

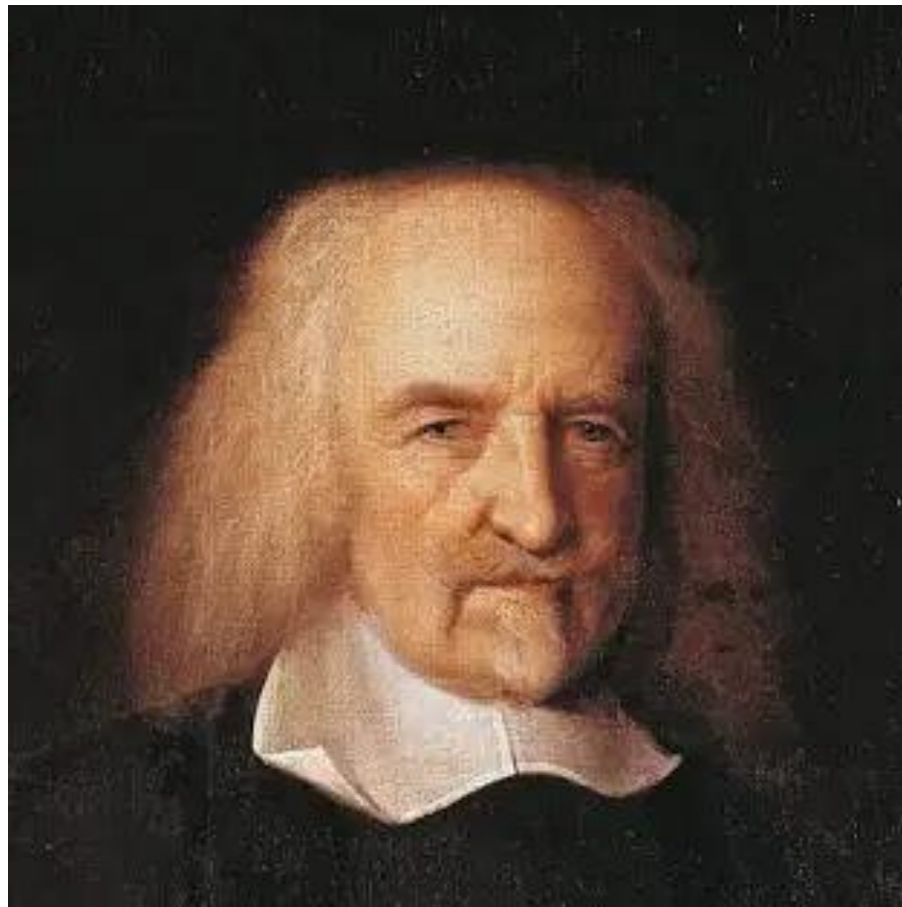
Artificial Intelligence

The Brave New World

Arlindo Oliveira

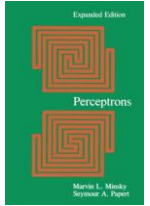
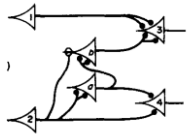


Thomas Hobbes



**For reason, in this sense, is nothing but reckoning (that is, adding and subtracting) of the consequences of general names agreed upon for the marking and signifying of our thoughts;
(Leviathan, 1651)**

Key Moments in the History of Artificial Intelligence



1943

McCulloch-Pitts Neuron Model

1950

Turing – Can machines think?

1952

Rosenblatt - Perceptron

1959

Samuel - Checkers

1967

Weizenbaum - ELIZA

1969



IBM DeepBlue



Rumelhart & McClelland PDP Book

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

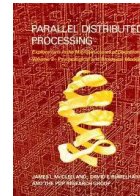
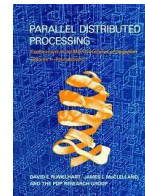
Fukushima Neocognitron

2006

1997



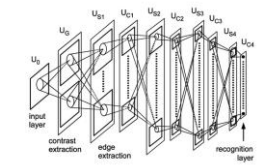
1987



1986



1979



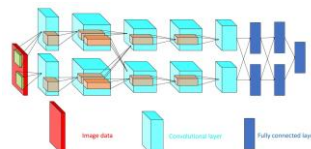
2009

Li Fei-Fei et. al - ImageNet



2012

Krizhevsky, Sutskever & Hinton AlexNET



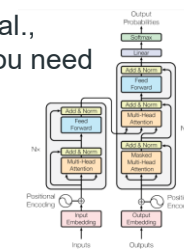
2016

Silver et al. AlphaGo



2017

Vaswani et al., Attention is all you need

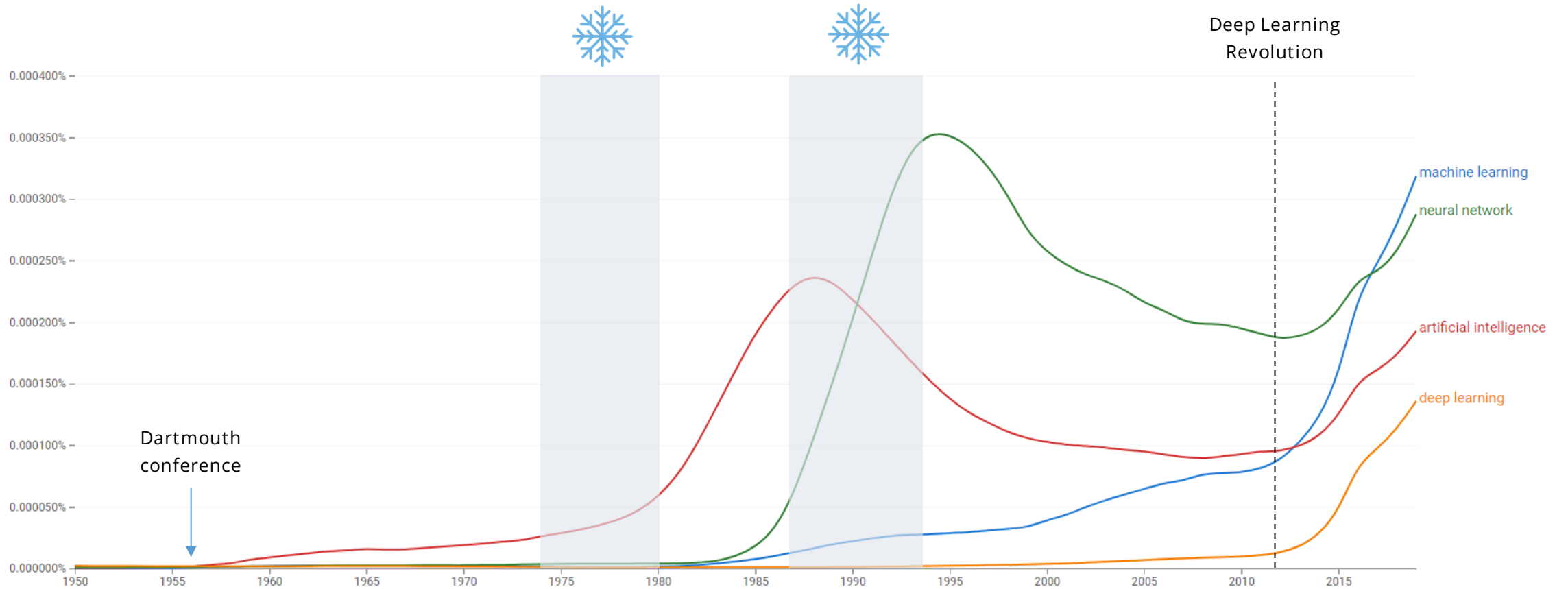


2020

Brown et al. GPT-3



70 years of history



source: [Google Books Ngram Viewer](#)

NEW NAVY DEVICE LEARNS BY DOING

Psychologist Shows Embryo
of Computer Designed to
Read and Grow Wiser

WASHINGTON, July 7 (UPI)—The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.

The embryo—the Weather Bureau's \$2,000,000 "704" computer—learned to differentiate between right and left after fifty attempts in the Navy's demonstration for newsmen.

The service said it would use this principle to build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of \$100,000.

Dr. Frank Rosenblatt, designer of the Perceptron, conducted the demonstration. He said the machine would be the first device to think as the human brain. As do human be-

ings, Perceptron will make mistakes at first, but will grow wiser as it gains experience, he said.

Dr. Rosenblatt, a research psychologist at the Cornell Aeronautical Laboratory, Buffalo, said Perceptrons might be fired to the planets as mechanical space explorers.

Without Human Controls

The Navy said the perceptron would be the first non-living mechanism "capable of receiving, recognizing and identifying its surroundings without any human training or control."

The "brain" is designed to remember images and information it has perceived itself. Ordinary computers remember only what is fed into them on punch cards or magnetic tape.

Later Perceptrons will be able to recognize people and call out their names and instantly translate speech in one language to speech or writing in another language, it was predicted.

Mr. Rosenblatt said in principle it would be possible to build brains that could reproduce themselves on an assembly line and which would be conscious of their existence.

1958 New York Times...

In today's demonstration, the "704" was fed two cards, one with squares marked on the left side and the other with squares on the right side.

Learns by Doing

In the first fifty trials, the machine made no distinction between them. It then started registering a "Q" for the left squares and "O" for the right squares.

