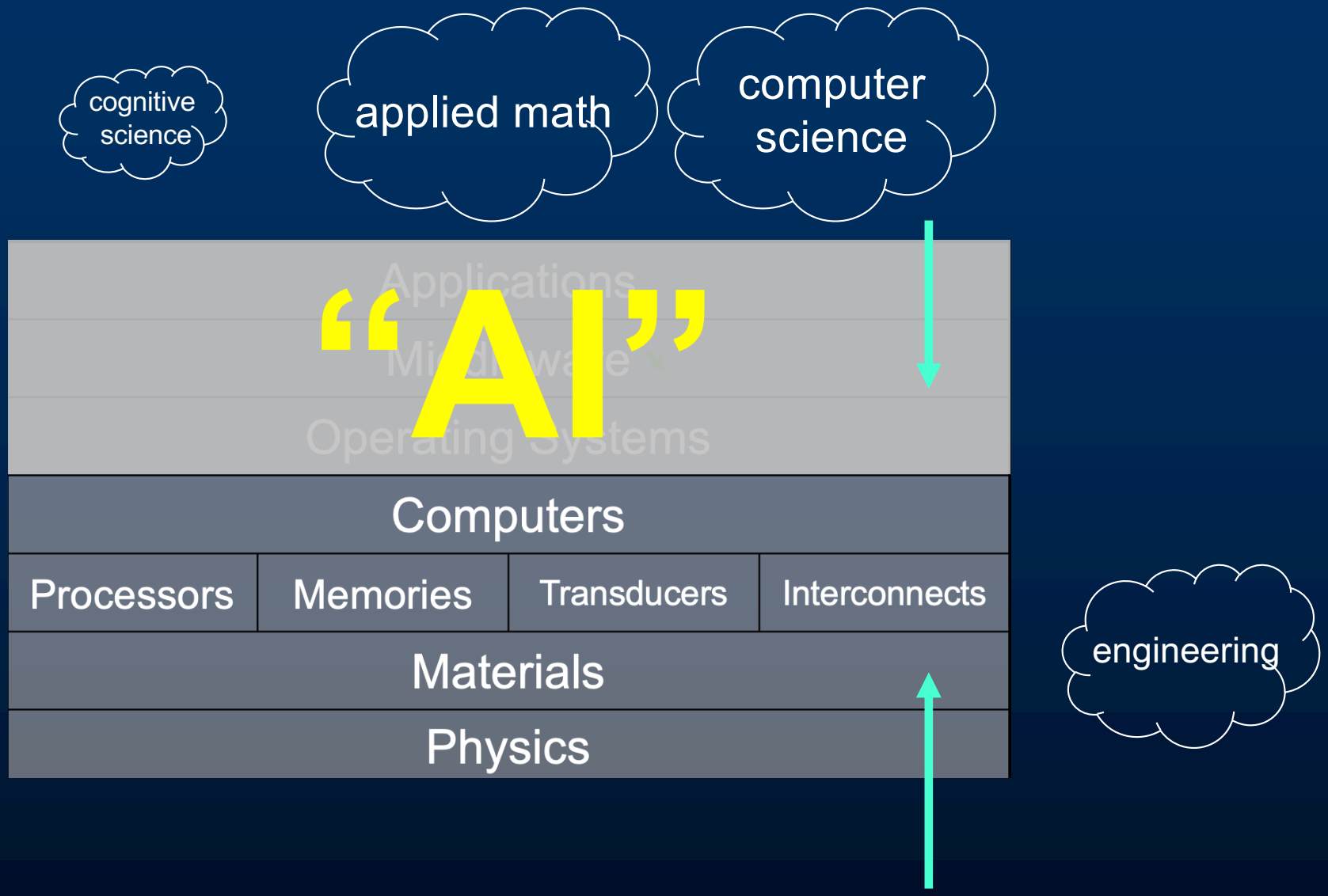
A Brief Graphic History of AI



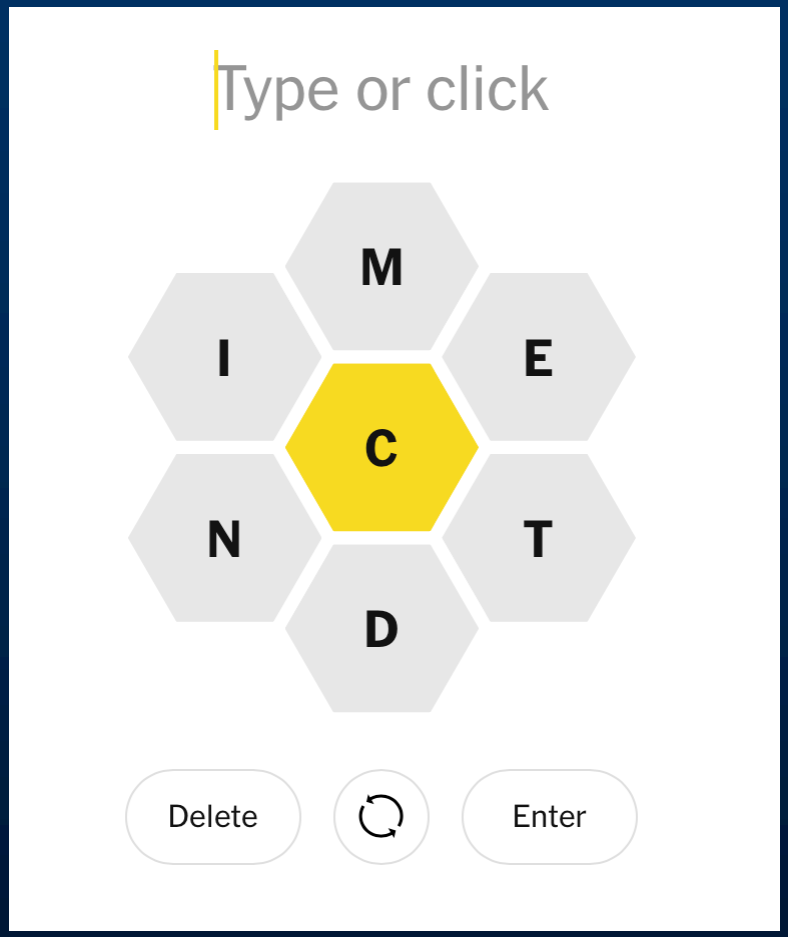
OpenAI market capitalization:
$157 Billion

