

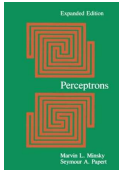
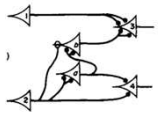
Artificial Intelligence

Technology and Applications

Arlindo Oliveira



Key Moments in the History of Artificial Intelligence



1943

1950

1952

1959

1967

1969

McCulloch-Pitts Neuron Model

Turing – Can machines think?

Rosenblatt - Perceptron

Samuel - Checkers

Weizenbaum - ELIZA

IBM DeepBlue

Rumelhart & McClelland PDP Book

Rumelhart, Hinton, Williams, MLP - Learning representations by back-propagating errors

Fukushima Neocognitron



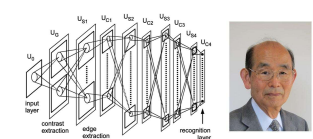
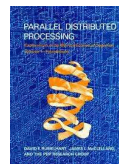
2006

1997

1987

1986

1979



2009

2012

2016

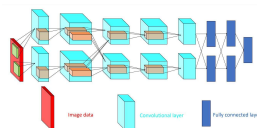
2017

2020

Li Fei-Fei et. al - ImageNet



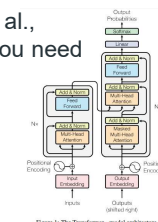
Krizhevsky, Sutskever & Hinton AlexNET



Silver et al. AlphaGo



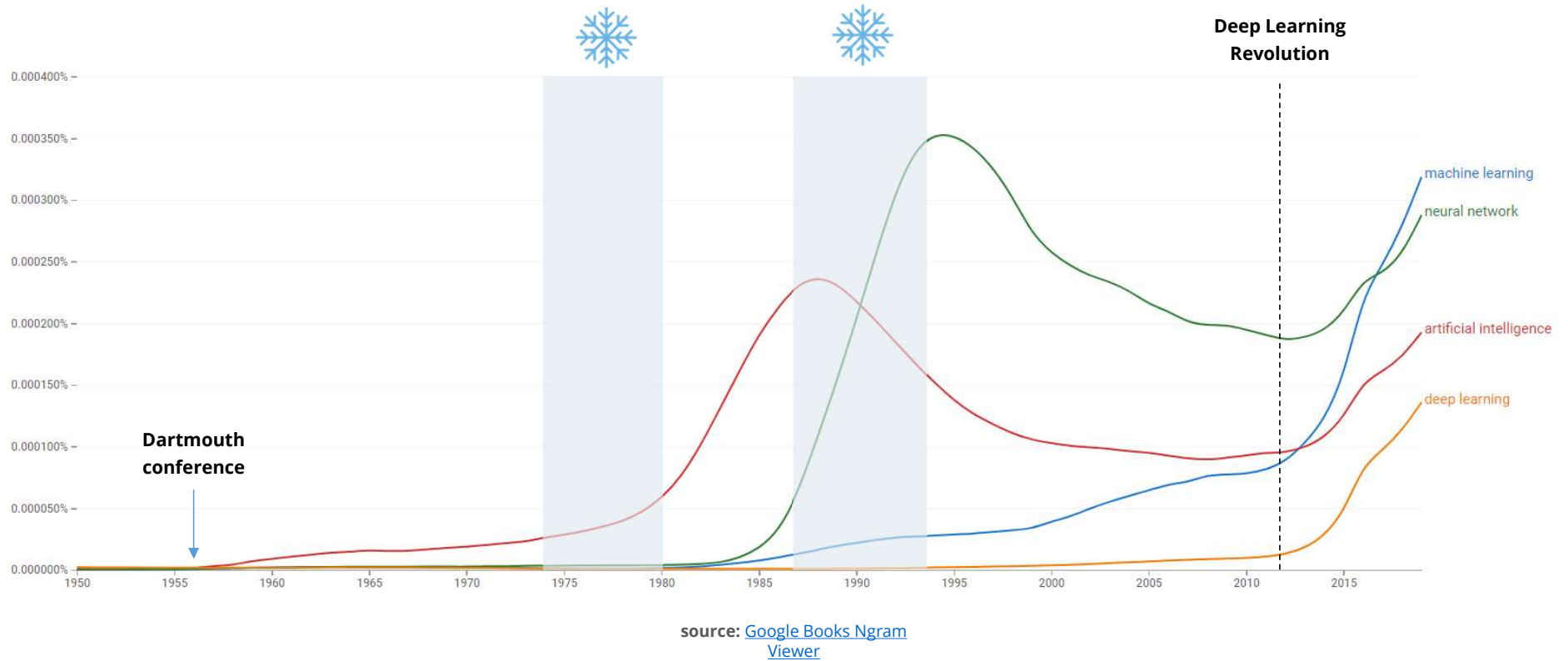
Vaswani et al., Attention is all you need



Brown et al. GPT-3



70 years of history



Moravec's Paradox

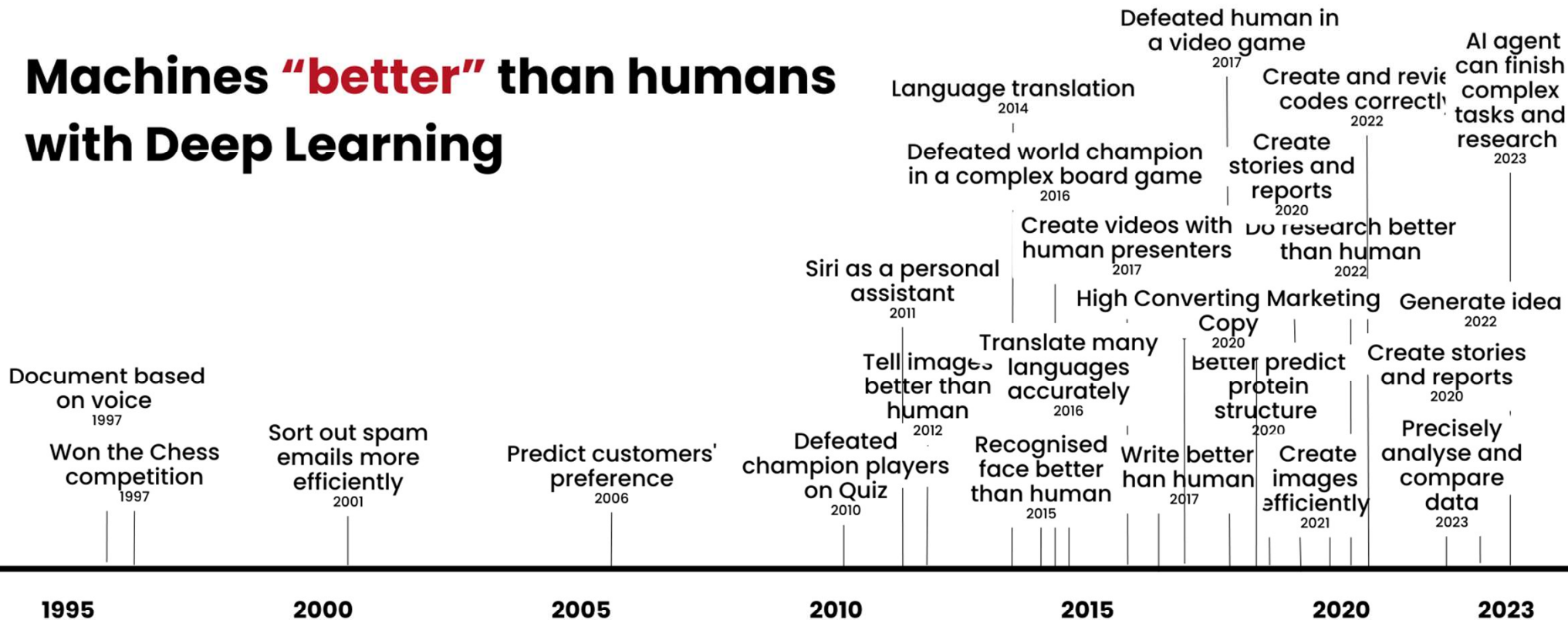


Robots and AI systems find the difficult things easy and the easy things difficult

Theorem proving, playing world-class chess, planning routes on a map, scheduling jobs, doing math, are relatively easy to program.

Recognizing faces, understanding scenes, walking in a room, understanding speech, are very, very hard to program.

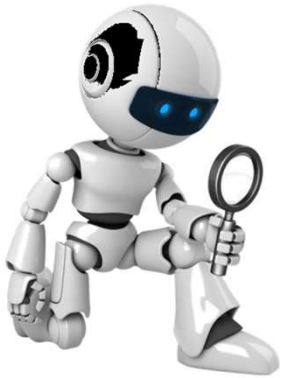
Machines “better” than humans with Deep Learning



Deep Learning (2010s-now)

Machine Learning (1995 - Now)

Artificial Intelligence Sub-Fields



Search



Reasoning



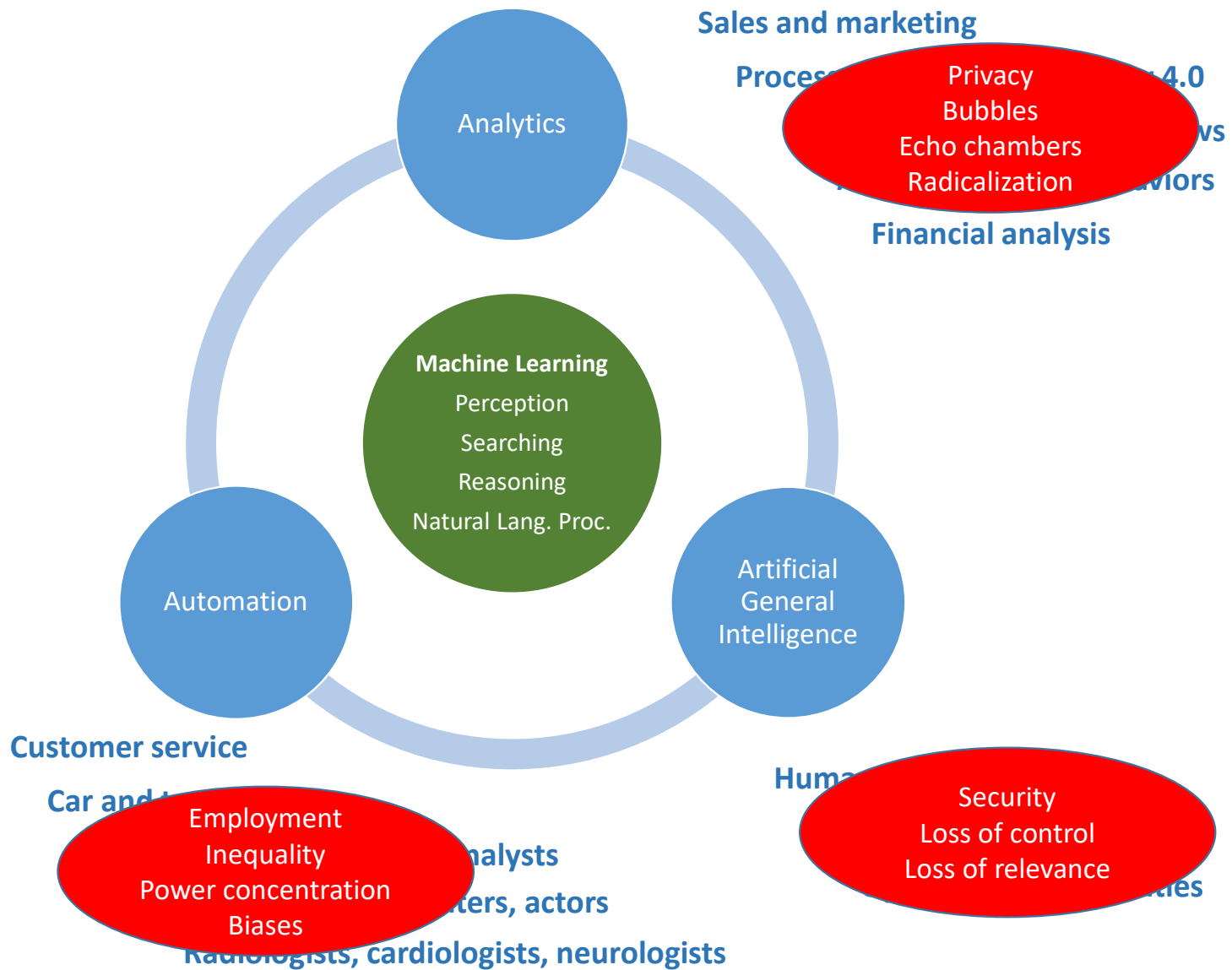
Natural Language Processing



Interaction and perception

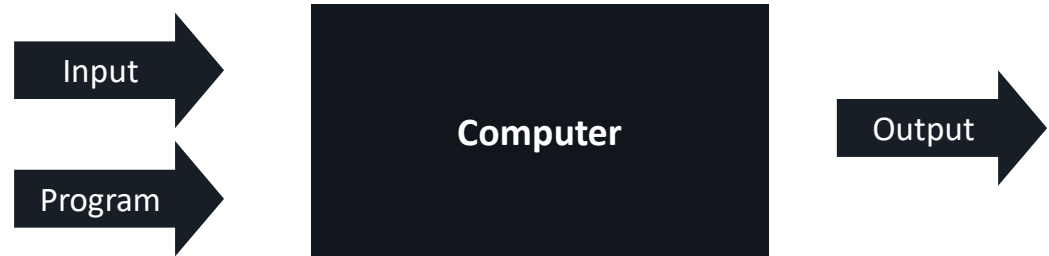


Machine Learning

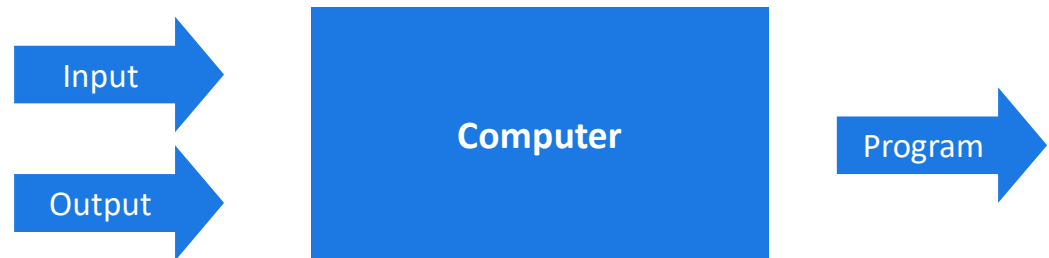


Machine Learning

Traditional programming



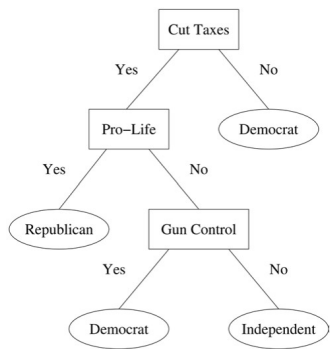
Machine Learning



Four Paradigms of Machine Learning

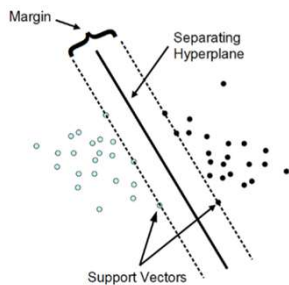
Symbolists

All intelligence can be reduced to symbol manipulation



Analogizers

The key to learning is recognizing similarities between situations



Statistical

All learned knowledge is uncertain, and learning itself is a form of uncertain inference