Dr. Rosenblatt said he could explain why the machine learned only in highly technical terms. But he said the computer had undergone a "self-induced change in the wiring diagram."

The first Perceptron will have about 1,000 electronic "association cells" receiving electrical impulses from an eye-like scanning device with 400 photo-cells. The human brain has 10,000,000,000 responsive cells, including 100,000,000 connections with the eyes.

Moravec's Paradox

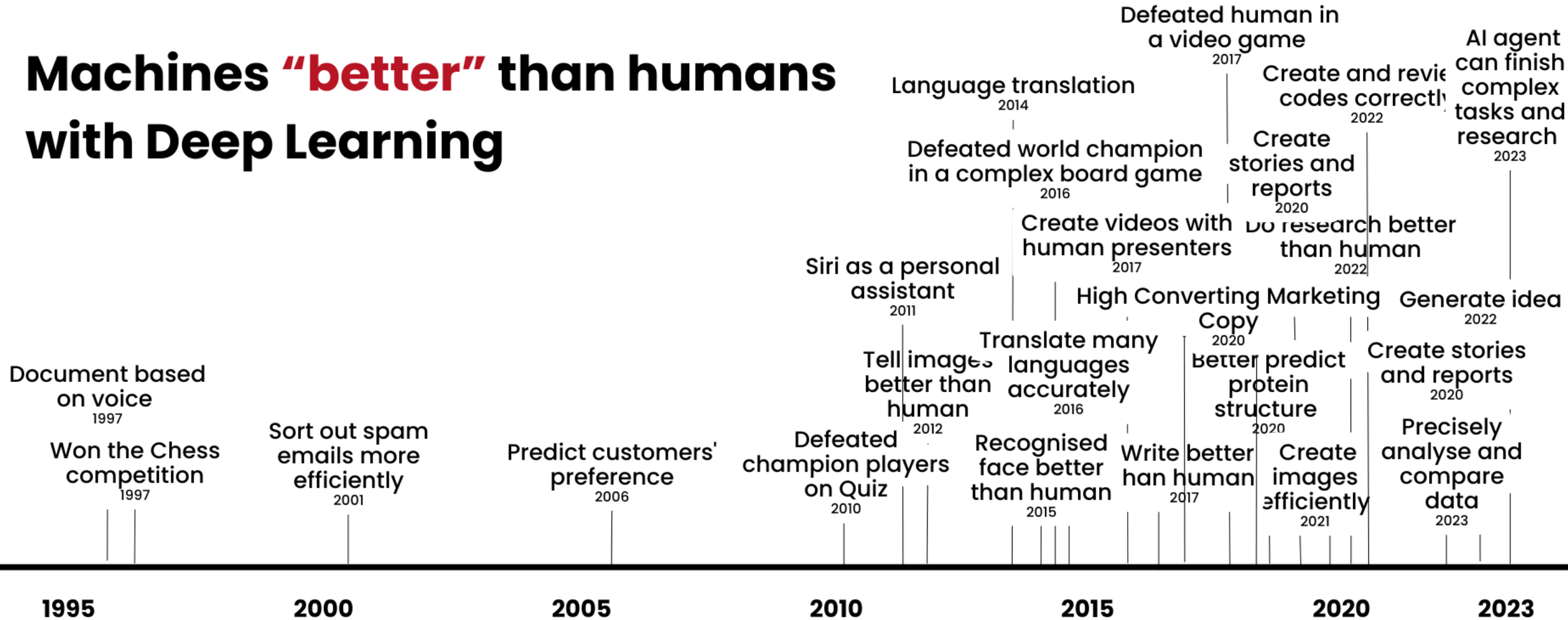


Robots and AI systems find the difficult things easy and the easy things difficult

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.

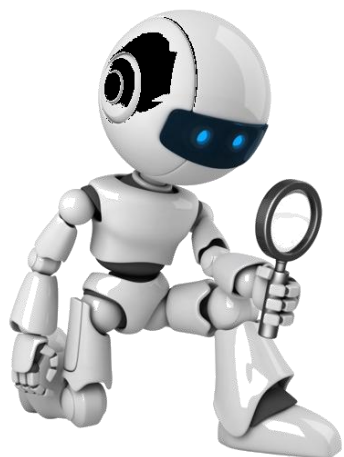
Machines “better” than humans with Deep Learning



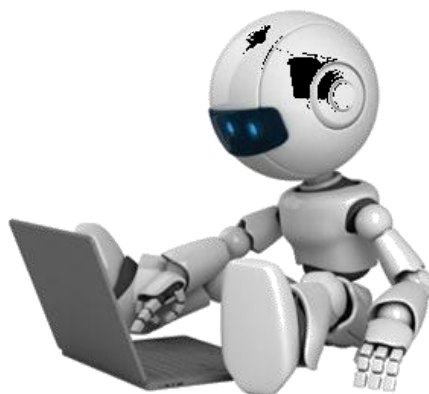
Deep Learning (2010s-now)

Machine Learning (1995 - Now)