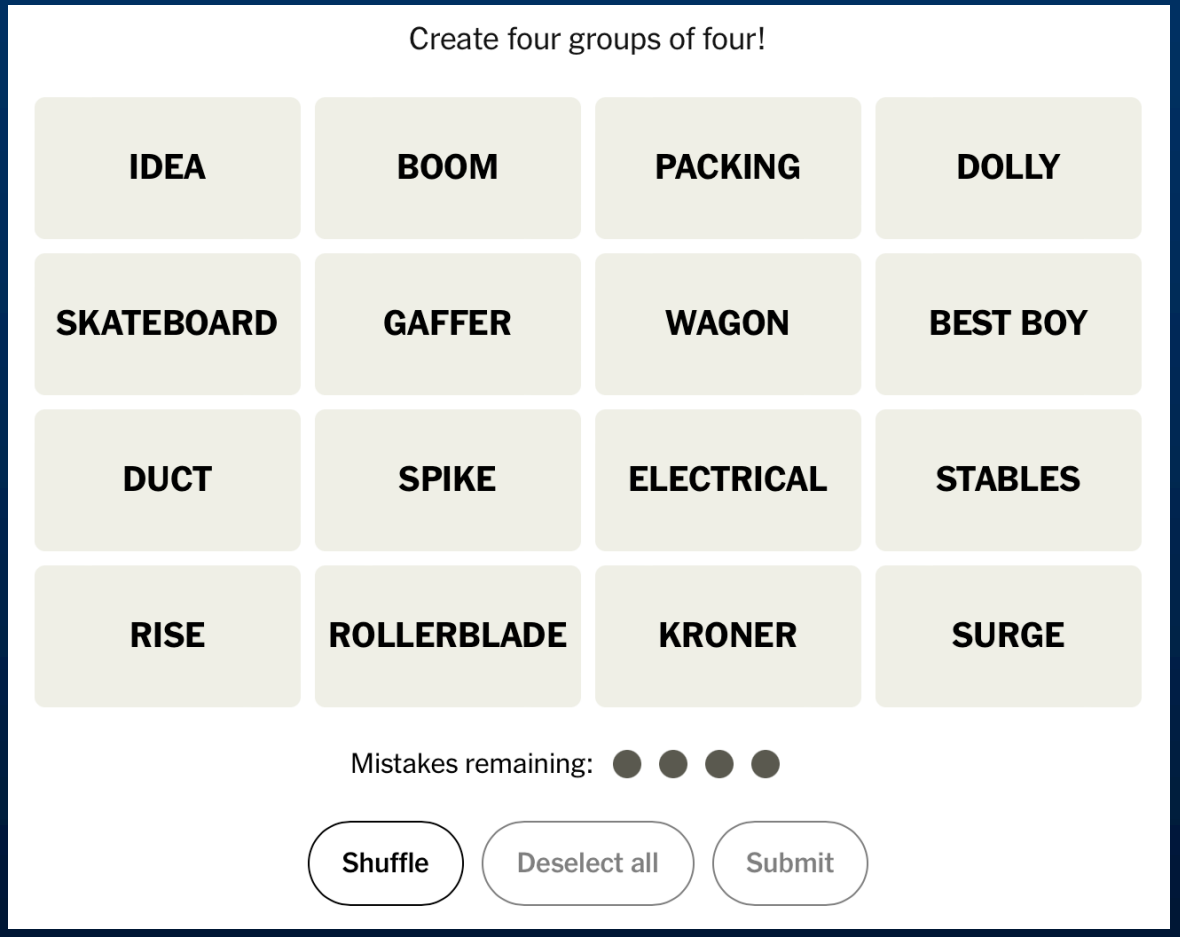
OpenAI 2024 revenue:
$3.7 Billion

AI - now


cognitive science
applied math
computer science
Applications
Middleware
Operating Systems
"AI"
Computers
Processors
Memories
Transducers
Interconnects
Materials
Physics