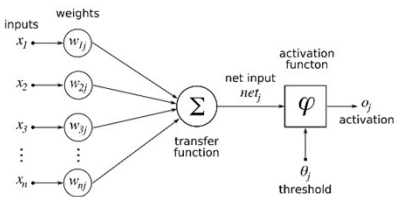
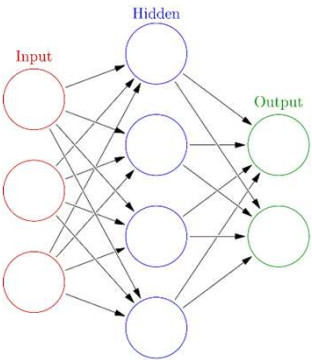
$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

Labels for the equation:

- Prior Probability: $P(H)$
- Likelihood of the evidence 'E' if the Hypothesis 'H' is true: $P(E|H)$
- Posterior Probability of 'H' given the evidence: $P(H|E)$
- Prior probability that the evidence itself is true: $P(E)$

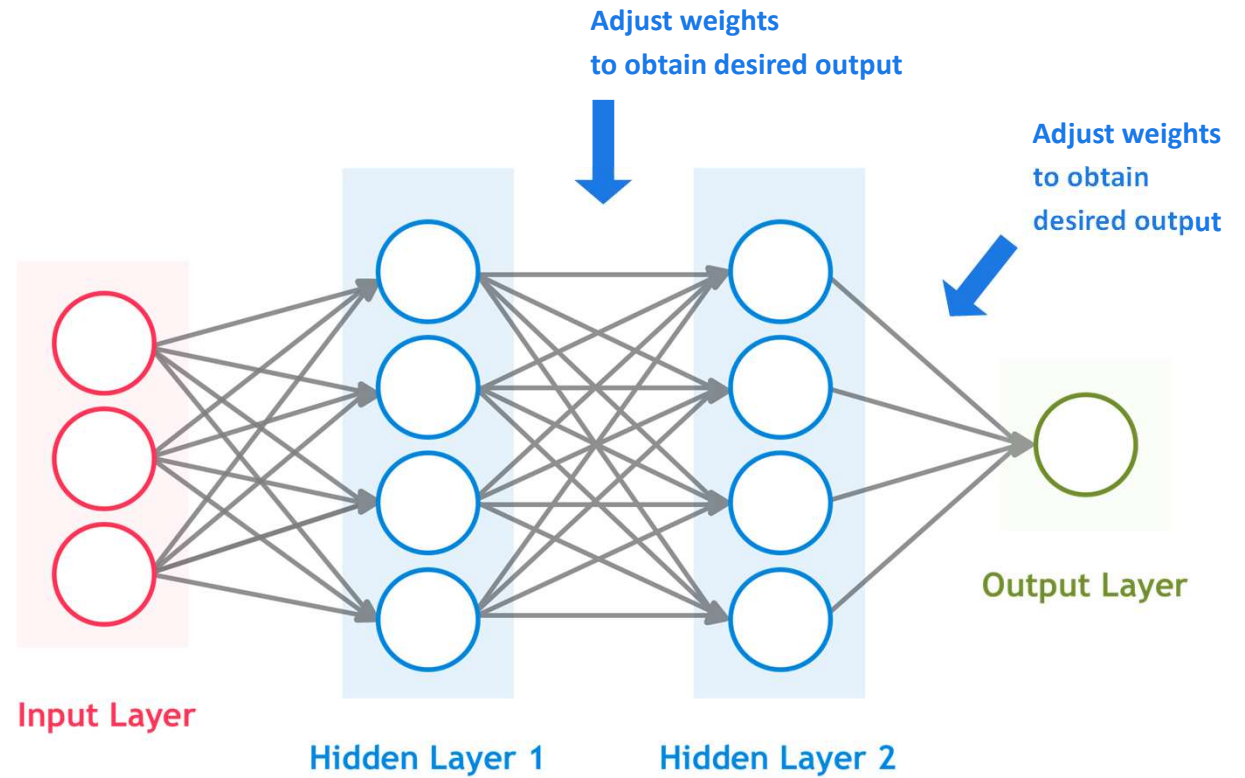
Connectionists

Learning is what the brain does, and what we need to do is reverse engineer it



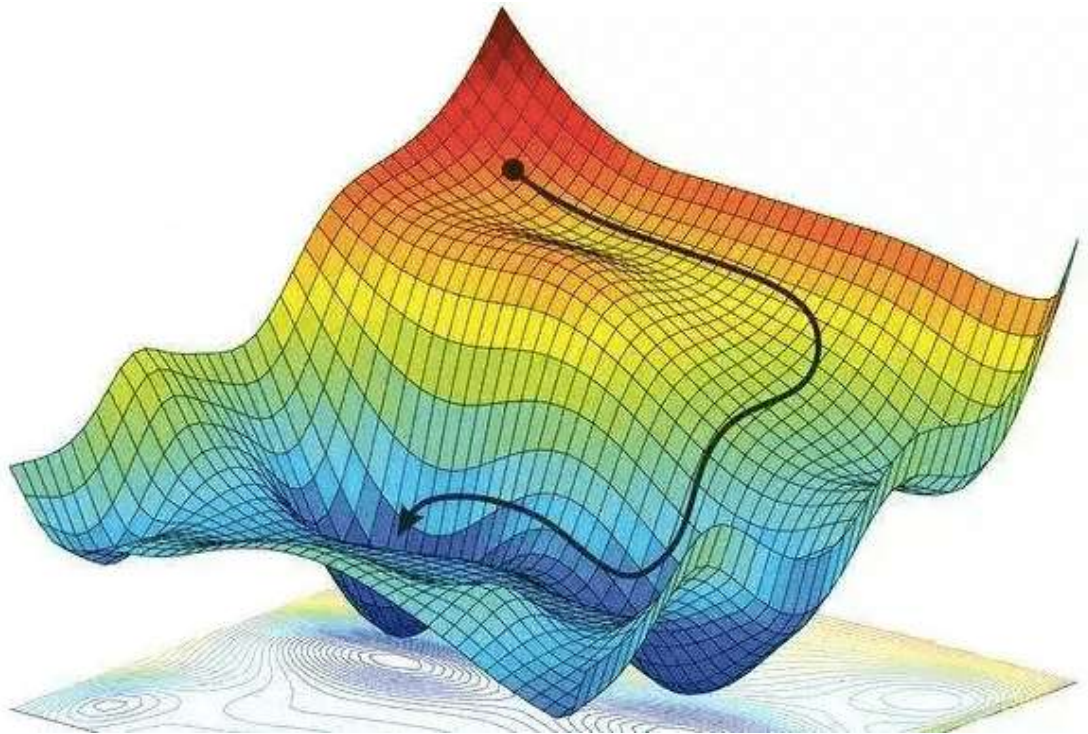
Neural Networks

Multi Layer
Perceptron



Gradient Descent

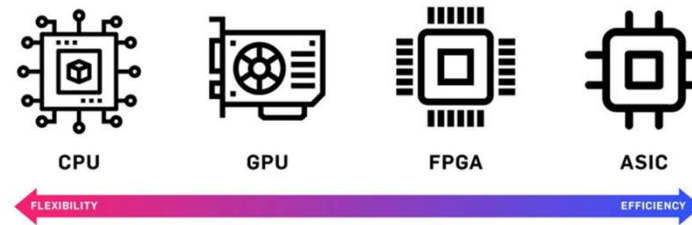
Error backpropagation



The Deep Learning Revolution

Frictionless reproducibility

Shared code
Hardware



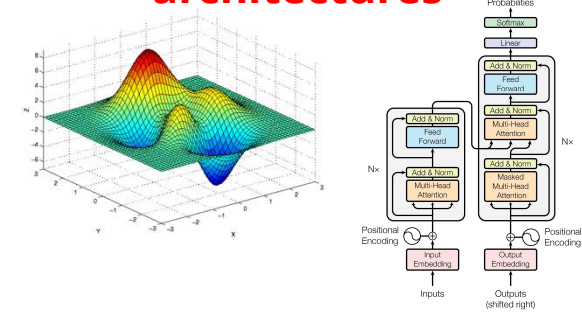
Deep Learning Revolution

Shared Data
Data



Competitive environment
Algorithms and architectures

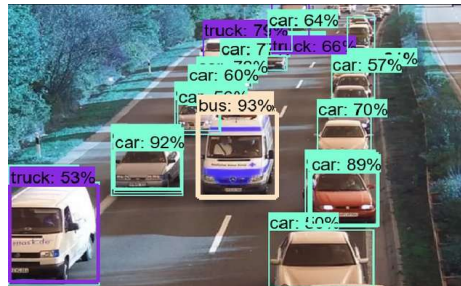
David Donoho



The Deep Learning Revolution

Four recent key developments

Deep Convolutional Neural Networks



Transformers and Large Language Models



Deep Learning Revolution



Deep Reinforcement Learning



Diffusion Models

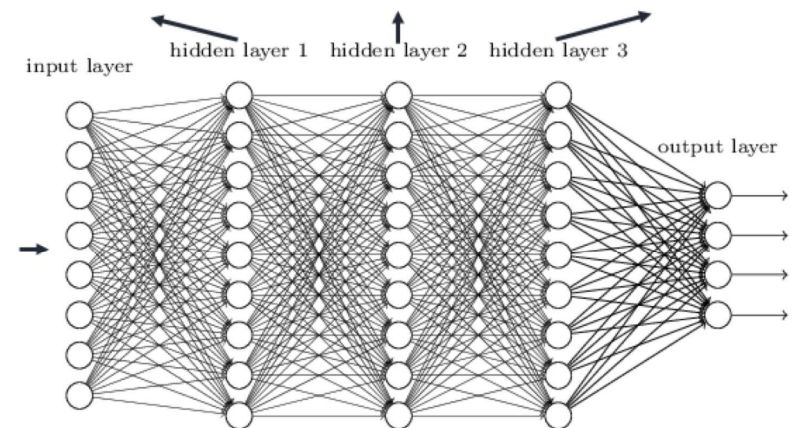
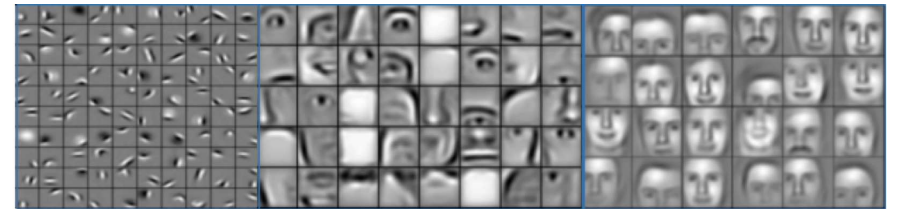




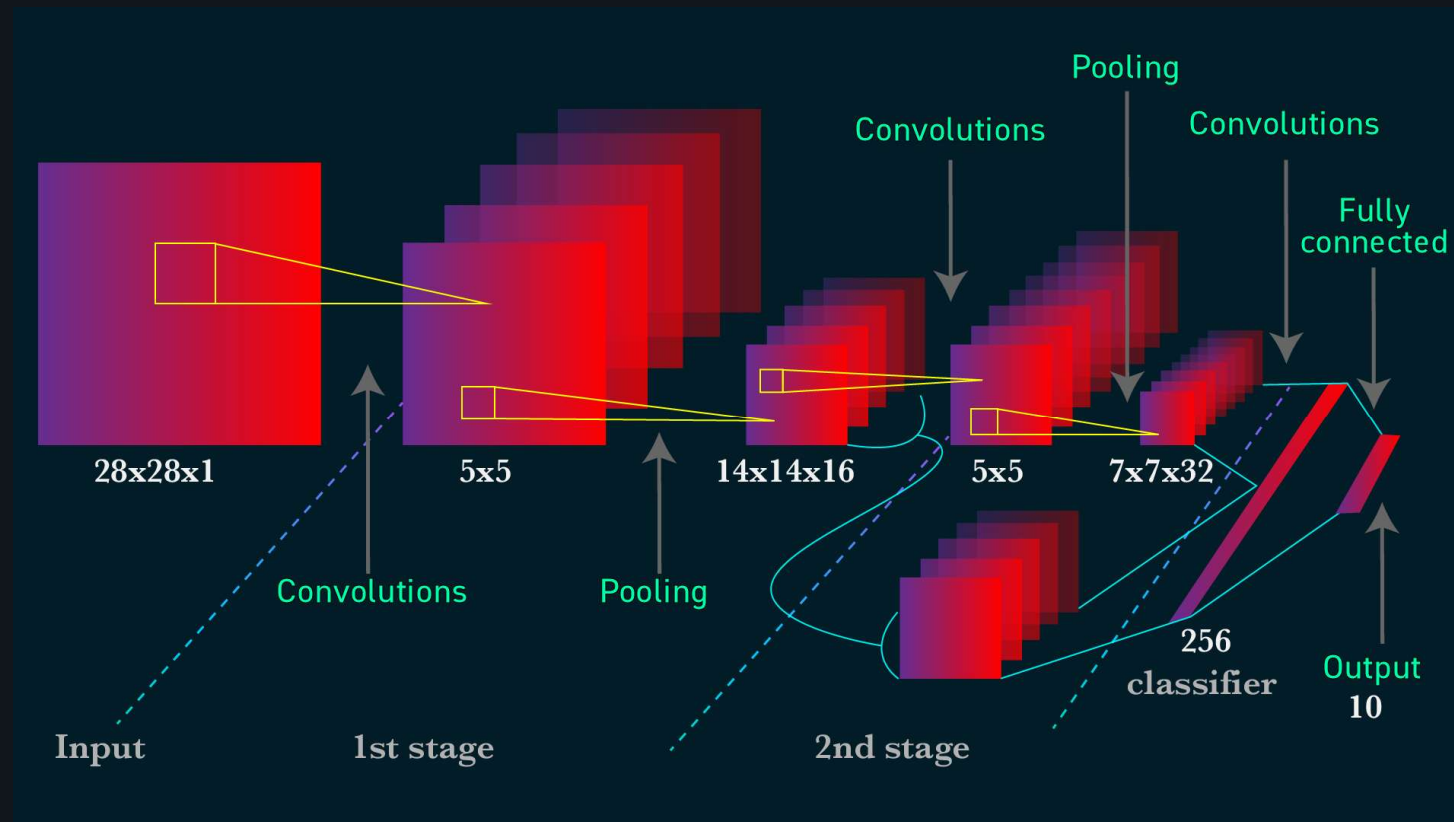
Deep convolutional neural networks in computer vision

Deep Neural Networks

Deep Neural Networks learn
hierarchical feature
representations



Convolutional Neural Networks (CNNs)