Artificial Intelligence Sub-Fields



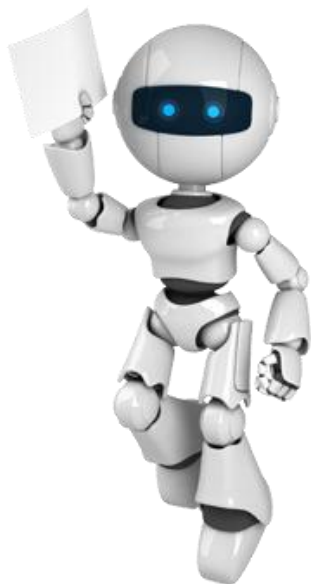
Search



Reasoning



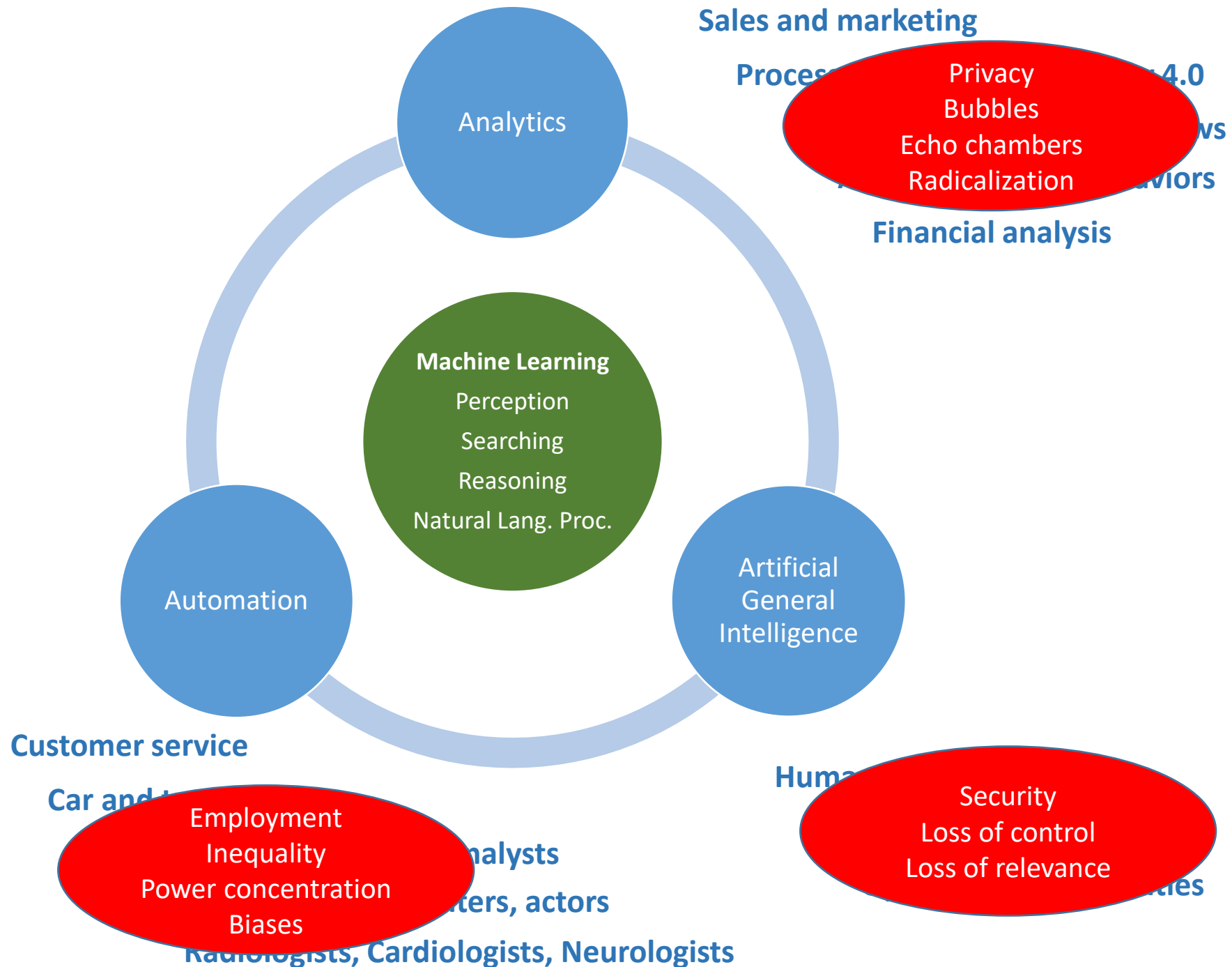
Natural Language Processing



Interaction and perception



Machine Learning

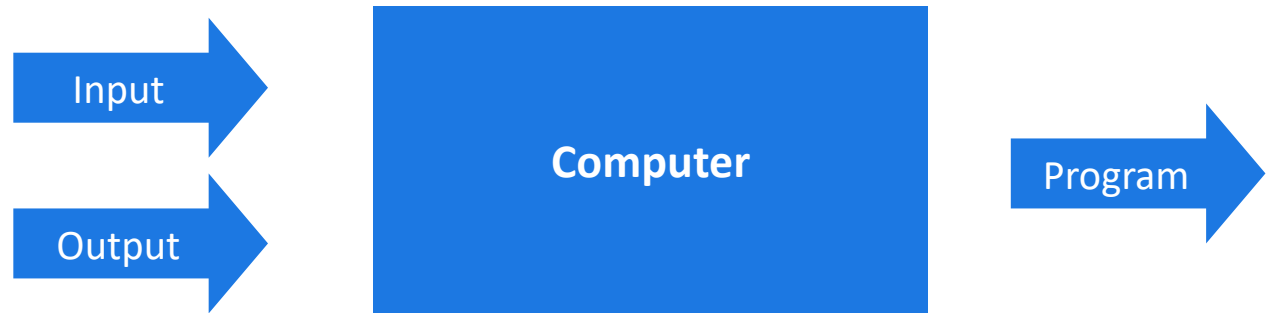


Machine Learning

Traditional programming



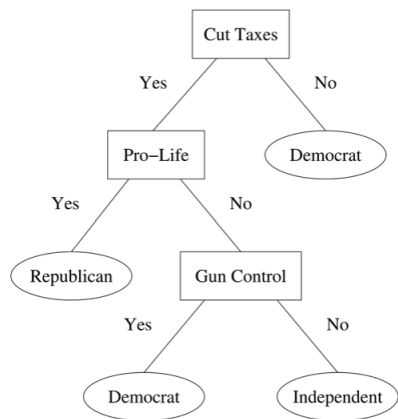
Machine Learning



Four Paradigms of Machine Learning

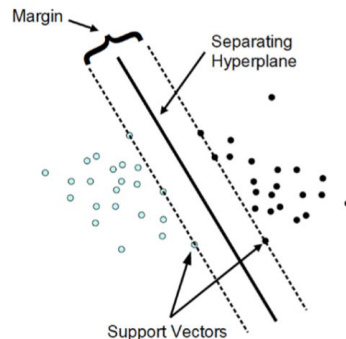
Symbolists

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



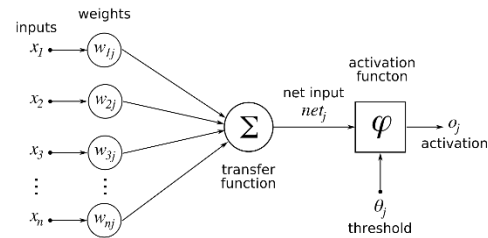
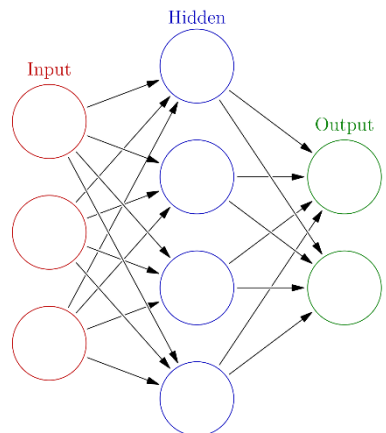
$$P(H|E) = \frac{P(H) * P(E|H)}{P(E)}$$

Annotations for the equation:

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

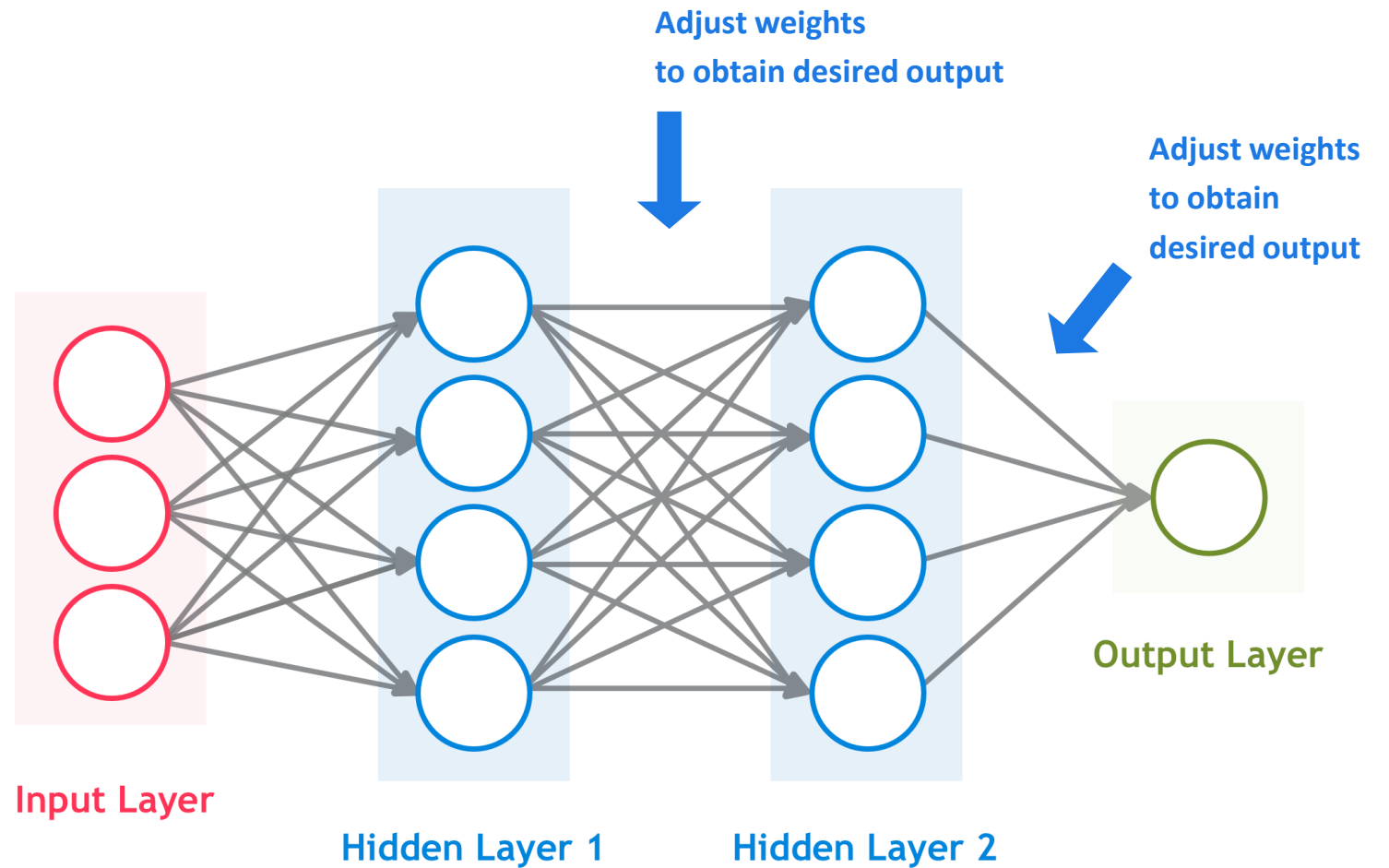
Connectionists

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



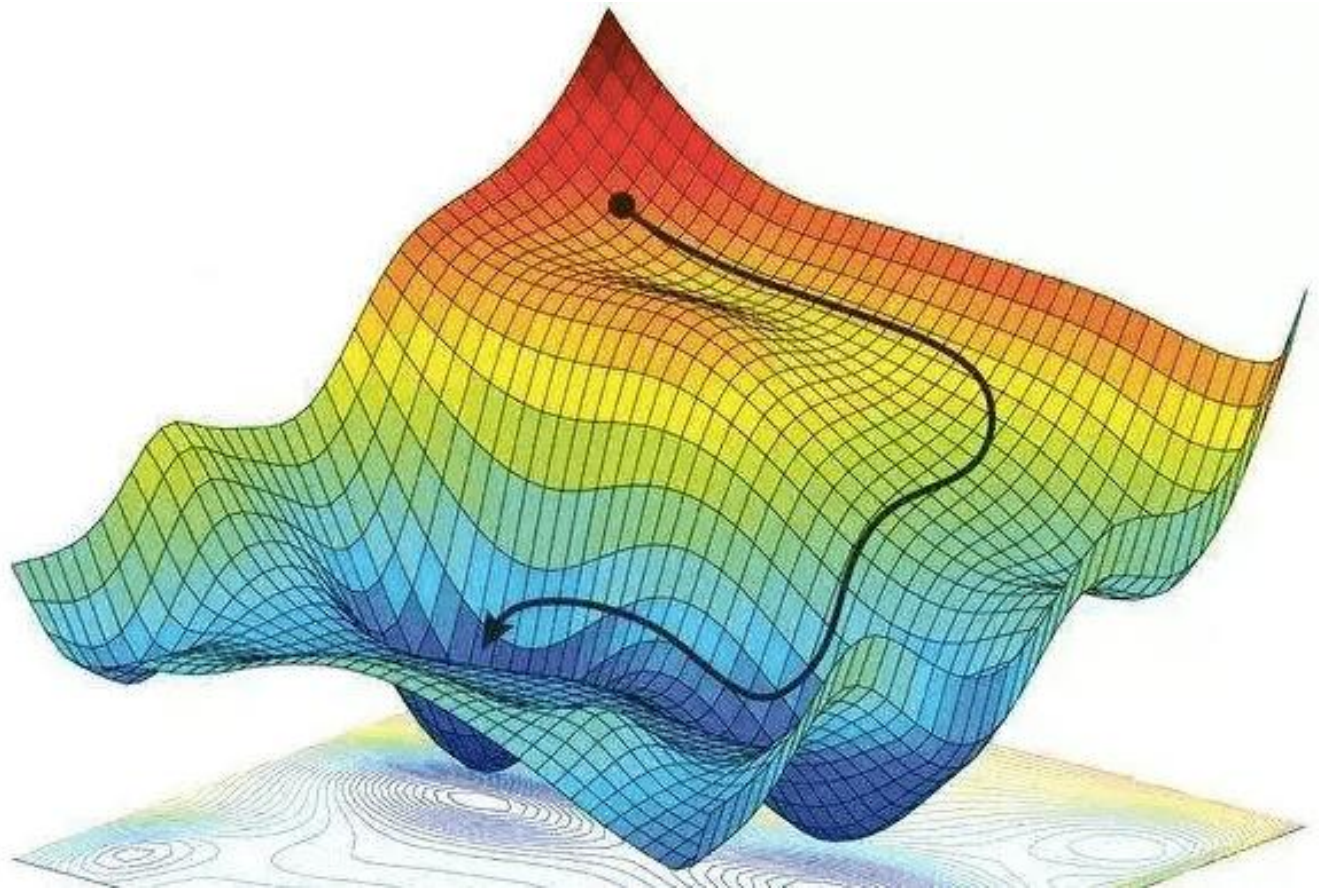
Neural Networks

Multi Layer
Perceptron



Gradient Descent

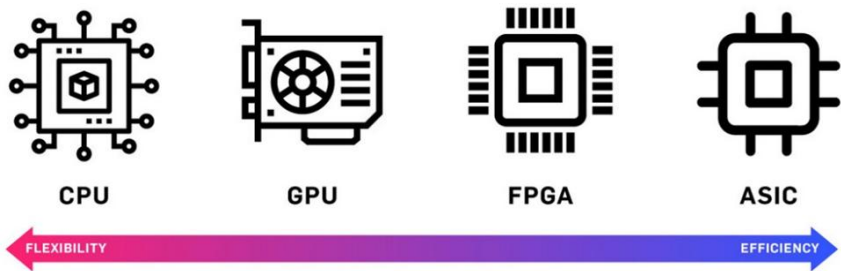
Error backpropagation



The Deep Learning Revolution

Frictionless reproducibility

Shared code
Hardware

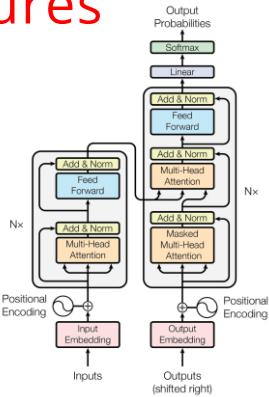
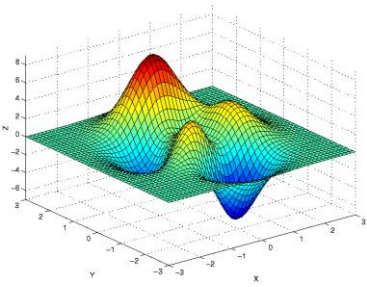


Deep Learning Revolution

Shared Data
Data



Competitive environment
Algorithms and architectures

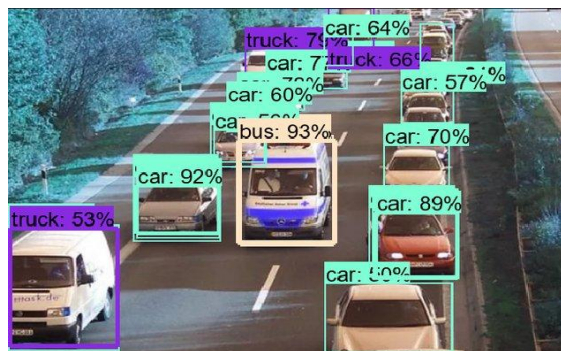


David Donoho

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

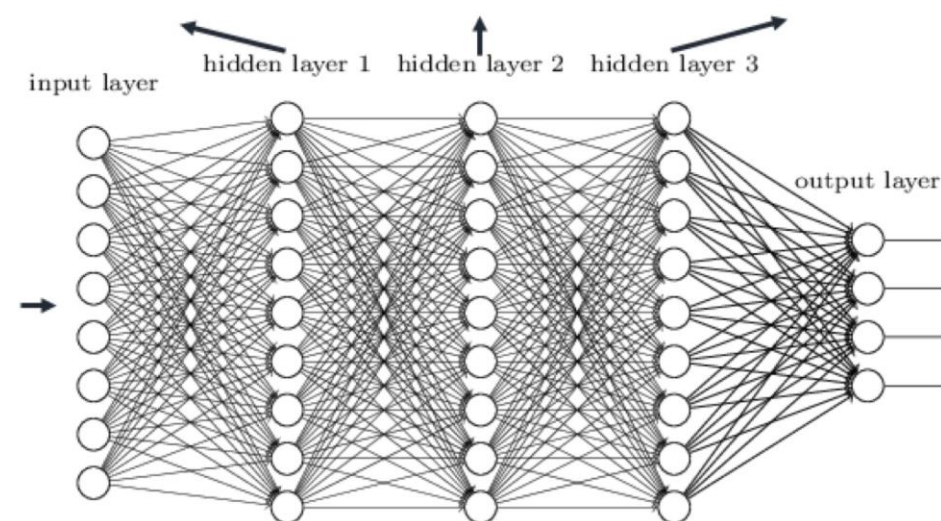
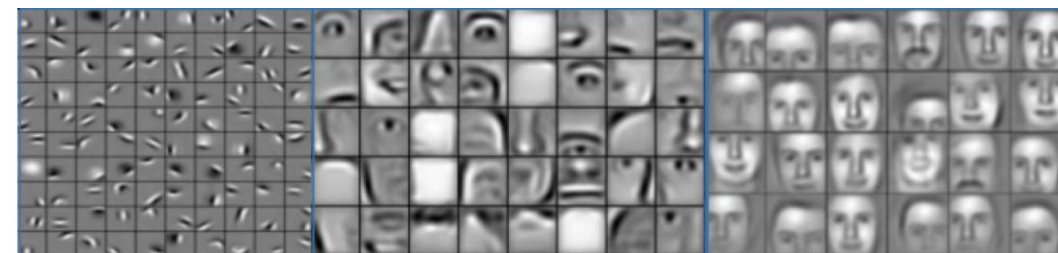




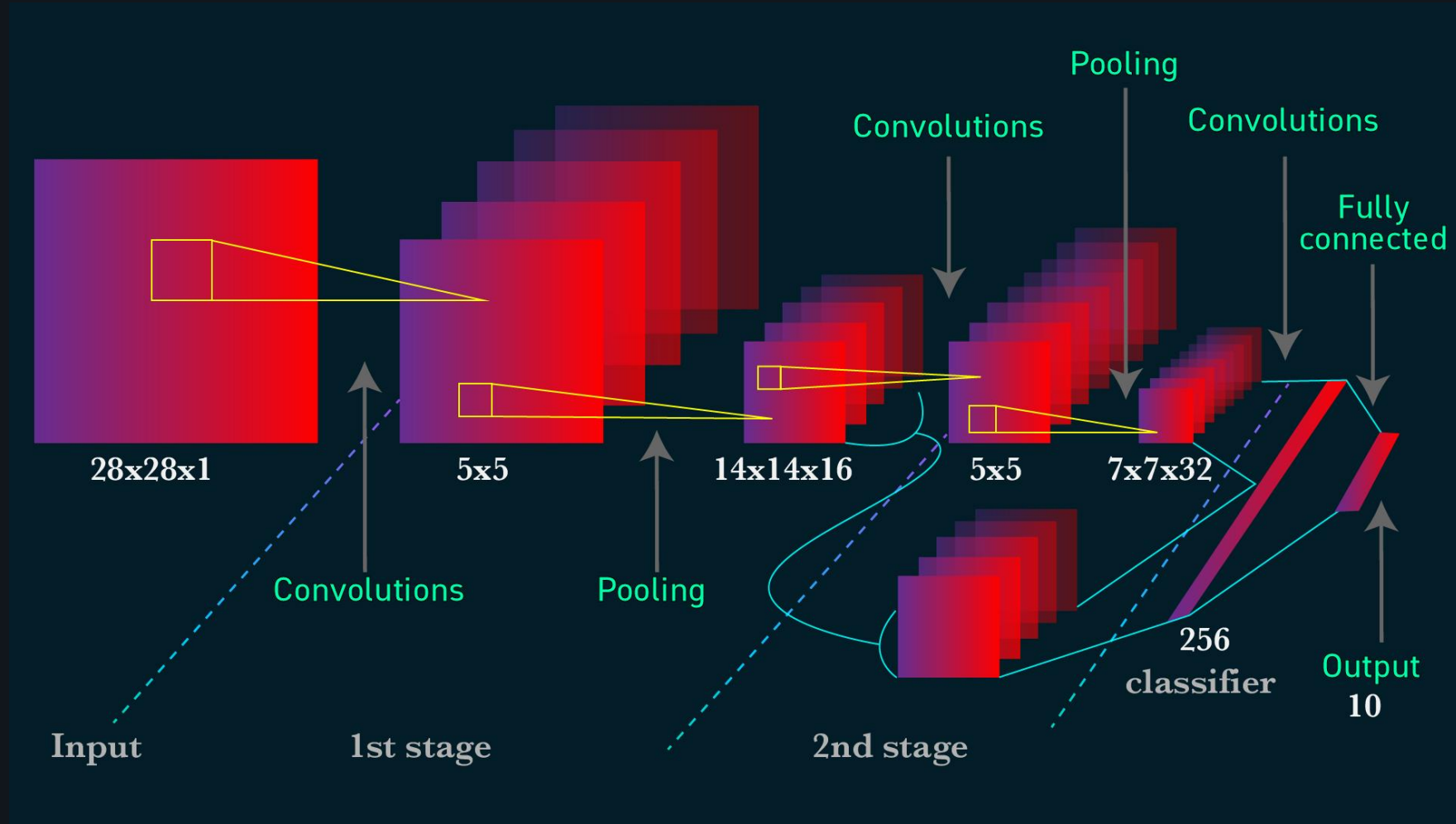
Deep convolutional neural networks in computer vision