engineering

Audience Participation





“The Singularity” – fiction or inevitability?

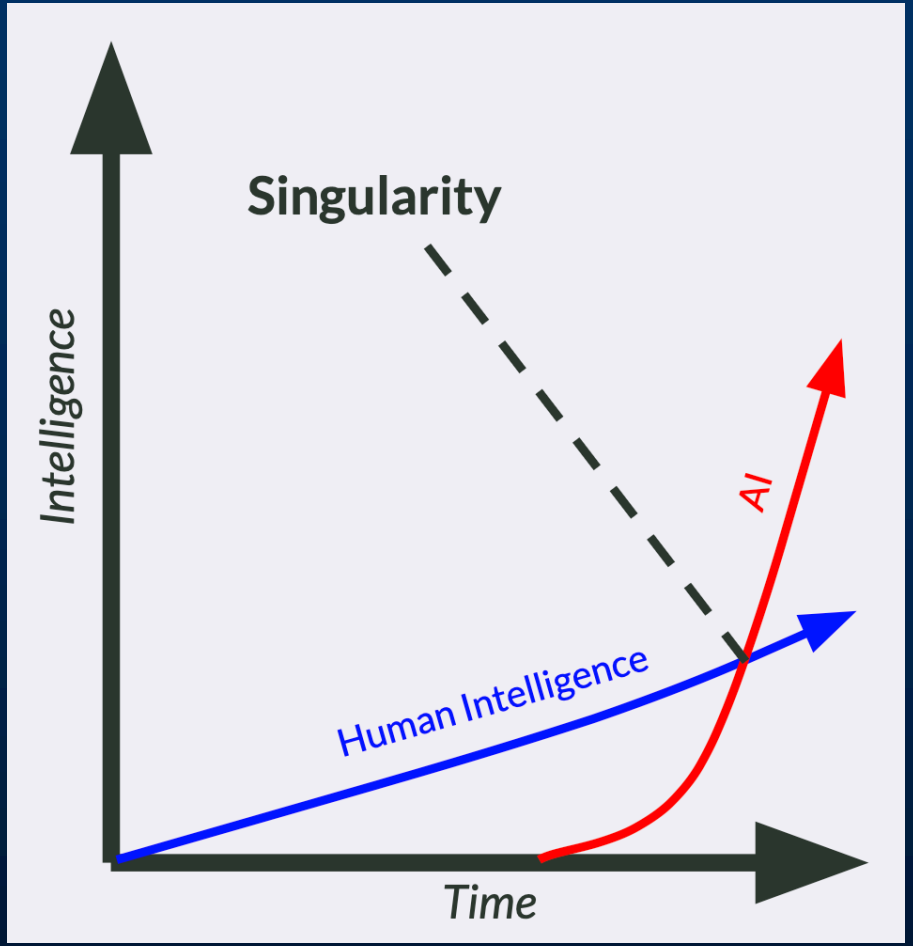
A hypothetical future point where AI surpasses human intelligence, potentially leading to a rapid, uncontrollable technological explosion.