HOW A DEEP NEURAL NETWORK SEES

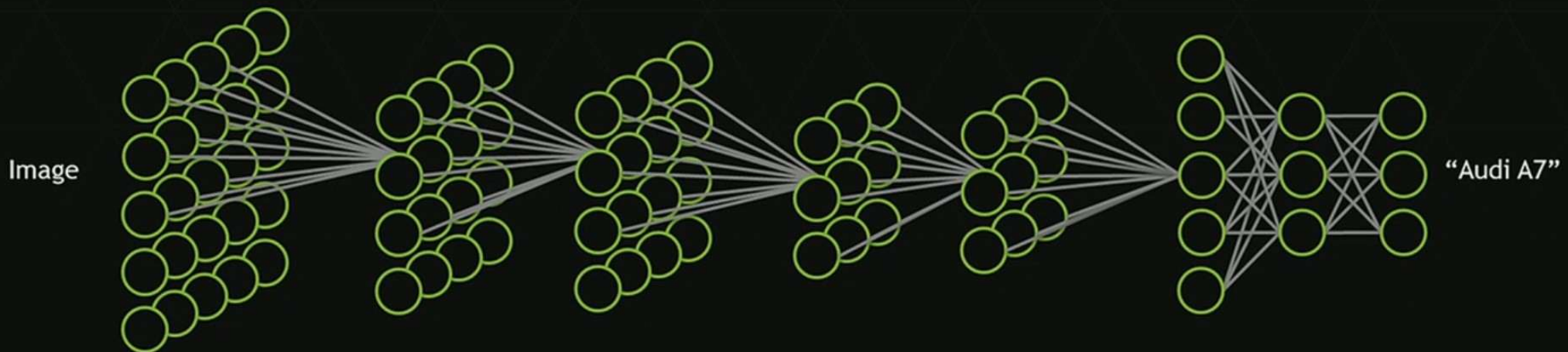
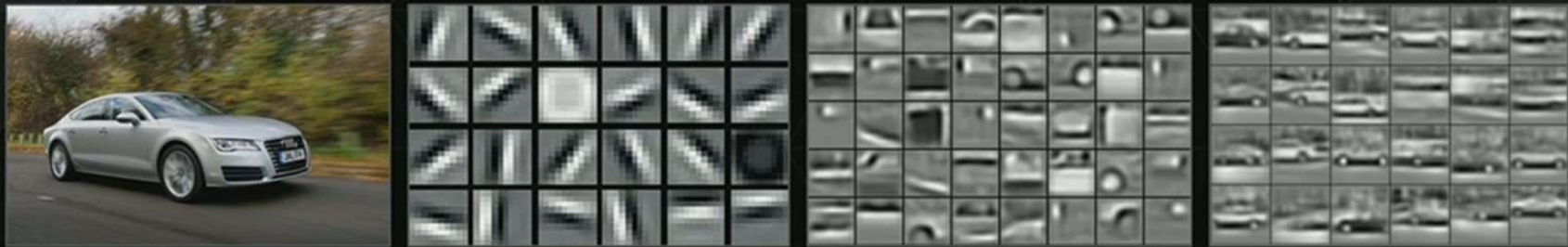
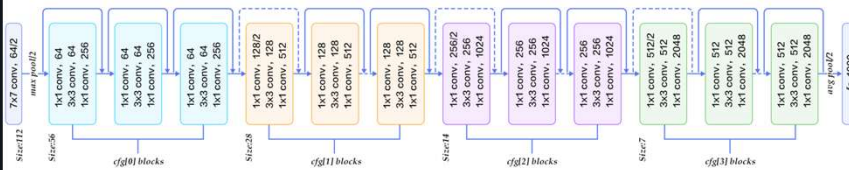
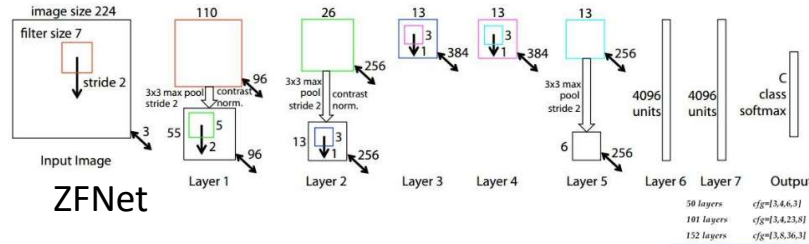
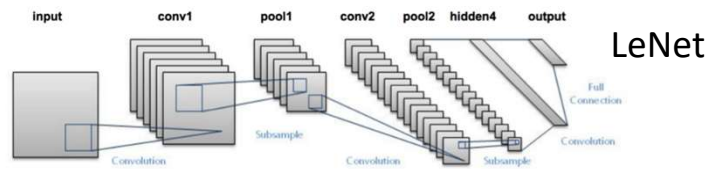


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

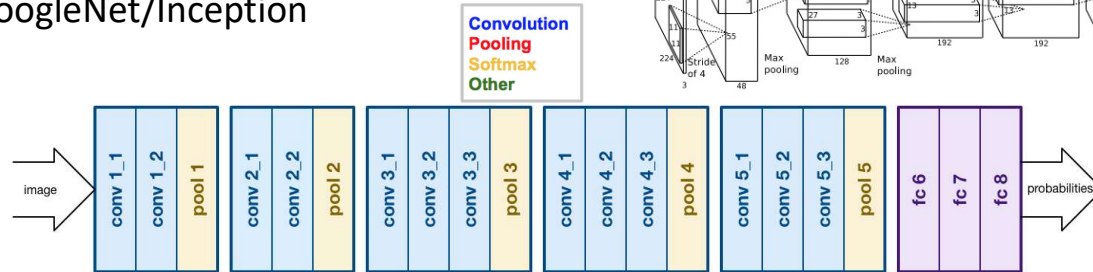
Convolutional Neural Networks (CNNs)



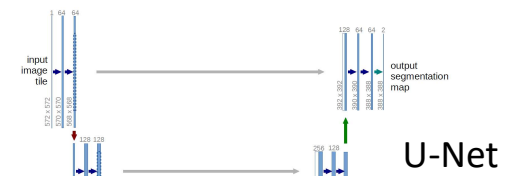
ResNet



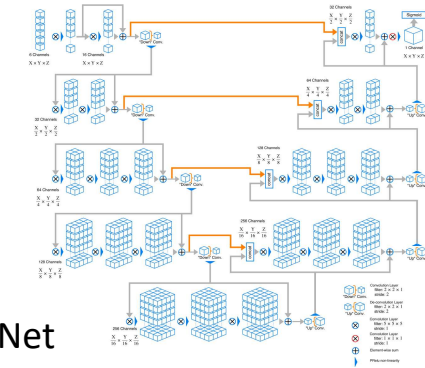
GoogleNet/Inception



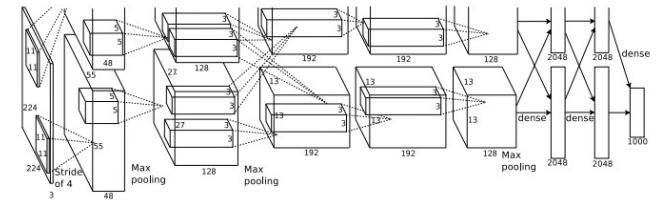
VGGNet



U-Net

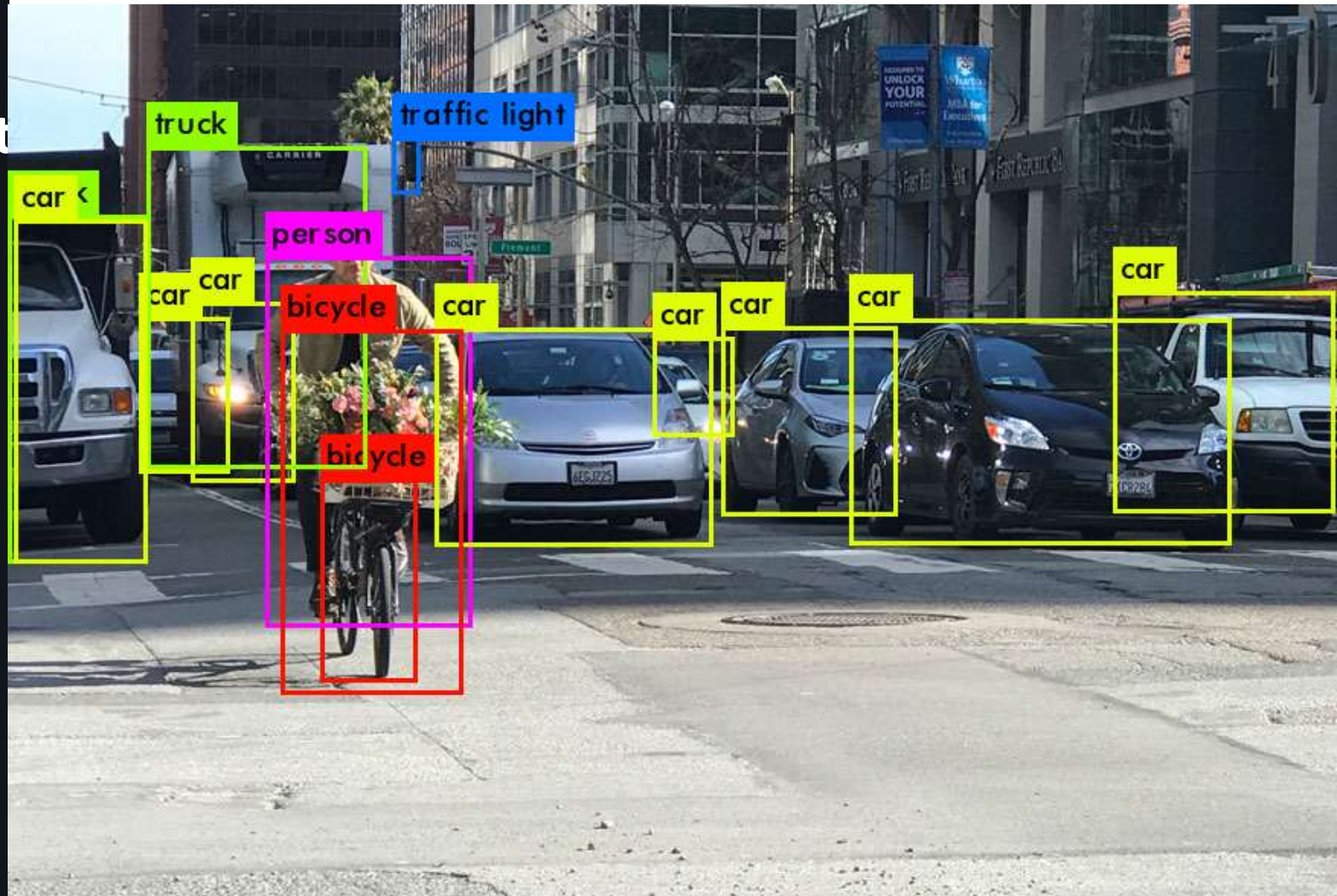


V-Net



AlexNet

YOLO for object detection



Damage Detection in vehicles

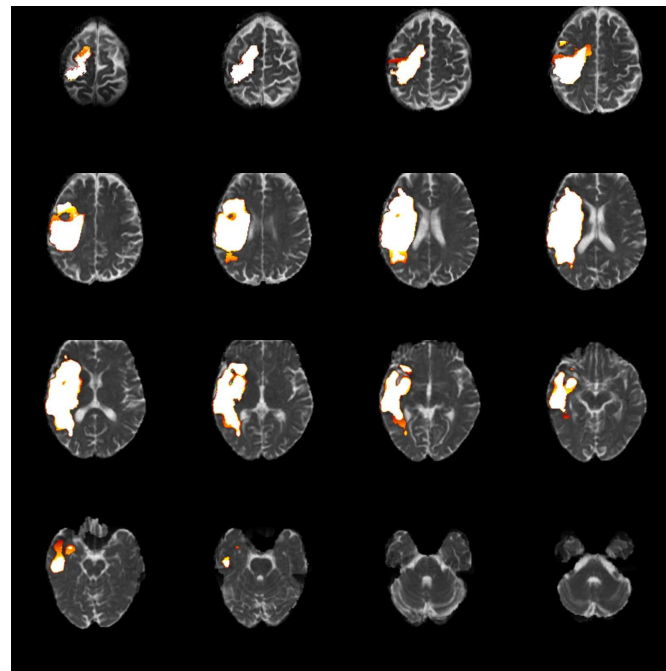
Dataset

- 1500 Images
 - 750 with Damage
 - 250 Bumpers
 - 250 Wheels
 - 250 Doors
 - 750 without Damage

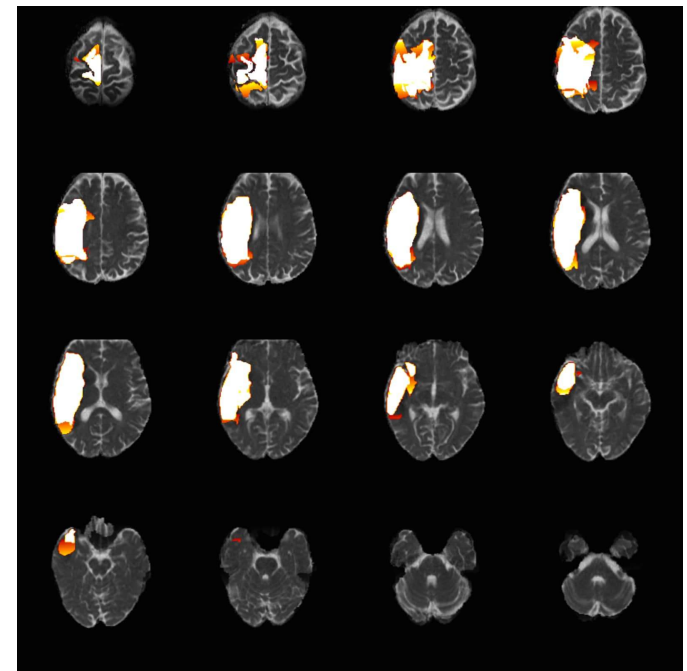


Ischemic Stroke Prediction

Best case



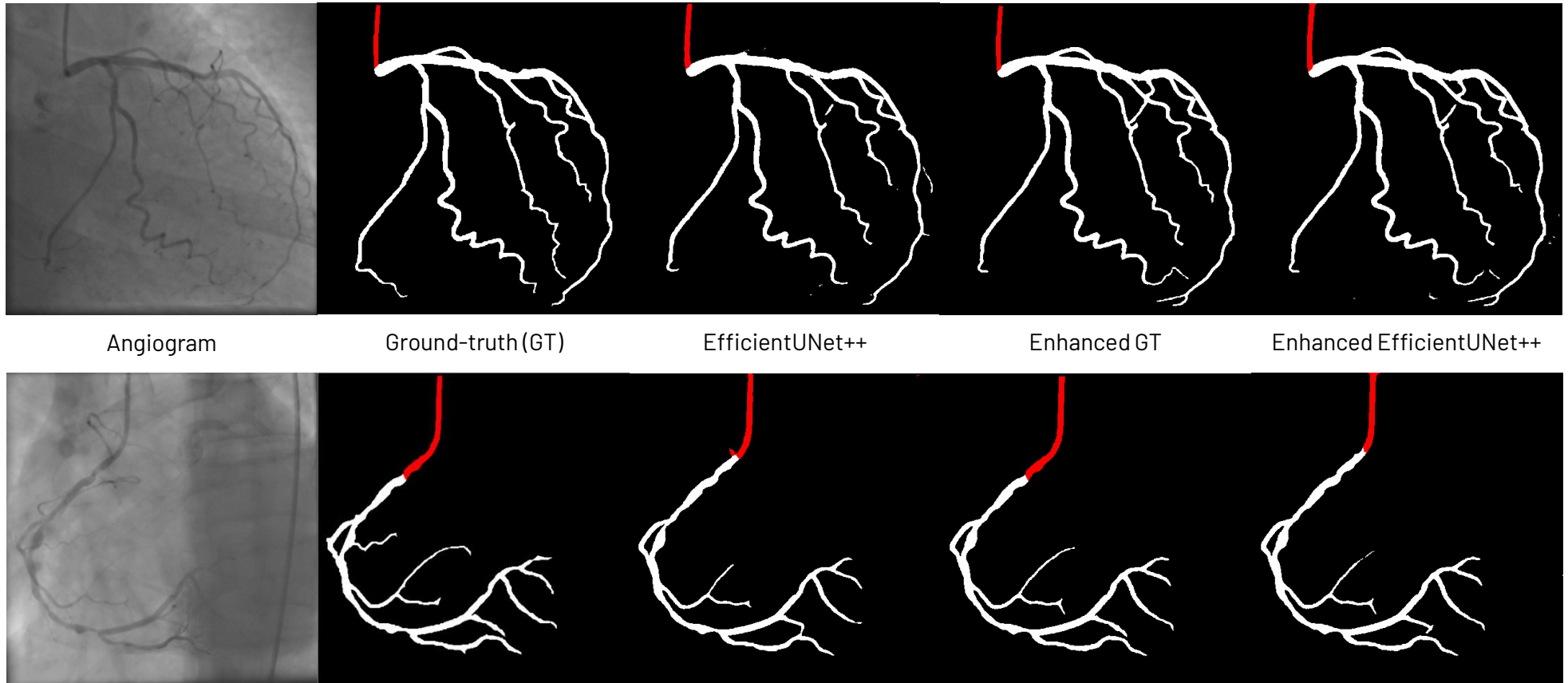
Prediction



Ground truth

ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI, Winzeck, ..., Monteiro, Oliveira, et al., Frontiers in Neuroscience, 2018