Deep Neural Networks

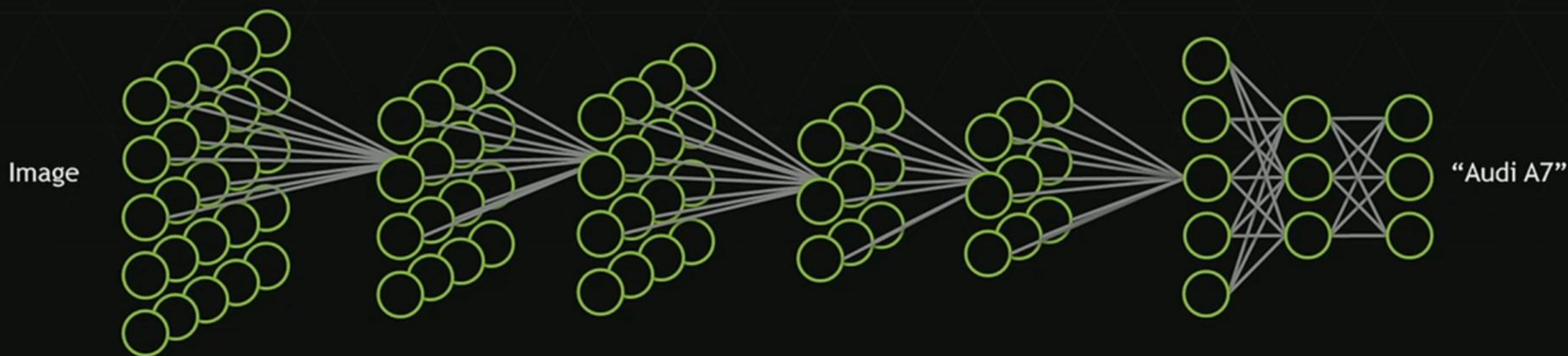
Deep Neural Networks learn
hierarchical feature
representations



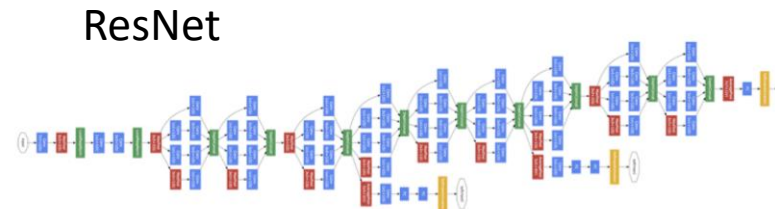
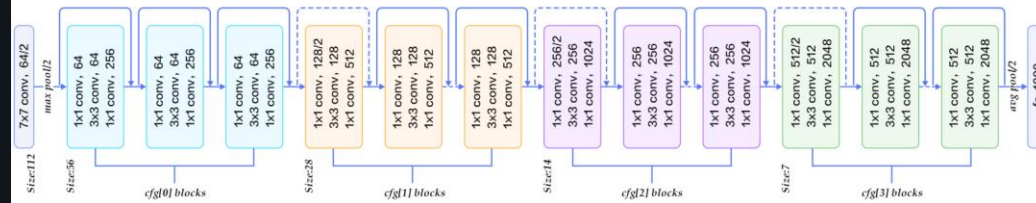
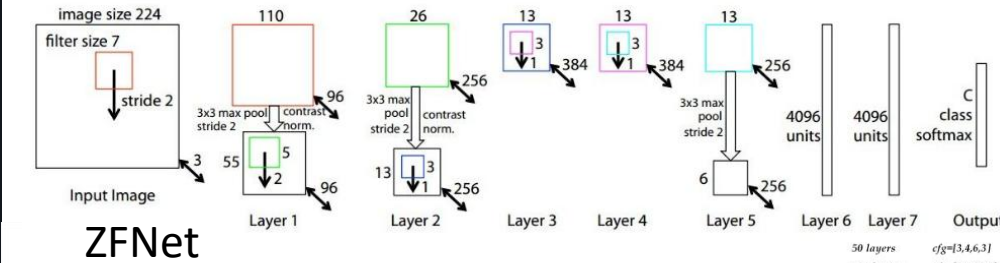
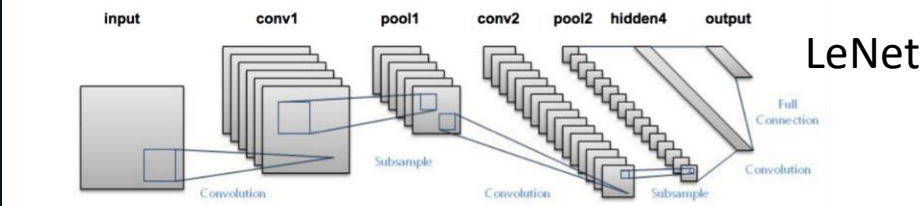
Convolutional Neural Networks (CNNs)



HOW A DEEP NEURAL NETWORK SEES



Convolutional Neural Networks (CNNs)