Pathway to Singularity:

- Recursive self-improvement:
  - AI enhancing its own algorithms
- Accelerating advancements in AI technologies (e.g., machine learning, neural networks)

What are some things we can do?

- Develop and enforce international regulations
- Promote transparency in AI algorithms
- Support interdisciplinary collaboration between AI, ethics, and social sciences
- Encouraging public input in shaping AI policies and governance



Option A, or Option B?

Option A

Superintelligence by 2100 is inevitable

Option B

Superintelligence by 2100 is impossible

Option A, or Option B?

Myth:
Superintelligence by 2100 is inevitable

Myth:
Superintelligence by 2100 is impossible

Mon	Tue	Wed	Thr	Fri	Sat	Sun
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	✓	22	23	24	25
26	27	28	29	30		

Fact:
It may happen in decades, centuries or never: AI experts disagree & we simply don't know

Worry A, or Worry B, or both?

Worry A

AI turning evil

Worry B

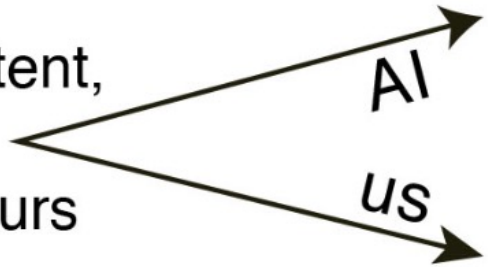
AI turning conscious

Worry A, or Worry B, or both?

Mythical worry:
AI turning evil

Mythical worry:
AI turning conscious

Actual worry:
AI turning competent, with goals misaligned with ours



‘Competent’ AI is like The Force

Light Side

- A powerful tool in service of human good

  Organizing / retreiving knowledge
  Discovery
  Interacting with the physical world