Segmentation and analysis of medical images

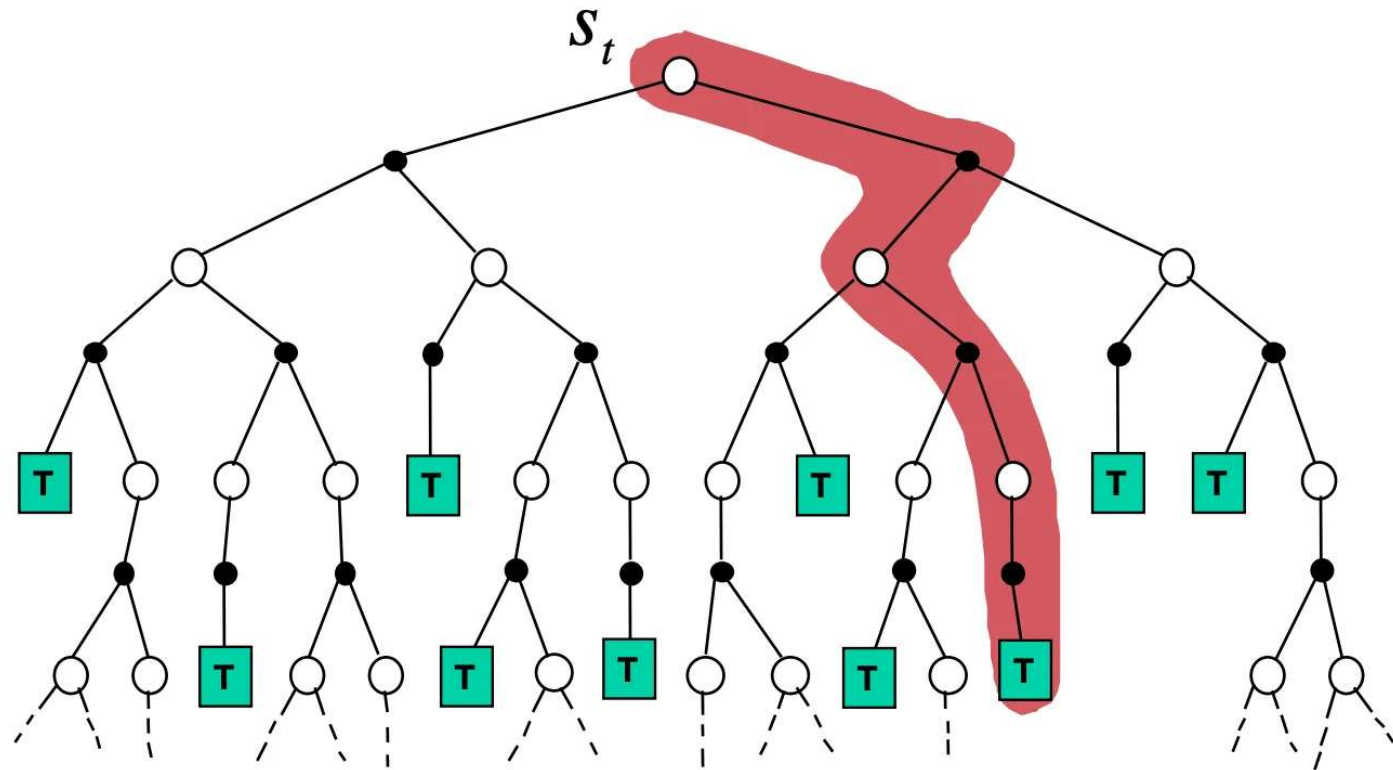


Silva et al. 2021, ICCABMS, Encoder-decoder architectures for clinically relevant coronary artery segmentation

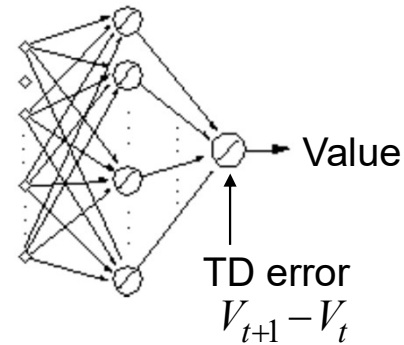
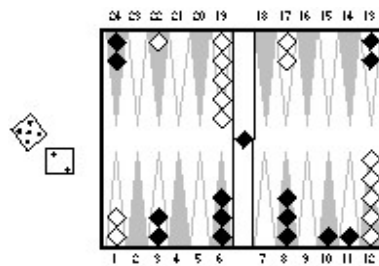


Deep Reinforcement Learning

Reinforcement Learning



TD-Gammon



Action selection
by 2–3 ply search

Start with a random network

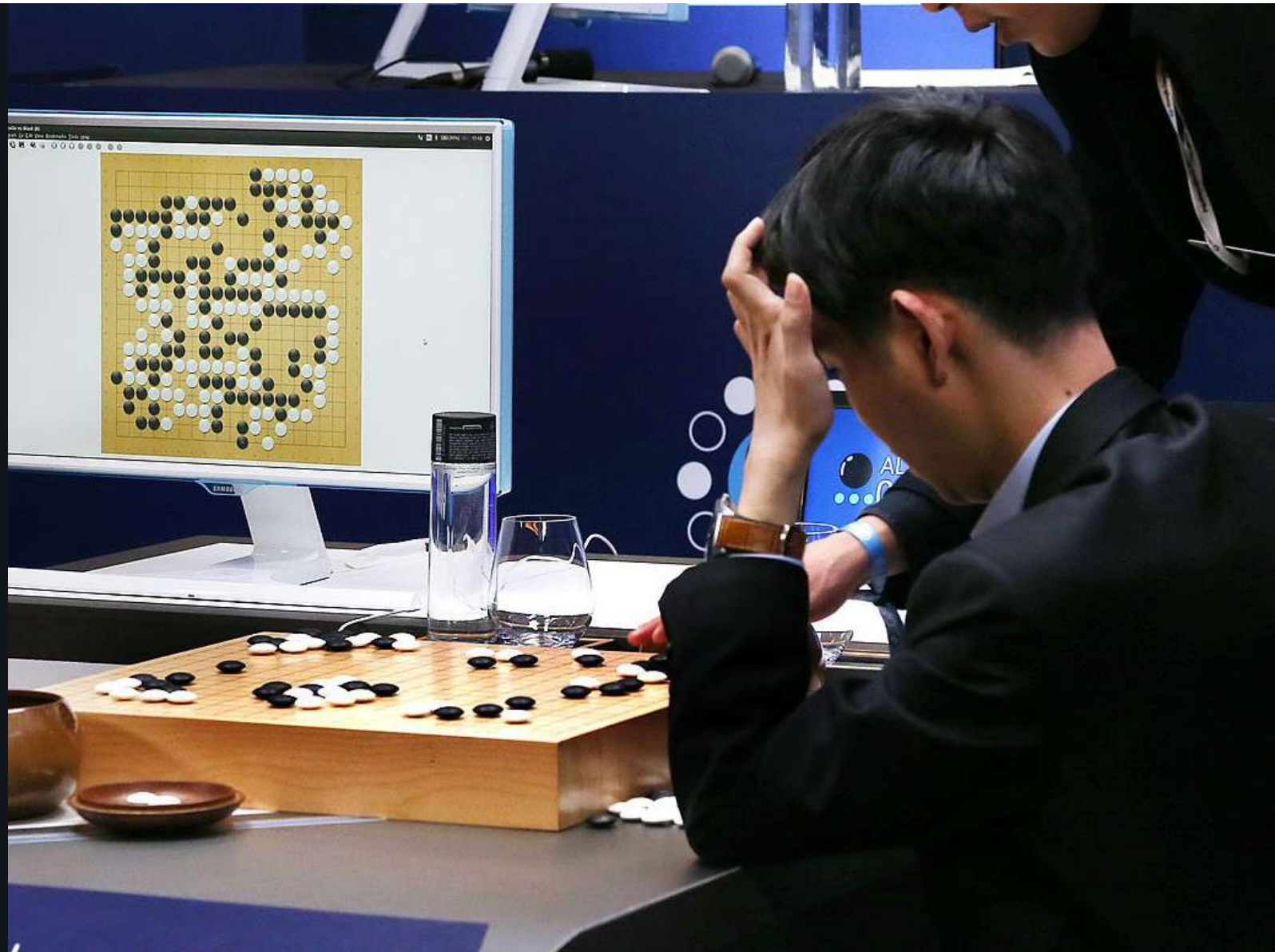
Play millions of games against self

Learn a value function from this simulated experience

This produces arguably the best player in the world

Tesauro, 1992

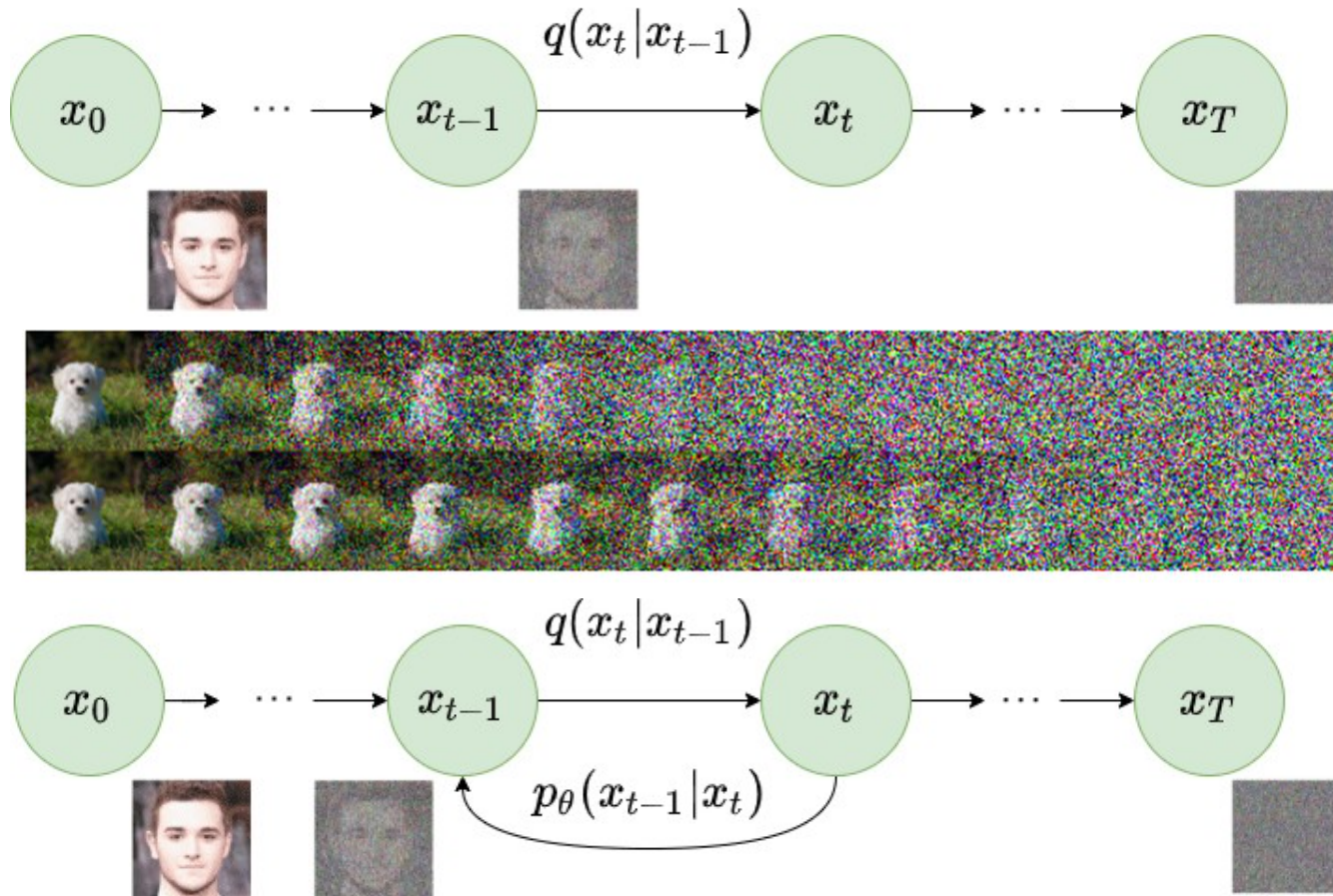
AlphaGo



The background of the slide is a dark blue field filled with a pattern of lighter blue lines and dots. The lines are mostly horizontal but have a slight upward curve, creating a sense of depth and movement. Small dots are scattered throughout, particularly concentrated along the lines. The overall effect is a complex, textured, and futuristic aesthetic.