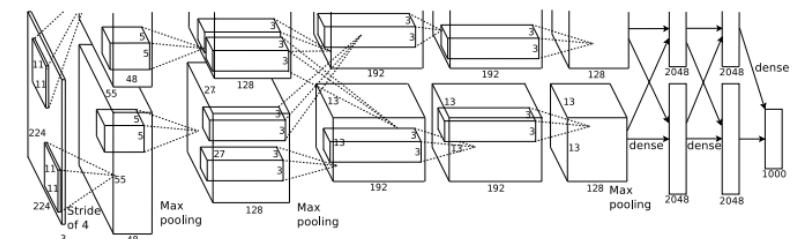
GoogleNet/Inception



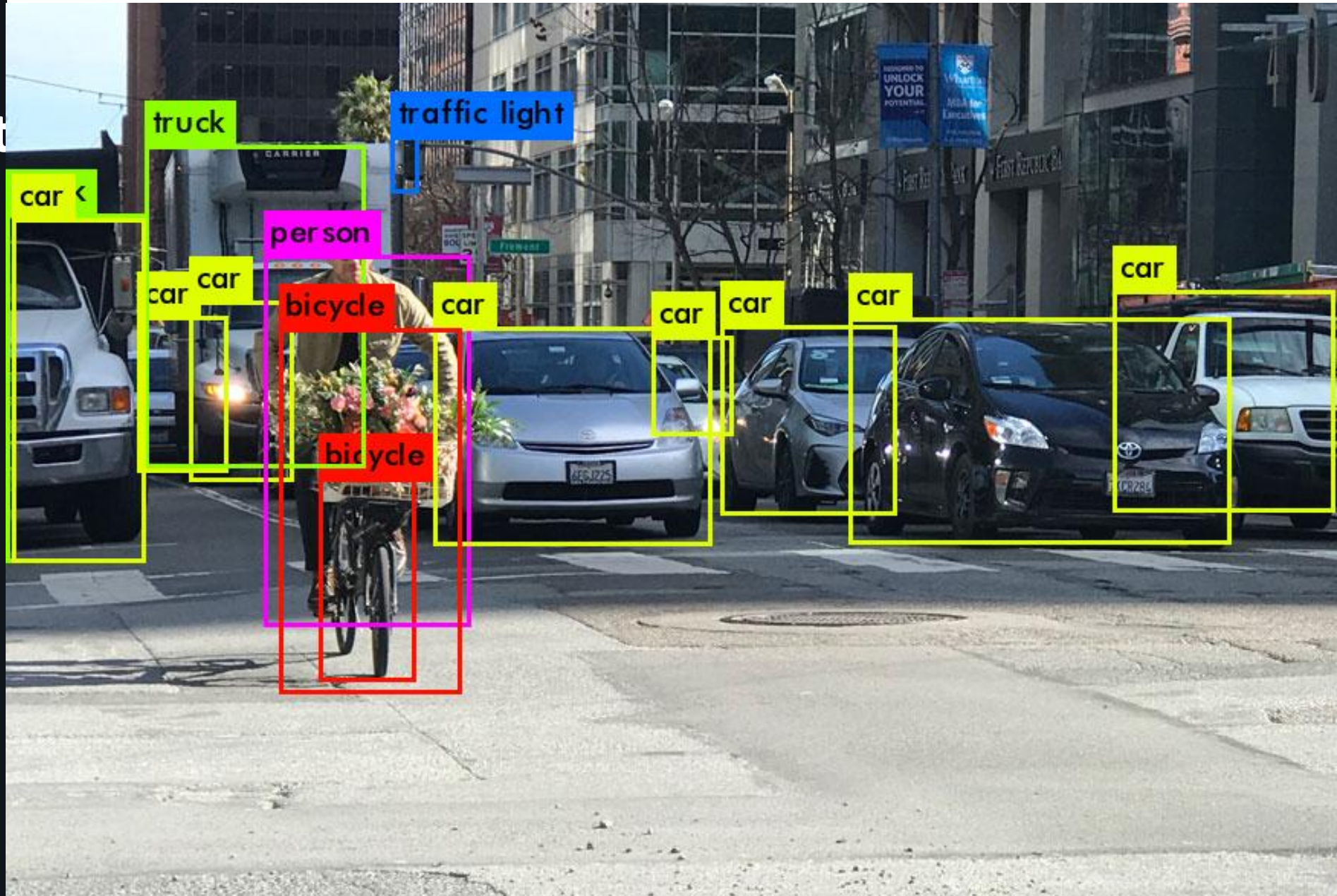
VGGNet



V-Net



YOLO for object detection



Damage Detection in vehicles

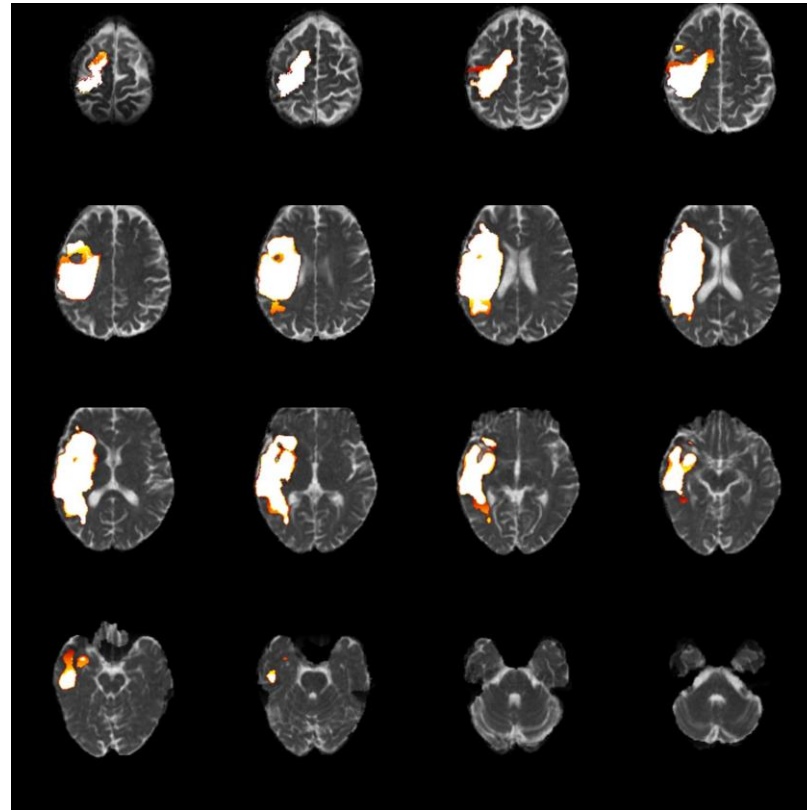
Dataset

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

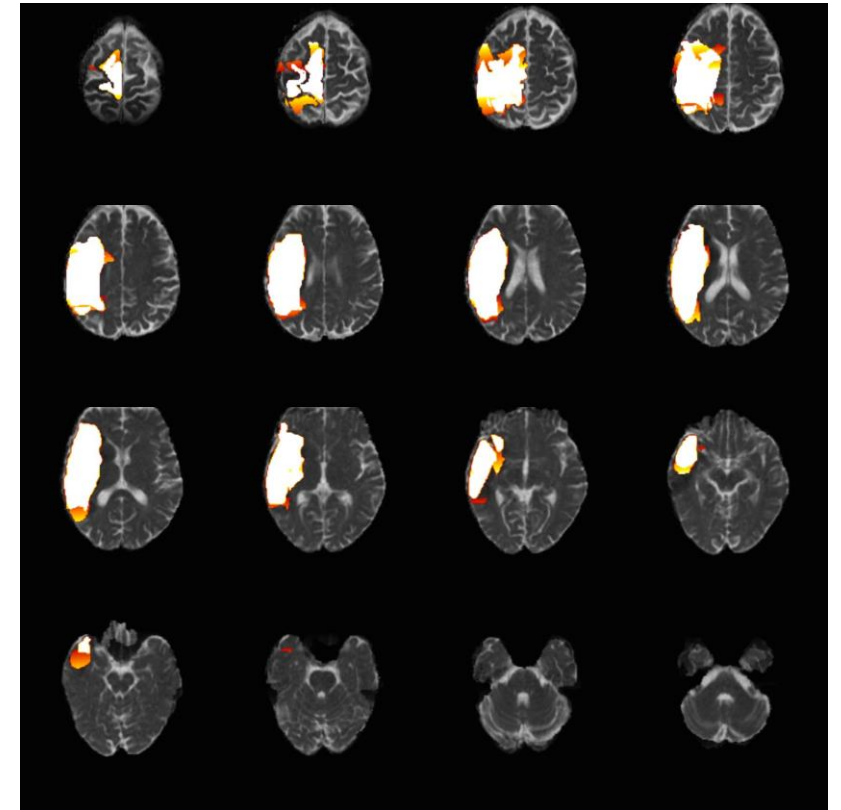


**Ischemic
Stroke
Prediction

Best case**



Prediction



Ground truth

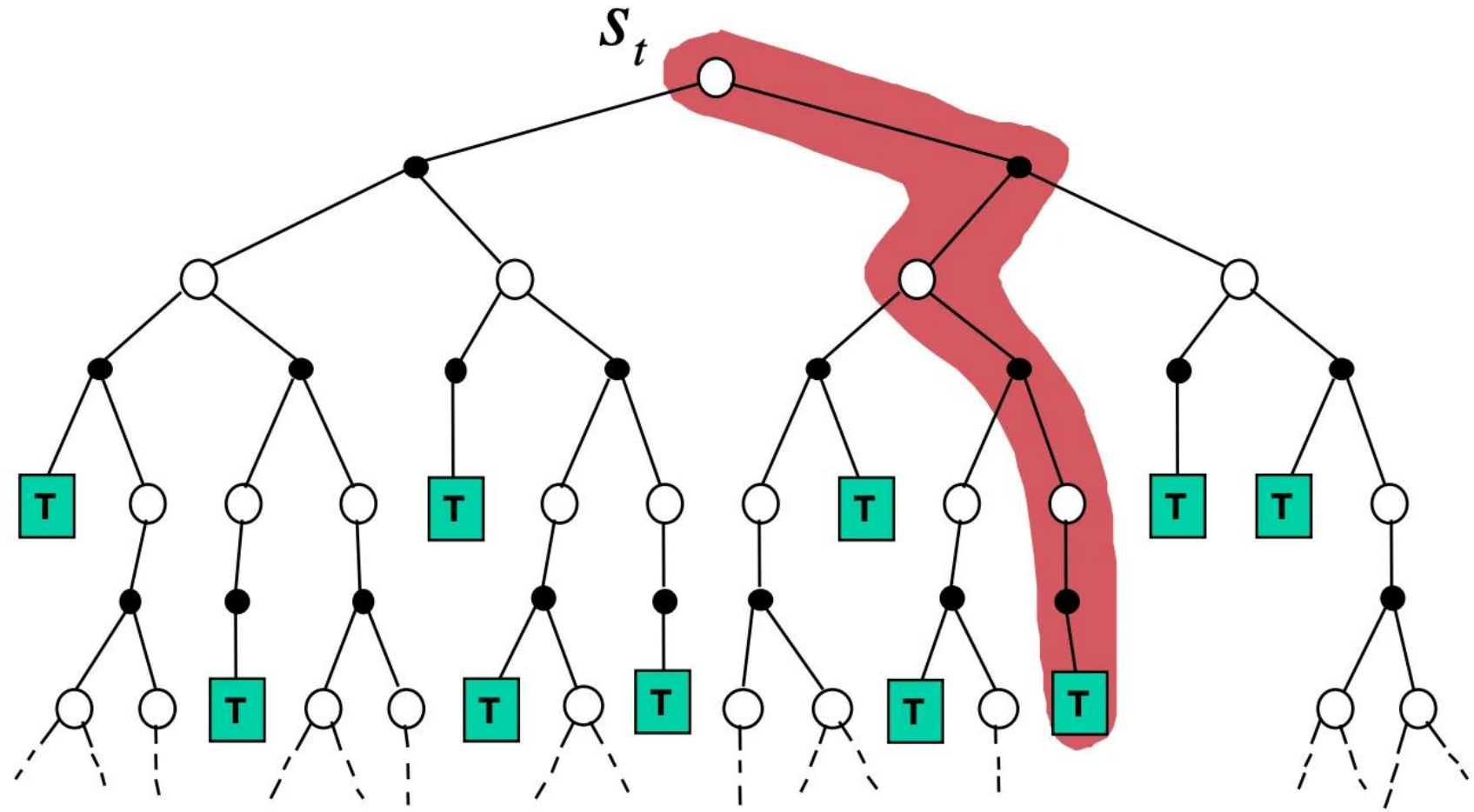
ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI, Winzeck, ..., Monteiro, Oliveira, et al., Frontiers in Neuroscience, 2018