‘Competent’ AI is like The Force

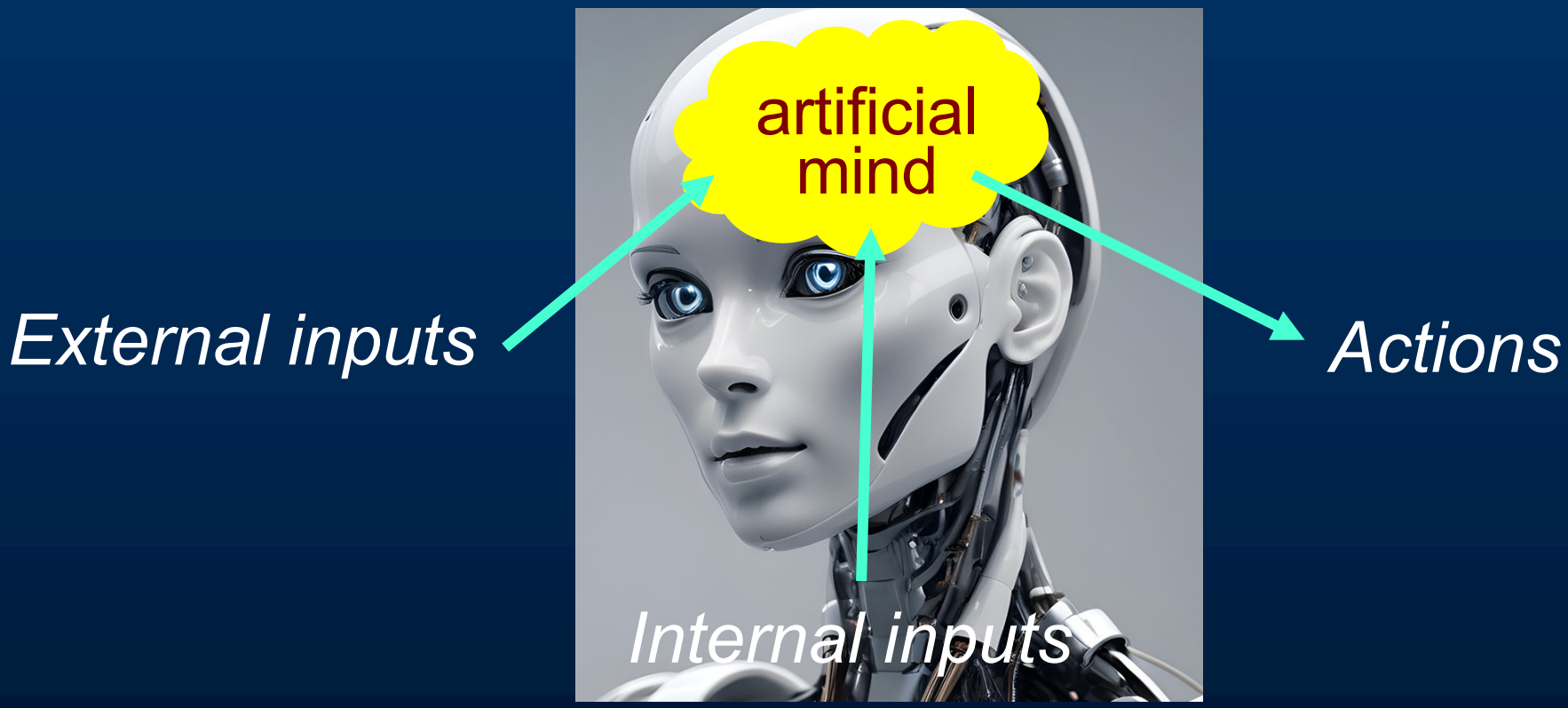
Light Side

- Organizing / retreiving knowledge
- Discovery
- Interacting with the physical world