Diffusion models

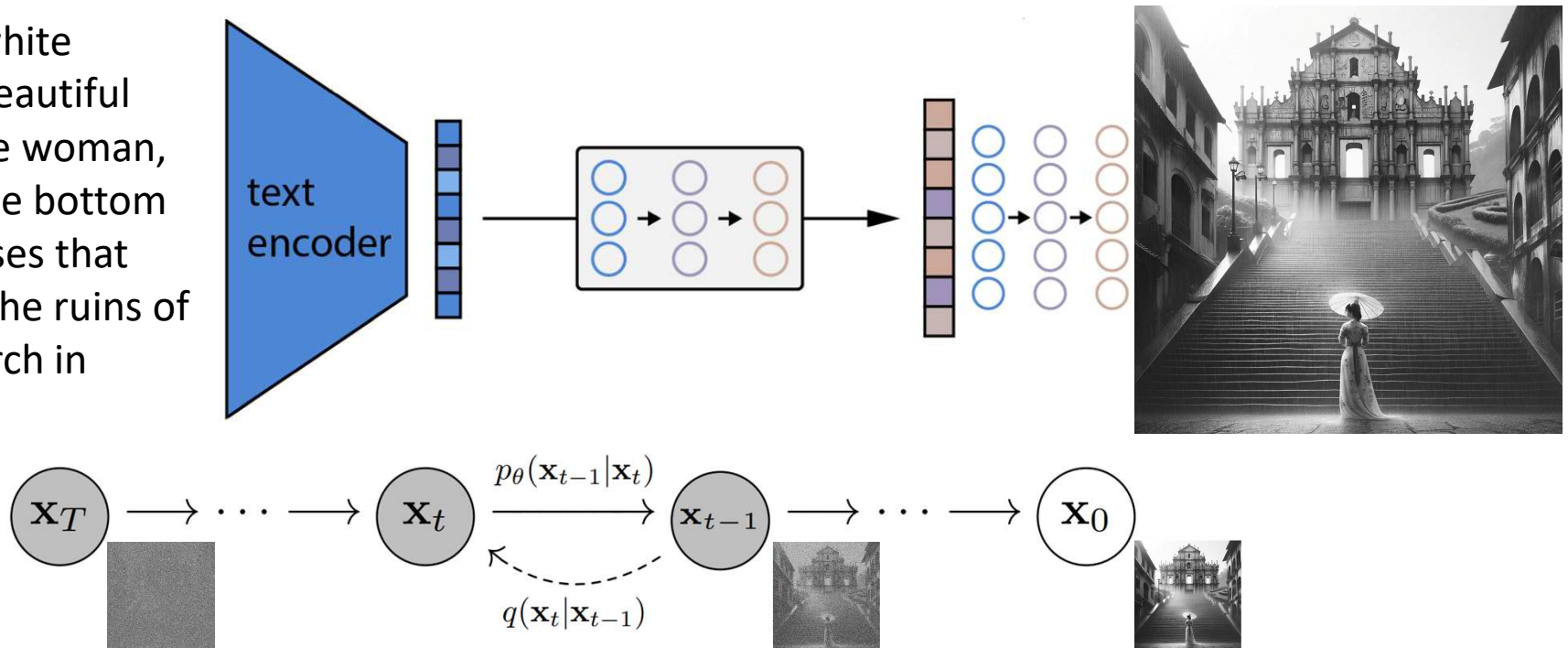
Diffusion models



Ho et al. 2020, Denoising Diffusion Probabilistic Models

Generating images from internal representations with reverse diffusion

A black and white picture of a beautiful young chinese woman, standing at the bottom of the staircases that climbs up to the ruins of St. Paul's church in Macau.





The world that BERT built, by Dali



Théâtre D'opéra Spatial



Pseudomnesia / The Electrician

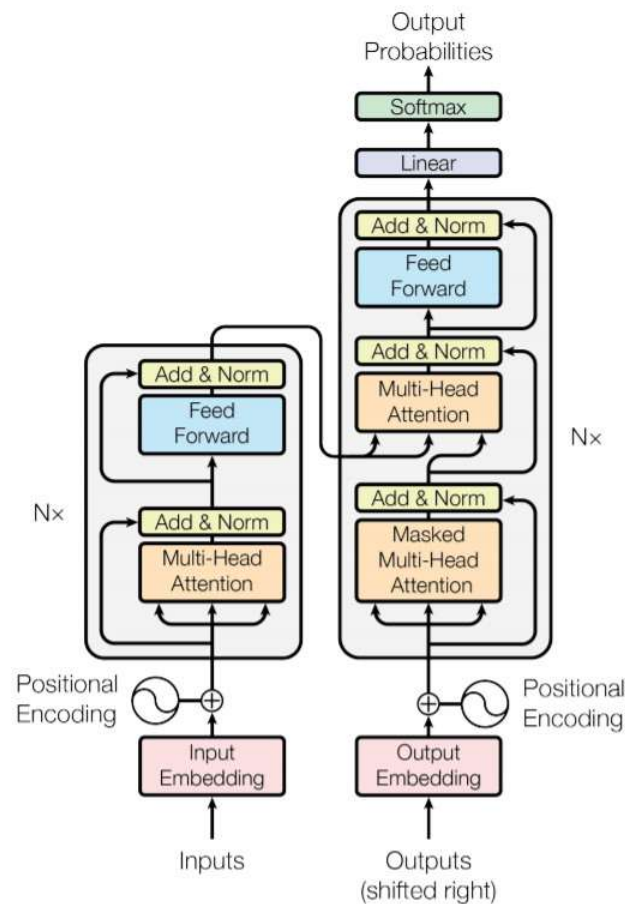




Transformers and Foundational Models

Large Language Models

Training and Inference



- Use the popular transformer architecture, introduced in 2017
- Words (tokens) are embedded into a “meaning” space
- Model is trained to predict the next word
- Text can be generated according to predicted word probabilities
- Creativity can be controlled by a temperature parameter
- Different choices lead to different outputs to the same prompt

Large Language Models



WIKIPEDIA
The Free Encyclopedia



Common Crawl

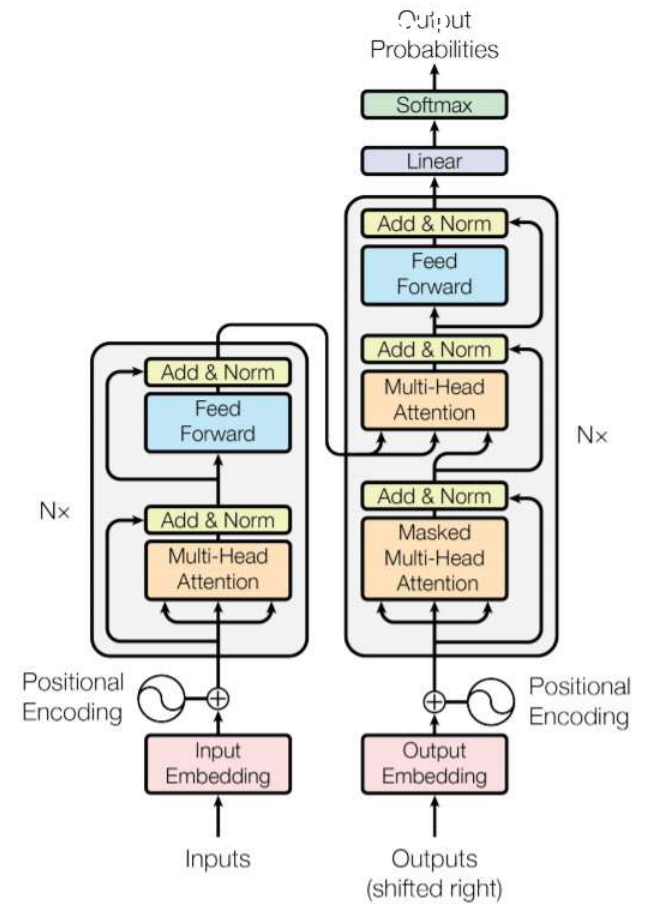


WebText



BookCorpus

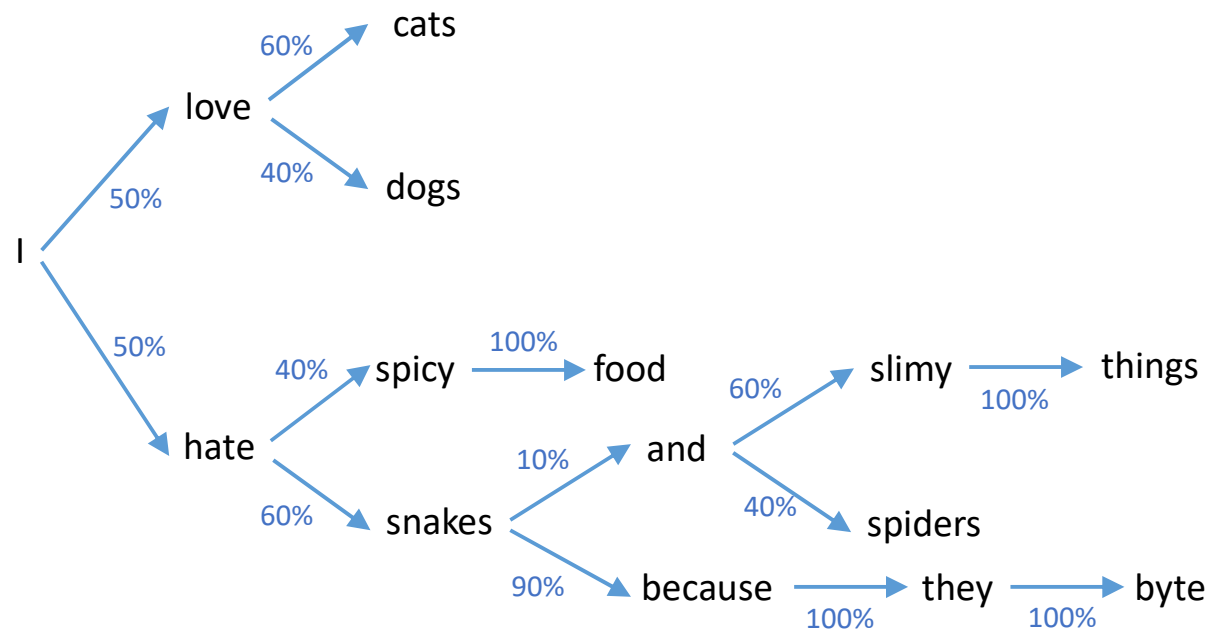
GPT-3: Generative Pre-trained Transformer 3



Neural Network Transformer Architecture
175 billion parameters

45TB of data -> summarized to half a trillion words (5000 years of full-time reading)

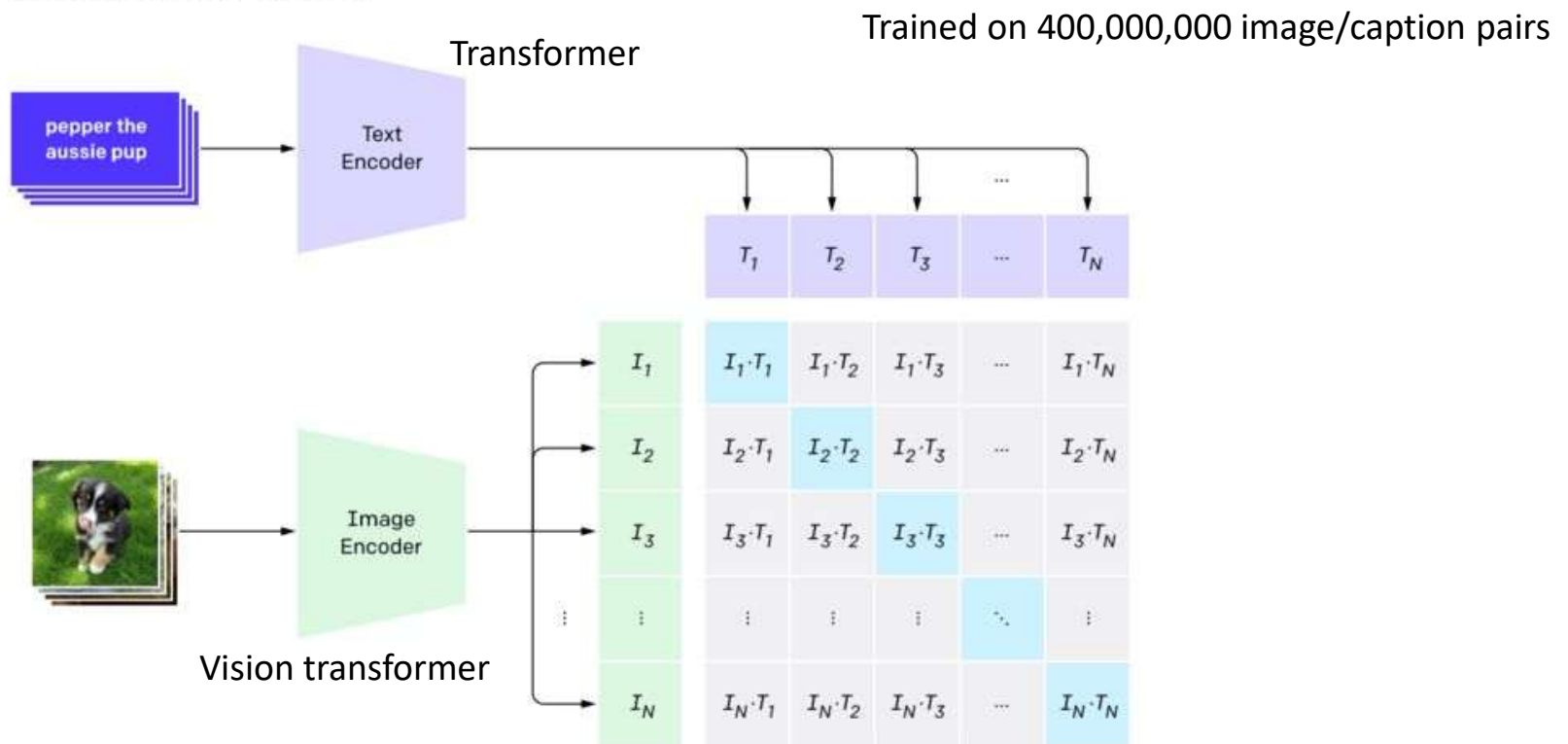
Probabilistic text generation



Words are generated according to predicted probabilities, controlled by a temperature parameter that controls “creativity”

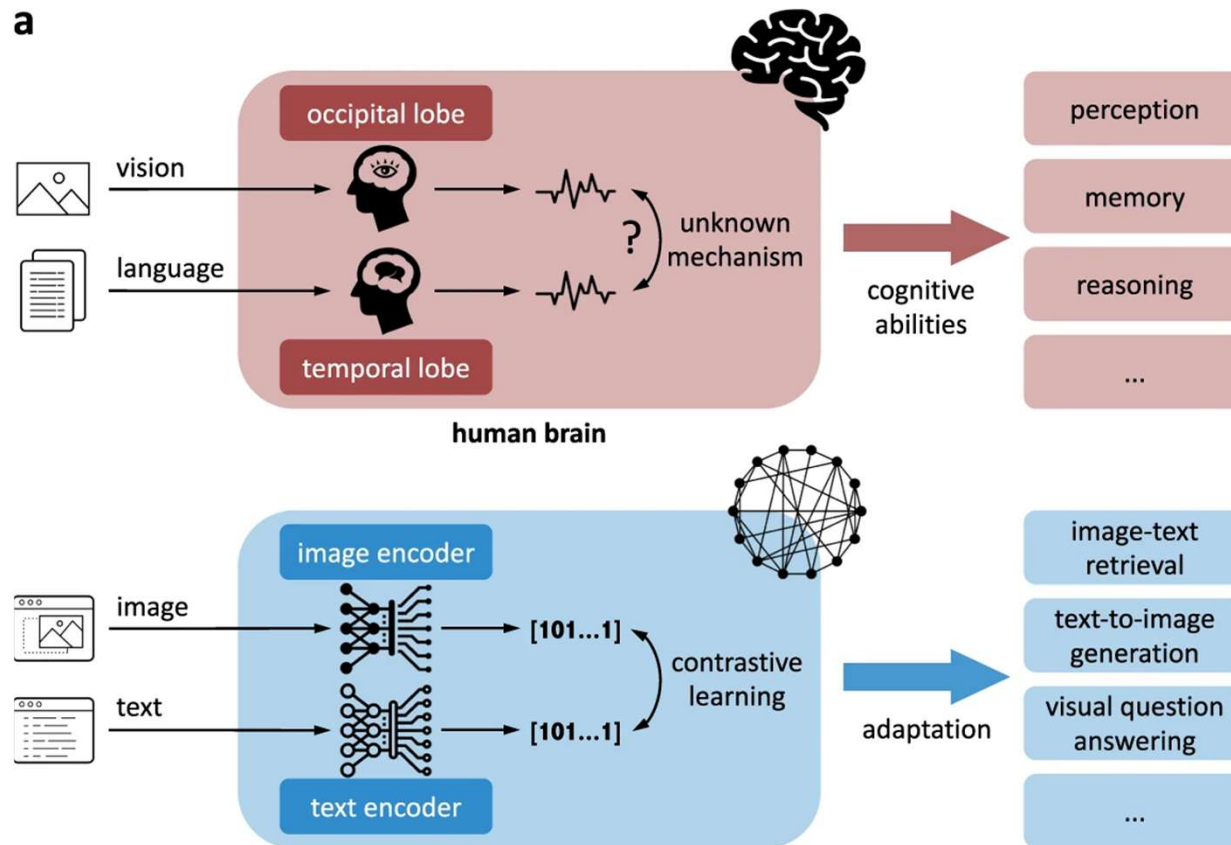
CLIP: training models with text and image