Segmentation and analysis of medical images

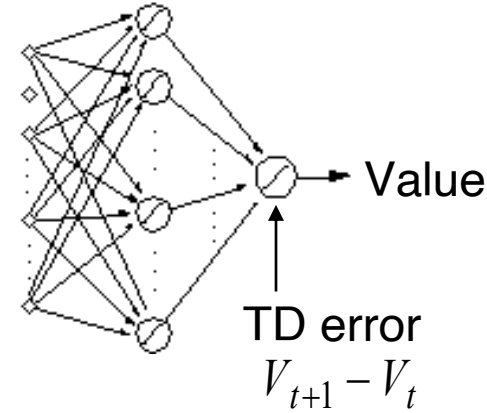
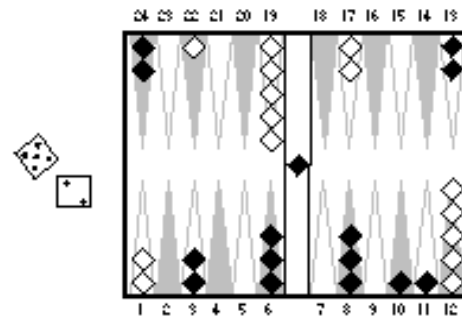


Deep Reinforcement Learning

Reinforcement Learning



TD-Gammon



Action selection
by 2–3 ply search

Start with a random network

Play millions of games against self

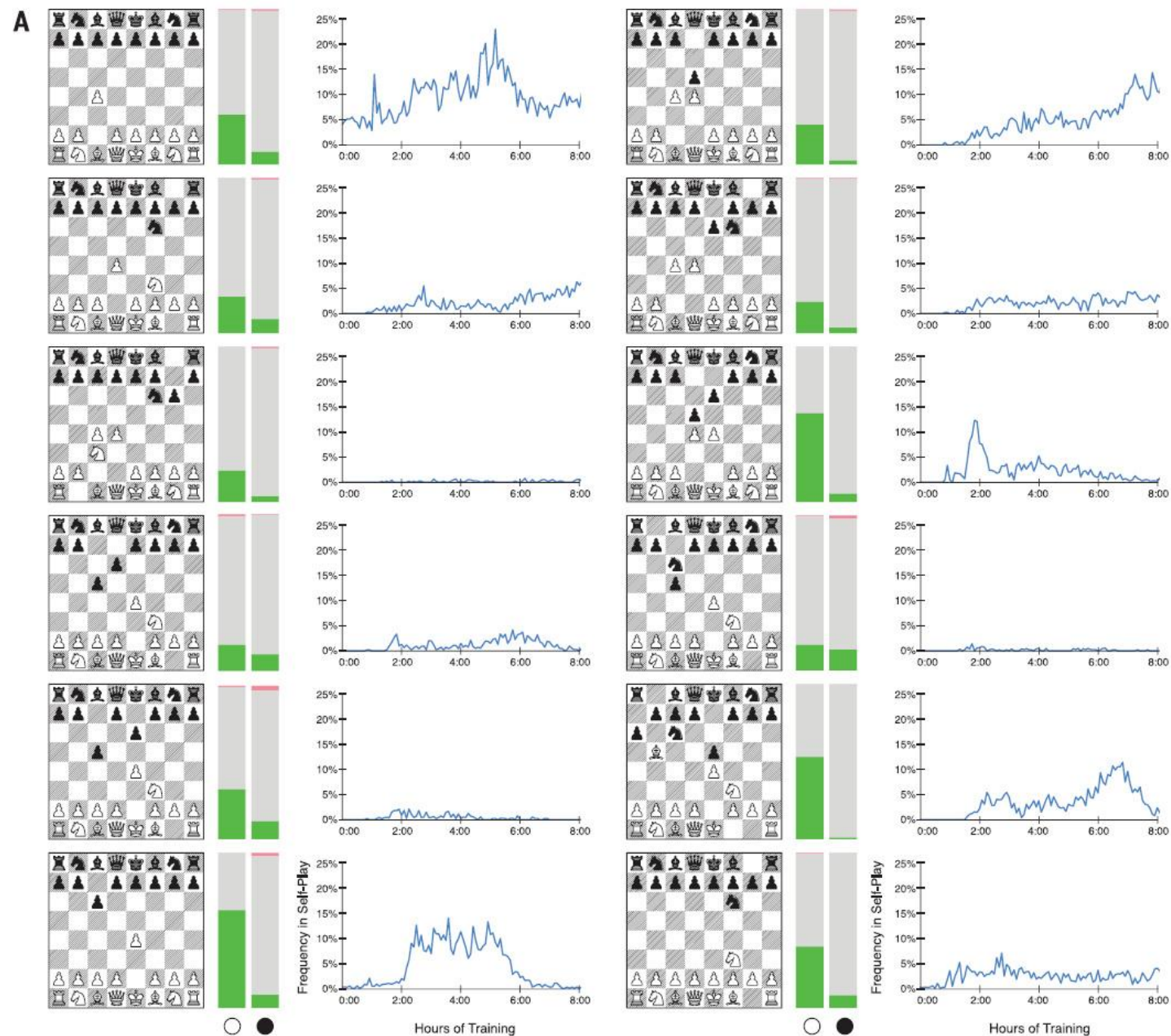
Learn a value function from this simulated experience