Dark Side

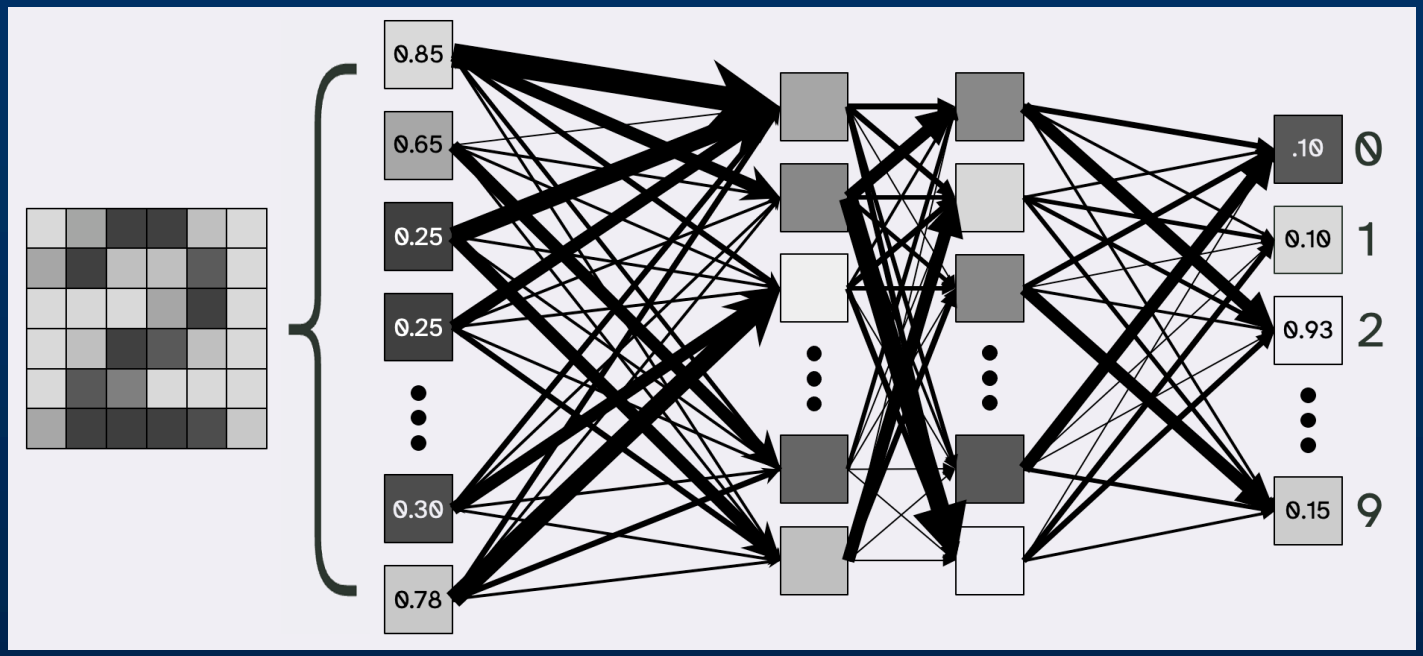
- Automation of totalitarianism
- Autonomous weapons
- Destruction of intellectual property (evil, or not?)
- Deep Fakes
- AI sludge

The holy grail?

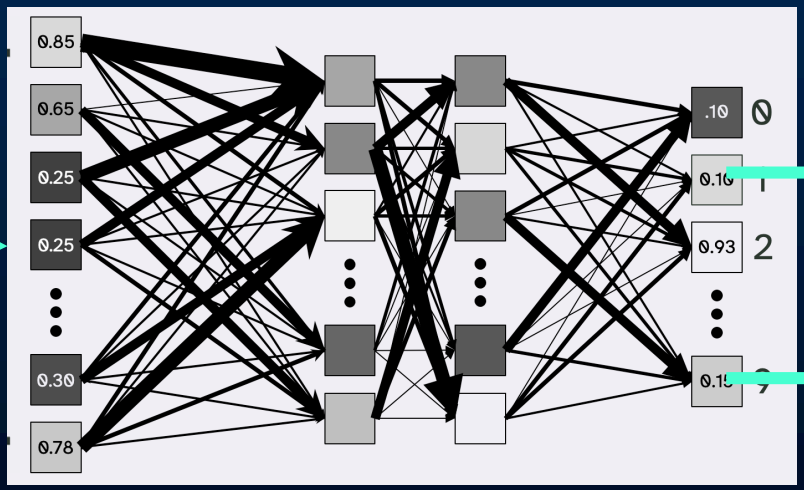


Neural Networks Are Amoral

*"…once the rocket goes up,
who cares where it comes down?
That's not our department…"*



???



Right

Wrong

‘Incompetent’ AI is like the plague

Tesla Using 'Full Self-Driving' Hits Deer Without Slowing, Doesn't Stop

How's that computer-vision-only approach working out for you, Tesla?

‘There are no guardrails.’ This mom believes an AI chatbot is responsible for her son’s suicide

By Clare Duffy, CNN

6 minute read · Updated 2:17 PM EDT, Wed October 30, 2024

An insidious threat?

- Massive workforce disruption based on perceived lower cost of AI solutions

An insidious threat?

"...but it's only non-skilled jobs."

- Disruption is across industries and job type:
  - financial reports
  - press releases
  - computer programs
  - color grading

An insidious threat?

"…but it's only at the entry level."

Where do most experienced (human) employees come from?

With threat comes opportunity

Build a society where AI remains a servant for good - it won’t happen by itself.

Questions & Comments