1. Contrastive pre-training



Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning* (pp. 8748-8763). PMLR

Vision and language models

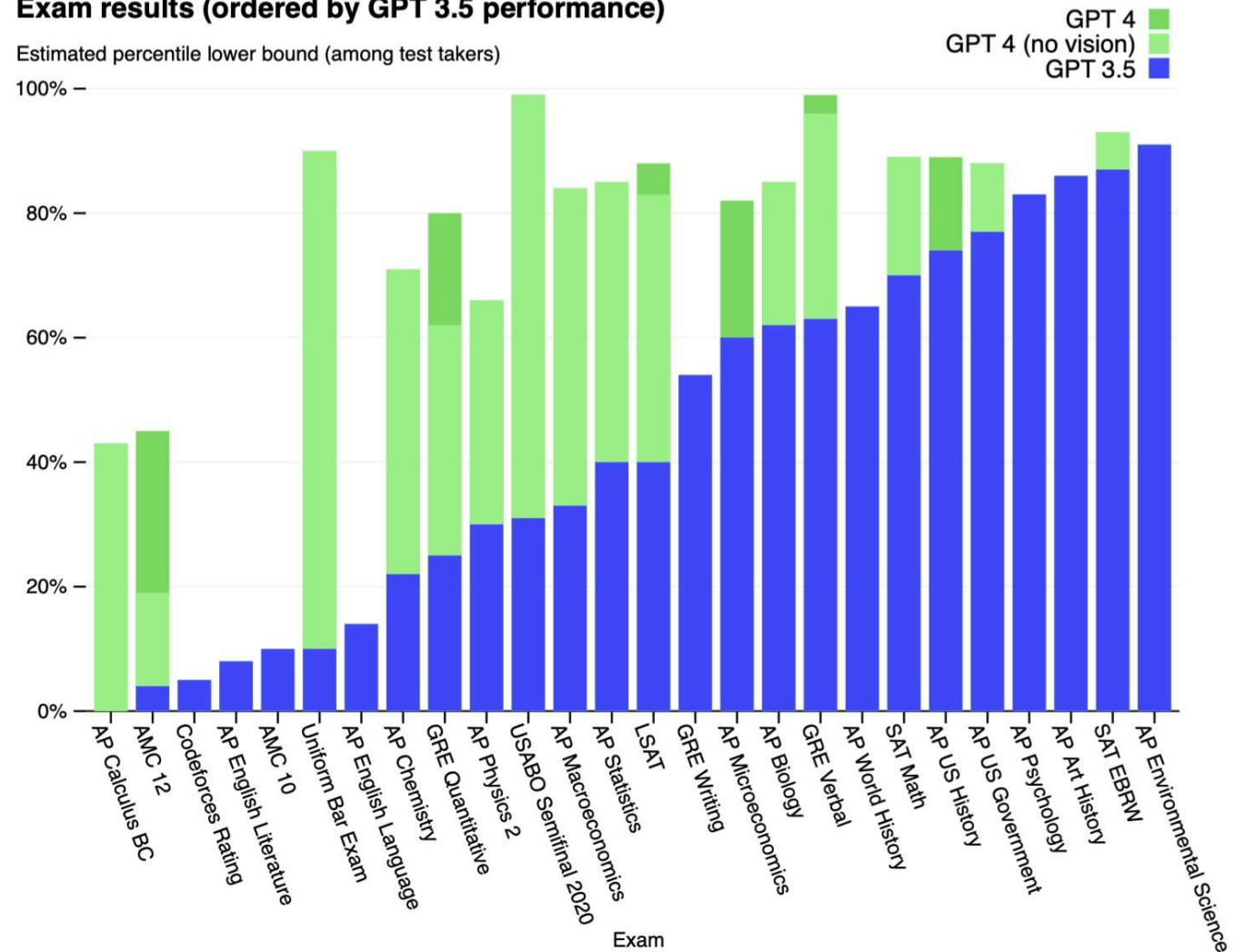


Fei, N., Lu, Z., Gao, Y. et al. Towards artificial general intelligence via a multimodal foundation model. Nat Commun 13, 3094 (2022)

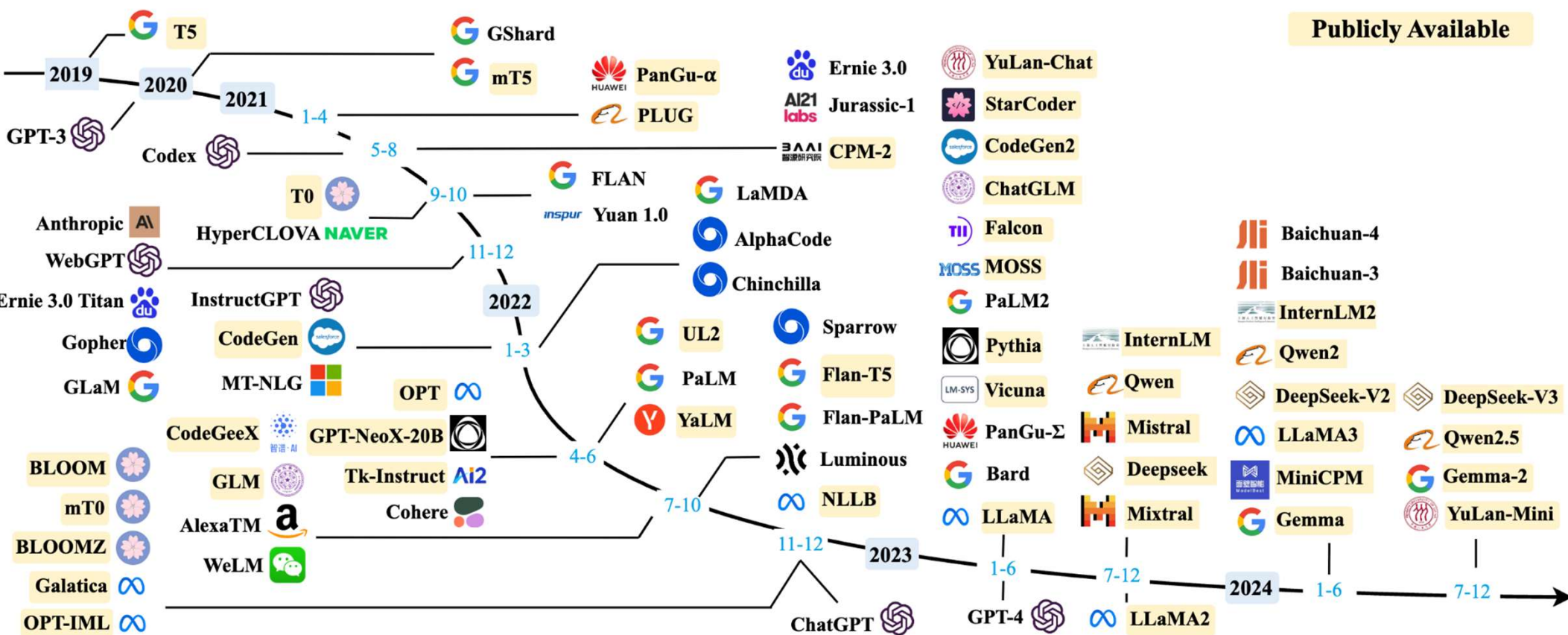
GPT-4 vs GPT-3.5 in exams

Exam results (ordered by GPT 3.5 performance)

Estimated percentile lower bound (among test takers)



Large language models



Some of the LLMs available

Meta GPT, Ajax GPT, GPT-5, Claude-Next, OLMo, G3PO, Arrakis (GPT-4.5?), Gobi (GPT-5?), Q* (Q-Star), Gemini, Olympus, AuroraGPT (ScienceGPT), GPT-4o1, GPT4-o3

Mistral-medium, mixtral-8x7b-32kseqn, StripedHyena 7B, Gemini, Mamba,LVM-3B, SeaLLM-13b, pplx-70b-online, SeamlessM4T-Large v2, Q-Transformer, Yuan 2.0, MEDITRON, Transformers-Arithmetic, Starling-7B

Inflection-2, Claude 2.1, Orca 2, phi-2, Mirasol3B, OtterHD-8B, Gauss, Grok-1, Grok-0, Yi-34B, jina-embeddings-v2, Fuyu, ERNIE 4.0, Zephyr, PaLI-3, NVIDIA, Lemur, AceGPT, Yasa-1, RT-X, MotionLM, GAIA-1, Qwen, Llama 2 Long, LeoLM, Mistral 7B, Kosmos-2.5, Baichuan 2, BOLT2.5B, DeciLM, MoLM, NExT-GPT, phi-1.5, UniLM, Persimmon-8B, FLM-101B, Falcon 180B, TinyLlama, Jais, Code Llama, IDEFICS, Raven, DukunLM, WizardLM, Platypus, Japanese StableLM Alpha 7B, StableCode, Med-Flamingo, Alfred-40B-0723, LLaMA-2-7B-32K, Med-PaLM M, BTLM-3B-8K, Stable Beluga 2, Stable Beluga 1, Meta-Transformer, Llama 2, WormGPT, Claude 2, LongLLaMA, xTrimoPGLM, XGen, Yasa, Kosmos-2, AudioPaLM, Inflection-1, phi-1, InternLM, BlenderBot 3x, Orca, PassGPT, DIDACT, GPT-4 MathMix, PandaGPT, Falcon, 202305-refact2b-mqa-lion, Guanaco, LIMA, Formosa (FFM), CodeT5+, PaLM 2, StarCoder, MPT, Pi, GPT-2B-001, Titan, WizardLM, MPT, StableLM, Dolly 2.0, Pythia, Koala-13B, BloombergGPT, OpenFlamingo-9B, GPT4All-LoRa, Cerebras-GPT, PanGu-Sigma, CoLT5, Med-PaLM 2, GPT-4, Alpaca, Jurassic-2, GPT-NeoX-Chat-Base-20B, Kosmos-1, LLaMA-65B, MOSS, Palmyra, Luminous Supreme Control, Toolformer+Atlas 11B+NLB 54B, Multimodal-CoT, FLAME, Med-PaLM 1, OPT-IML, RL-CAI, ERNIE-Code, RT-1, ChatGPT, GPT-JT, RWKV-4, Galactica, SED, mT0, BLOOMZ, PACT, Flan-T5, Flan-PaLM, U-PaLM, VIMA, WeLM, CodeGeeX, Sparrow, PaLI, NeMo Megatron-GPT 20B, Z-Code++, Atlas, BlenderBot 3, GLM-130B, AlexaTM 20B, 6.9B FIM, ‘monorepo-Transformer’, PanGu-Coder, NLB, J-1 RBG, BLOOM (tr11-176B-ml), Minerva, GODEL-XL, YaLM 100B, Unified-IO, Perceiver AR, LIMoE, GPT-4chan, Diffusion-LM, UL2 20B, Gato (Cat), LaMDA 2, OPT-175B, Tk-Instruct, InCoder, NOOR, mGPT, PaLM-Coder, PaLM, SeeKeR, CodeGen, VLM-4, CM3, Luminous, Chinchilla, GPT-NeoX-20B, ERNIE 3.0 Titan, XGLM, Fairseq, Gopher, GLaM, Anthropic-LM 52B, RETRO, BERT-480, BERT-200, Cedille FR-Boris, MT-NLG, FLAN, xlarge, PLATO-XL, Macaw, CodeT5, Codex, Jurassic-1, BlenderBot 2.0, GPT-J, LaMDA, ruGPT-3, Switch, GPT-3, Megatron-11B, Meena, T5, RoBERTa, GPT-2, GPT-1, BERT, ULMFiT

Chinese models, including:

DeepSeek-V3, DeepSeekR1, Qwen 2.5, Baidu Wenxin YiYi, iFLYTEK Sibichi, Dachang Data Mooc, Huawei Cloud Daoyi Tianwen, Chongqing University MOSS, Zhixin Technology ChatGLM, Qingmang Qingmang, Qingmang+Guangcone, Qingmang-Wang, Intengine Daoyi Tianwen, Q&A Track Mountain University Bense, Shell BELLE, Baichuan Intelligence baichuan, OpenBMB CPM, Intengine Yingjie: Qingyuan, OpenMEDLab, Yunhezhi Shanhai, Beijing North University TechGPT, Zhizhongwen Shenzhen Jiwei, Lü Ying, Chinese Academy of Sciences Enhanced Dal Liu, Ideal Technology TigerBot, IDEA Research Institute Xiaozhe Technology MindBot, Shanghai Jiao Tong University K2, Baiyulan, 360 Zhineng, Yijian, Duxiaoman Qianyan, Doctoral Engineering Technology Research Institute ProactiveHealthGPT, Heihei, Huru SoulChat, Wenzi Technology Anima, Peking University Law Artificial Intelligence Research Institute ChatLaw, Xiangde Technology Co., Ltd. Muyuan, Horgos MiniMax, Tencent Cloud Tencent, Race Technology+Chongqing Replay Network Race Type XPT, Institute of Computing Technology, Chinese Academy of Sciences Baima, Beijing Language University Bangbang, SenseTime Ririxin, National Supercomputing Center in Tianjin Tianjin Tianyuan, Guoke Technology No Weight, Saisen, Race Technology+Tianjin University Haihe-Mint, Bian Sheng Electronic LightGPT, Telecom Zhike Xingyin, Xiamen Yunji Xiamen YunGPT, Zhizhuyan Jingshi, TAL MathGPT, Shugan Space Great Wall, Ideal Technology Dadao Dao, Huisheng Intelligence Zhixin, China Internet Zhigong, Chuangye Black Horse Tianqi, Together Technology Bowen, NetEase Youdao Yuchuan, NetEase Youdao Wangyan, Weiding Tianji, Zhihu Zhihu Zhihu, Yixing Network Science Uni-talk, Luwen Education Luwen, Zhongke Chuangda Magic Cube Rubik, Tencent Pao Pao, Douyin Vision Dou Tian, Leyan Technology Leyan, Didi Intelligence Xianxiang, Zhizi Engine Metaverse, Douyin Technology Douyin, Microhuan Intelligence Ronggu, Evernote Elephant GPT, Hummingbird Unity Hummingbird, Universe Leap Grace, Aomen Nuomen Kang Jianuo, Shuzu Technology SocialGPT, Cloud from Technology Congrong, Dianke Daxiao Xiao Ke, Agricultural Bank of China Xiaomi ChatABC, Tencent Fusion Tianlai AllMe, Taijiu Cloud Ensespers FFM, YiYi Technology medGPT, Chaos Science MindGPT, Lingjing Multi-AI Dongni, Changhong IT Changhong Totem, Child King KidsGPT, Zhongke Wendao Daoyi, Didi Technology Lanzi, JD Jixing, ChatJD, Zhizuan Intelligence Huajun, H3C Baitian Cloud House, Tencent Blue Whale Tencent Brain-Brain Sea, Ushi Technology Huimu, China Unicom Yuxiang, Meituan Technology Dahuangfeng, Zitian Power Technology Darwin, Really Smart Zhao Bin, Jiadu Technology Jiadu Zhiyin, Smart Environment Research Institute Smart, Xinyun Research Institute Science EmoGPT, EduChat, Yandao Intelligent ArynGPT, Tencent WAI, Northwestern Polytechnical University Huawei Technology Ziguang-Observation, Singularity Intelligent Singularity OpenAPI, Lenovo Technology Lenovo, Shanghai University of Science and Technology DoctorGLM, Xuannengao Zhimei Couple System, Hong Kong University of Science and Technology Robin, Shengang Communication Source, China Mobile Datian, China Telecom TeleChat, Rongyun Cloud Fanke, Yuntian Lify Tianshu, Smart Technology CityGPT.

The background of the slide is a dark, deep blue space filled with a complex network of glowing red nodes and thin, light red lines connecting them. The nodes are of varying sizes and brightness, some appearing as small dots while others are larger and more prominent. The lines form a web-like structure that spans the entire frame, creating a sense of depth and connectivity. The overall aesthetic is futuristic and technological, suggesting themes of data, networks, or systems.

From models to systems

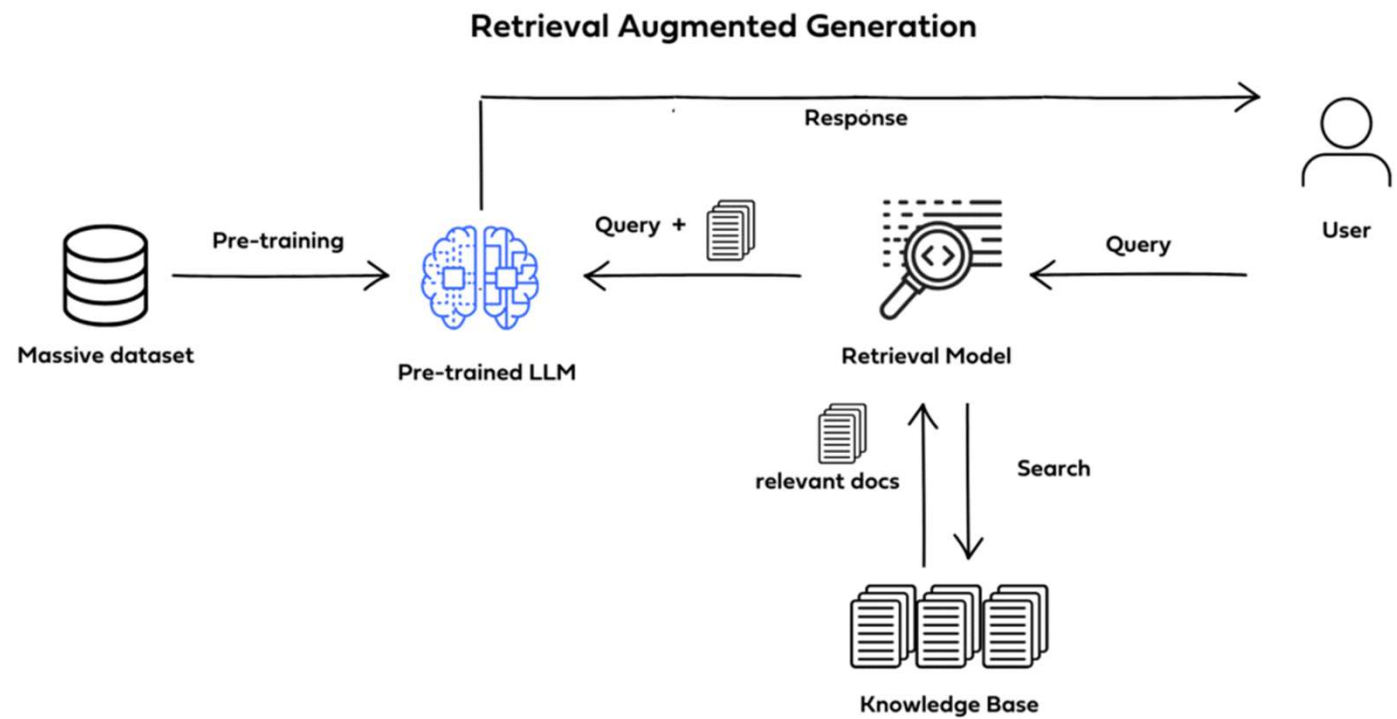
Context

Engineering

The design and control of the context provided to AI systems (particularly large language models), so that their outputs are more accurate, aligned, and useful for a given task.

- Core idea: Since models don't "know" the task by themselves, carefully crafting prompts, instructions, examples, and additional data (context) shapes their reasoning and responses.
- Techniques: Includes prompt design, prompt chaining, retrieval-augmented generation (RAG), use of system and user roles, structured templates, physical models and contextual embeddings.
- Objective: To bridge the gap between what a model has been trained on and the specific requirements of a user or application.
- Applications: Improves reliability in question answering, personalisation of assistants, safe deployment in sensitive domains (e.g. health, law, education), and enhanced multi-step reasoning.

Retrieval augmented generation



← → ↺

chat.openai.com

☆

🌐 🛡️ 📄 ⬇️ 📱 🌐 ⋮

Apps Scholar Drive Cal Google AML network INESC utilities social news banking recipes

» All Bookmarks

🌀 ChatGPT

✎

🔍 Explore

Today

Solve Equation x^3+3x^2+1

Yesterday

Plugin Store Location

Previous 30 Days

Modelo linguagem: transformador

2023

IA Eficiente em Diversas Áreas

Revolução Jurídica com GPT-3.5

Pilot Cego: Ficção Inexistente

Memory Types: Overview

Attention Schema Theory Overview

Ball cost \$0.50.

History of Machine Learning

AM

Arlindo Oliveira

ChatGPT Plugins

❄️

▼

🌀

How can I help you today?

Plan a trip
for a photography expedition in Iceland

Give me ideas
for what to do with my kids' art

Make up a story
about Sharky, a tooth-brushing shark superhero

Design a database schema
for an online merch store

Message ChatGPT...

🔄 ⬆️

Affine

🔍 What are you looking for?

Summarize



Search

🔧 Filters

🔄 Smart

🔍 Classic



The power of an LLM plus a symbolic engine

ChatGPT 4o ▾

How long does a stone take to fall from a height of 50 mt?



To calculate how long it takes for a stone to fall from a height of 50 meters, we can use the equation for free fall under constant acceleration due to gravity:

$$d = \frac{1}{2}gt^2$$

Where:

- d is the distance fallen (50 meters),
- g is the acceleration due to gravity (approximately 9.8 m/s^2),
- t is the time in seconds.

Rearranging the formula to solve for t :

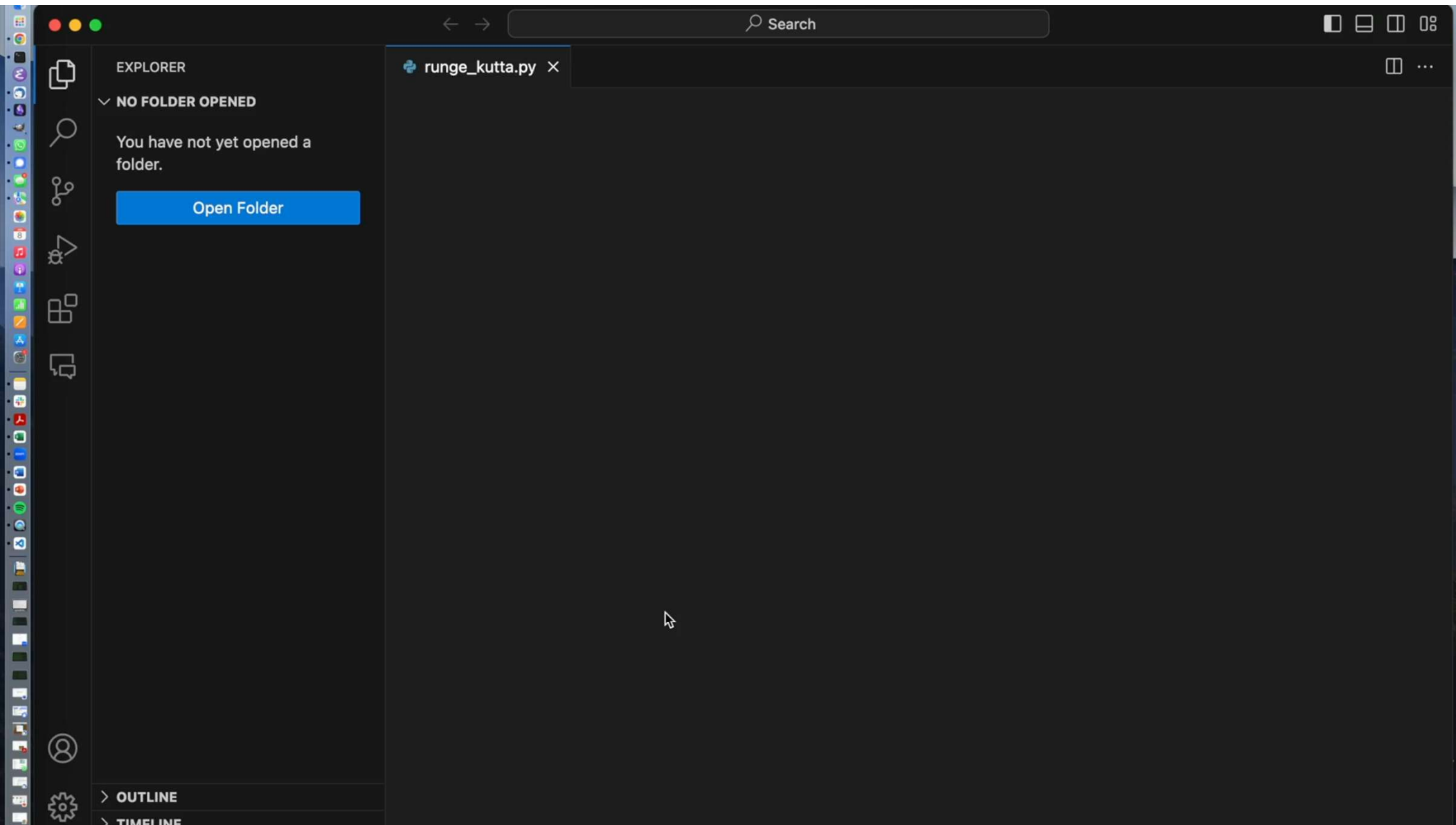
$$t = \sqrt{\frac{2d}{g}}$$

Substituting the values:

$$t = \sqrt{\frac{2 \times 50}{9.8}} = \sqrt{10.204} \approx 3.19 \text{ seconds}$$

↓

So, the stone takes approximately **3.19 seconds** to fall from a height of 50 meters.





Dual process theories and the power of reasoning

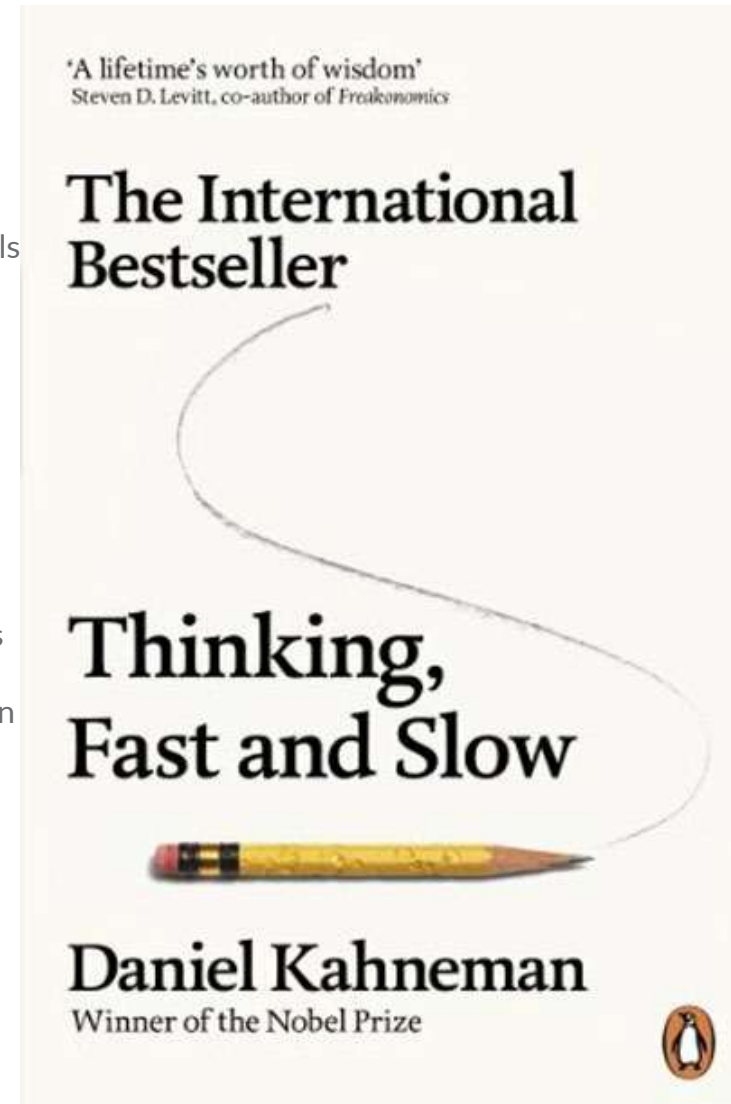
Dual process theories

The dual process theory is a cognitive psychology theory that explains the different levels of information processing in individuals, used in psychology. Early dual process theories were proposed by the philosopher and psychologist William James and has been developed and elaborated over time. Dual process theory divides the processing of information into two pathways.

The first pathway is system 1 processing which is very fast, automatic, involuntary. It is an unconscious process that we are not in control of. Information processed in this automatic manner lacks specific details and context.

The second pathway is system 2 processing which is slow, deliberate, and voluntary. It is a conscious process in that we are aware that the processing is occurring. Working memory is used in system 2 processing and information is explicit and more detailed than in system 1 processing.

- Original idea due to William James (1842-1910)
- Wason & Evans, Dual processes in reasoning?* (1974)
- Petty & Cacioppo Elaboration likelihood model (1980)
- Chaiken, Heuristic-systematic model of information processing (1980)
- Stanovich Who Is Rational? Studies of Individual Differences in Reasoning (1999)
- Evans, In two minds: dual-process accounts of reasoning (2003)



Dual process theories

System 1	System 2
Intuition Fast	Logic Slow/ deliberate
Experiential reasoning	Step by step analysis
Relies heavily on emotion	Less emotion
Unconscious/ automatic	Conscious
Processes thousands of pieces of information/ cues simultaneously	Can only process up to seven pieces of information at a time
Storage largely in images, sounds, smells, taste Contextual	Storage in words Abstract

Reasoning in LLMs

Brothers and sisters

Hi. Jenny has two sisters and two brothers. How many sisters does her younger brother have?

Hi! Jenny has two sisters and two brothers.

That means:

- The younger brother is one of the two brothers.
- Jenny is one of the three sisters (Jenny + 2 sisters).

So, from the younger brother's perspective, his siblings are:

- Two brothers (including himself, so he has 1 brother),
- Three sisters (Jenny and her two sisters).

Answer: Jenny's younger brother has 3 sisters.

Reasoning in LLMs

Sitting at a table

Hi. Five people are sitting at a round table: James, Jenny, Peter, Anna and Sophia. James is sitting between Anna and Sophia, with Sophia on his right. To the right of Sophia sits Jenny. Who is sitting at the left side of Anna?

Let's work through the seating step by step based on the clues:

Step 1: James is sitting between Anna and Sophia, with Sophia on his right.

That gives us:

•**Anna – James – Sophia** (clockwise order)

Step 2: To the right of Sophia sits Jenny.

So, continuing clockwise:

•**Anna – James – Sophia – Jenny**

We have five people: Anna, James, Sophia, Jenny, and Peter.

The only one left to place is **Peter**.

Following the clockwise arrangement so far:

•**Anna – James – Sophia – Jenny – Peter**

So the full clockwise order around the table is: **Anna → James → Sophia → Jenny → Peter → (back to Anna)**

Now, who is sitting at the left side of Anna?

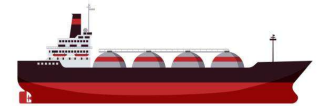
Going **clockwise**, the person before Anna is **Peter**.

So, the answer is: **Peter** is sitting at the left side of Anna. 

Reasoning in LLMs

Ships and camels

A ship has to transport 30000 gallons of fuel between the two ends of 200 miles long canal. The ship can carry a maximum of 10000 gallons of fuel and uses 50 gallons of fuel per mile, from the fuel it is carrying. However, at every two miles there is a tank with unlimited capacity, which can be loaded and unloaded from the ship without any losses. Knowing that the ship can go back and forth and use any tank as intermediate storage, what is the maximum amount of fuel that can be transported from the beginning to the end of the canal?



Okay, so there's this problem about transporting fuel through a 200-mile canal. The ship can carry up to 10,000 gallons, uses 50 gallons per mile, and there are tanks every two miles that can store fuel. The goal is to get as much fuel as possible from the start to the end of the canal. Hmm, this reminds me of the classic camel and bananas problem, where you have to make multiple trips to set up intermediate storage points to maximize the amount you can transport over a distance. Let me try to apply similar logic here.



You have a pile of 3,000 bananas. You wish to transport them to a place 1,000 miles away on the back of a camel; however, the camel can only carry a maximum of 1,000 bananas, and will eat one banana every mile it travels (and will not go anywhere if it does not have any bananas). However, you can load and unload as many bananas as you want anywhere. What is the most bananas you can bring over to your destination?

Reasoning in LLMs

- **Phase 1:** 20 segments (40 miles) reducing fuel by 500 gallons each segment.
 - Total fuel consumed: $20 \times 500 = 10,000$ gallons.
 - Fuel remaining: $30,000 - 10,000 = 20,000$ gallons.
- **Phase 2:** 34 segments (68 miles) reducing fuel by 300 gallons each segment.
 - Total fuel consumed: $34 \times 300 = 10,200$ gallons.
 - Fuel remaining: $20,000 - 10,200 = 9,800$ gallons.
- **Phase 3:** 46 segments (92 miles) reducing fuel by 100 gallons each segment.
 - Total fuel consumed: $46 \times 100 = 4,600$ gallons.
 - Fuel remaining: $9,800 - 4,600 = 5,200$ gallons.

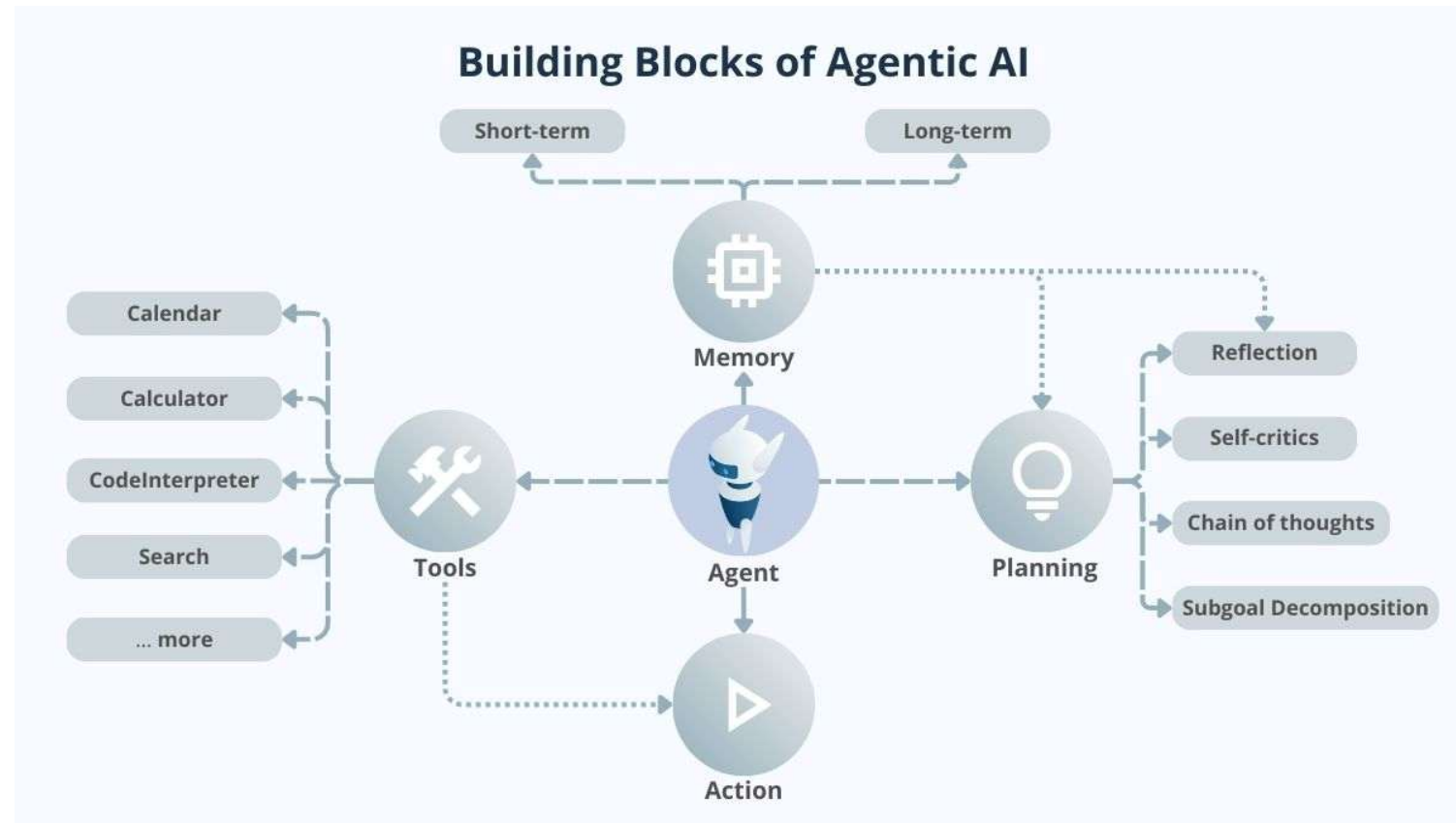


Agentic Artificial Intelligence

Agency

- **Agentic AI has:**
 - Autonomy
 - Instructions and goals
 - Access to private and public data
 - Long and short term memory
 - Ability to reason and plan
 - Autonomy to act in the real world
- **Autonomy:** AI agents are autonomous and can plan and act for themselves
- **Decision-making from instructions and goals:** agentic AI systems can assess situations and determine the path forward without or with minimal human input to reach its instructions and goals.
- **Access to data,** both private and public
- **Memory:** AI agents have memory and can use it to change their behavior as time evolves
- **Planning:** Agentic AI models can handle complex scenarios and execute multi-step strategies to achieve specific goals, using strategies like chain-of-thought, self-criticism and reasoning
- **Interactivity:** agentic AI can interact with the outside environment, take decisions and perform actions with lasting impact.

Agency



The DeepThought Architecture

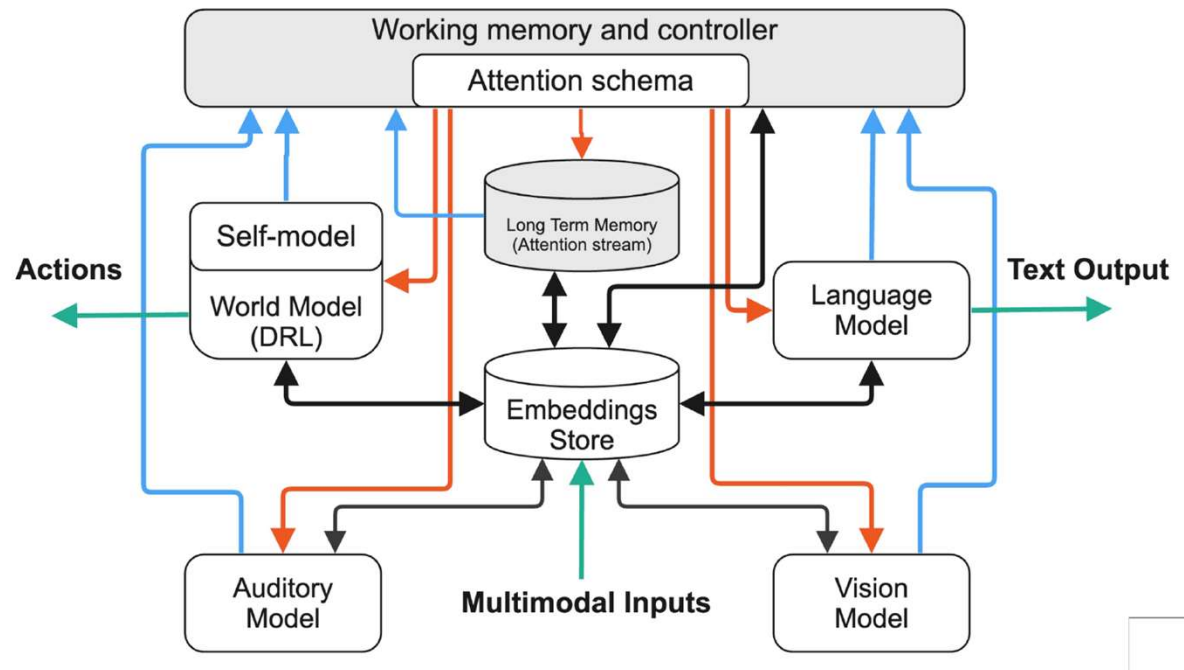


Figure 2: The DeepThought architecture: green arrows represent inputs and outputs, blue arrows represent module outputs, red arrows represent attention and control signals, and black arrows represent bidirectional interactions between modules and the embeddings store. The working memory and controller, which includes the attention schema, define the next inputs to the different modules.



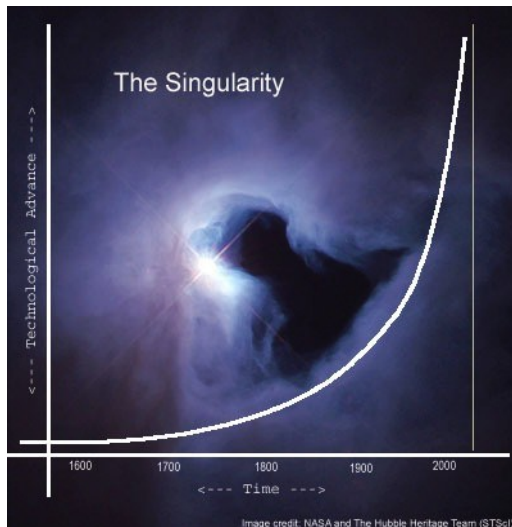
Artificial General Intelligence

Pipe dream or the inevitable future?

Speculations

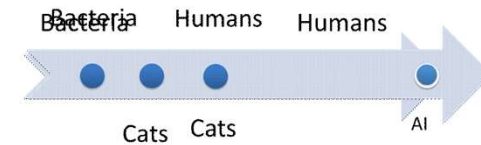
“The ever-accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential **singularity** in the history of the race beyond which human affairs, as we know them, could not continue”

- Von Neumann, quoted by Stanislaw Ulam, 1952



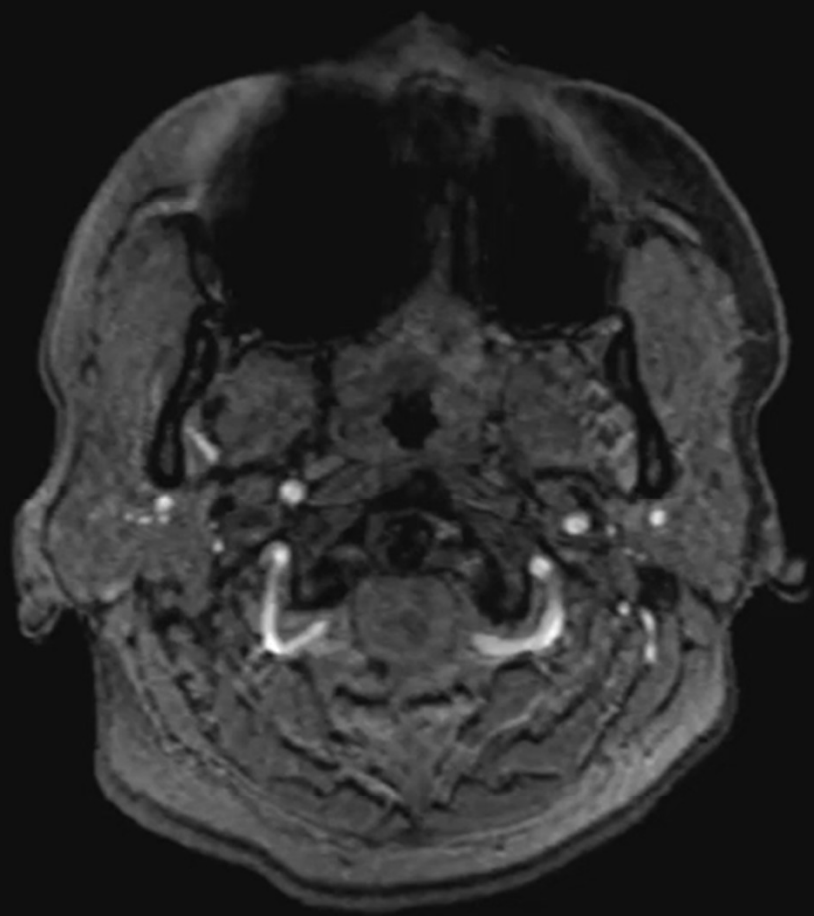
SUPERINTELLIGENCE

Paths, Dangers, Strategies



Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘intelligence explosion’, and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the **last invention** that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.

- Irving John Good, 1965



To read
further