This produces arguably the best player in the world

AlphaGo



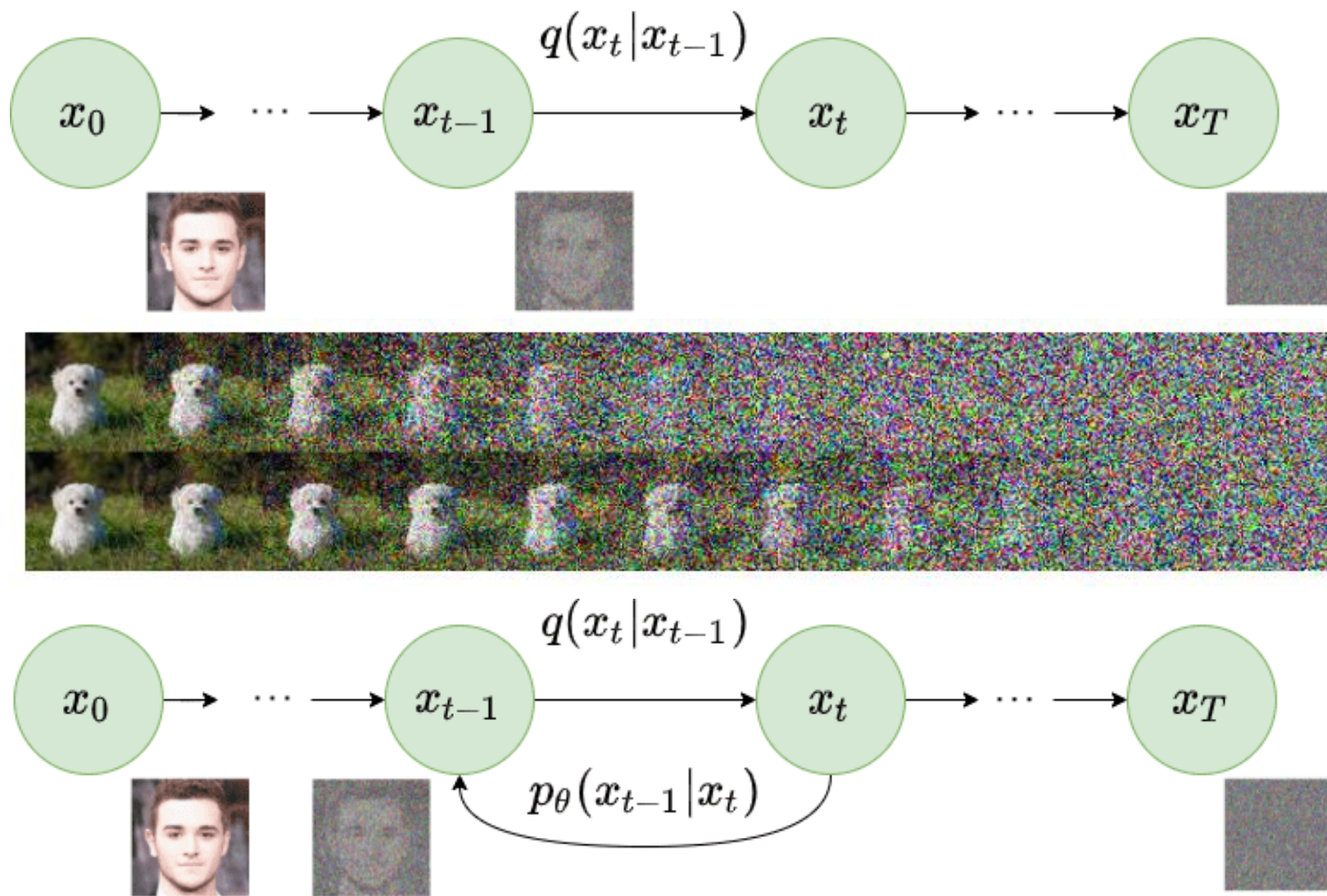
AlphaZero



Diffusion models

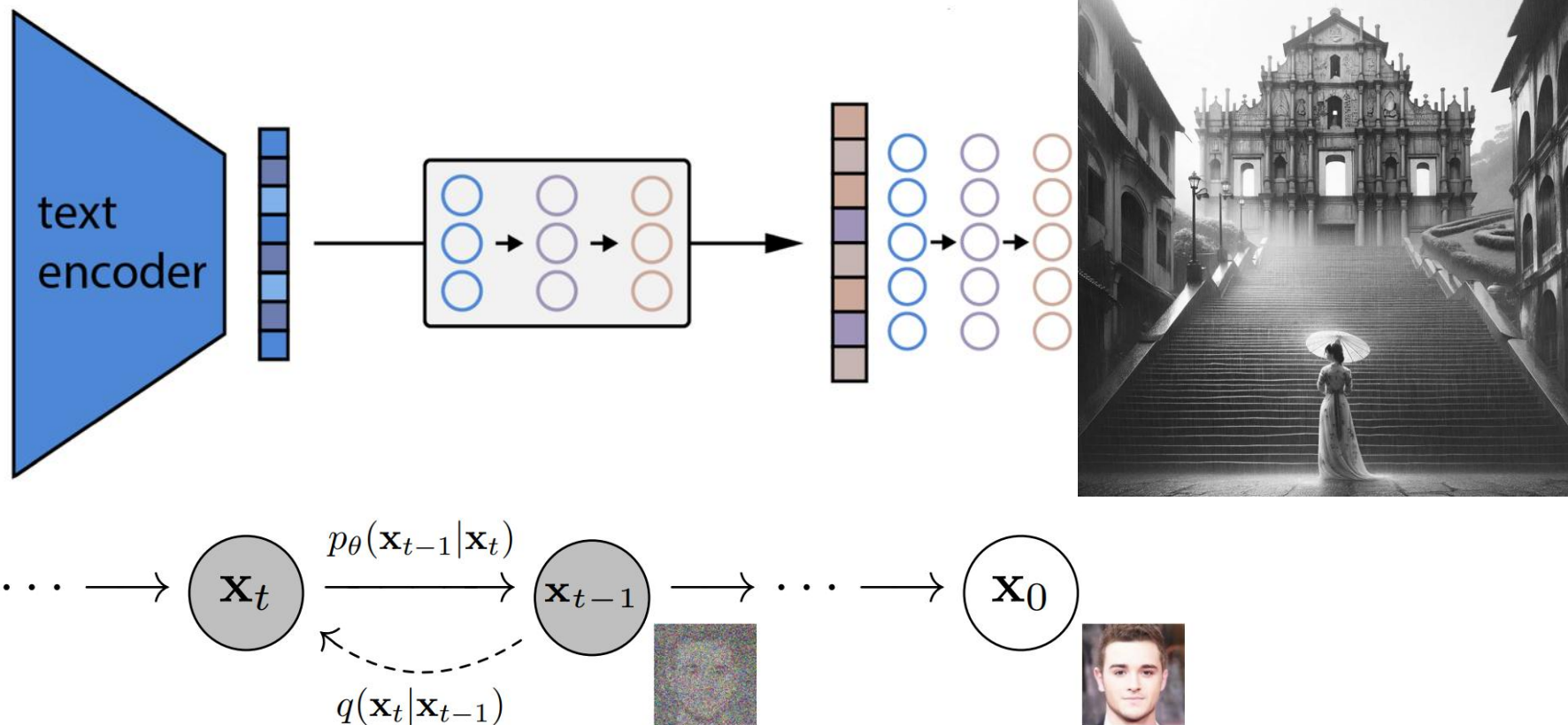
The background of the slide is a dark blue gradient. It features a complex pattern of thin, light blue lines and dots. These elements are arranged in a way that suggests a grid or a series of intersecting paths, possibly representing a diffusion process or a data visualization. The lines and dots are more densely packed on the left side and become sparser towards the right, creating a sense of depth and movement.

Diffusion 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



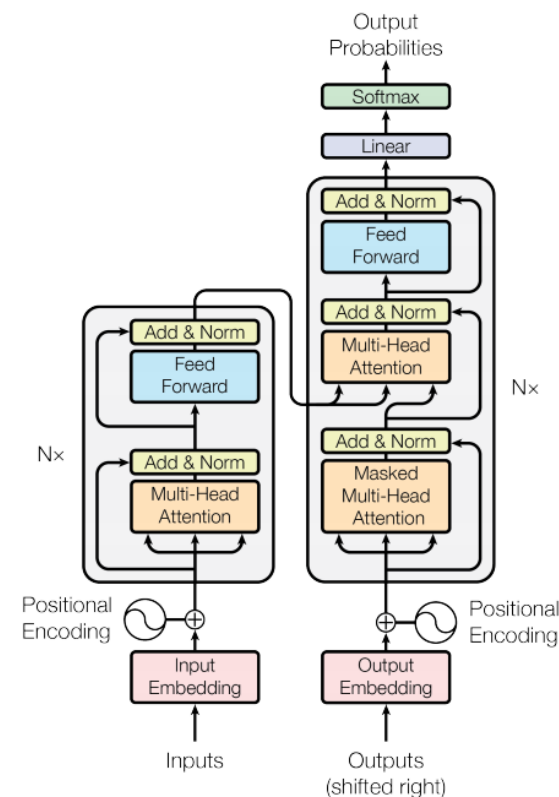
Pseudomnesia / The Electrician





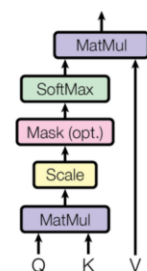
Transformers and Foundational Models

Transformers

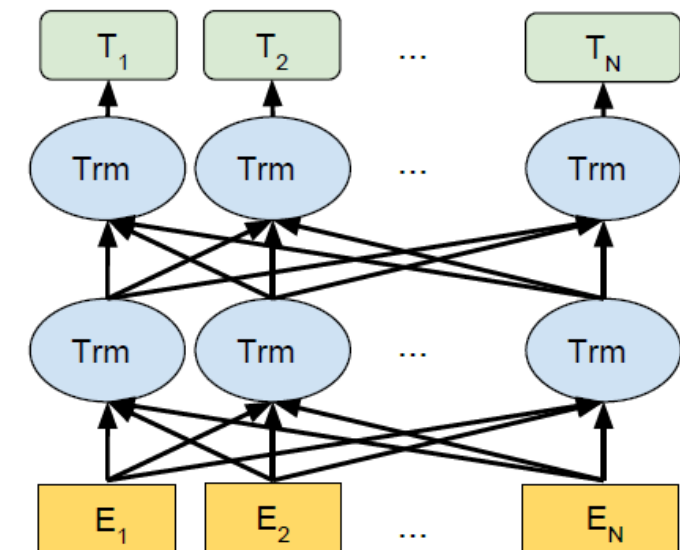
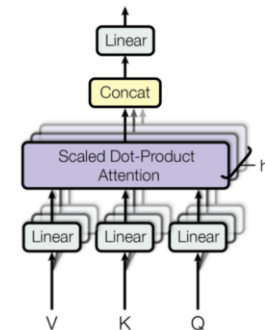


Transformer

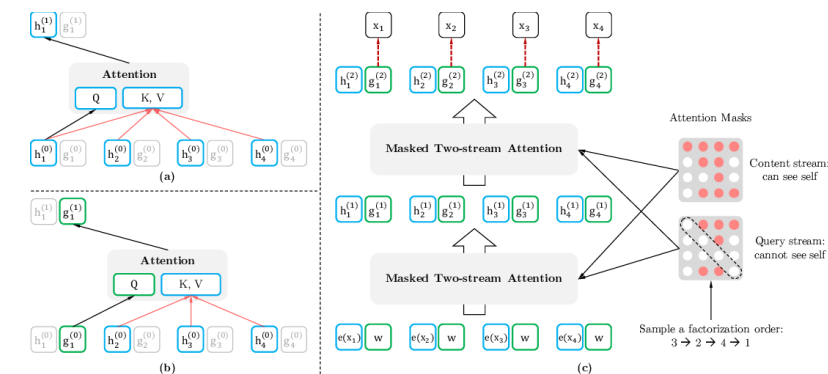
Scaled Dot-Product Attention



Multi-Head Attention



BERT



XLNet

Language Models

- Language models are statistical models of the human language

Prediction and Entropy of Printed English

By C. E. SHANNON

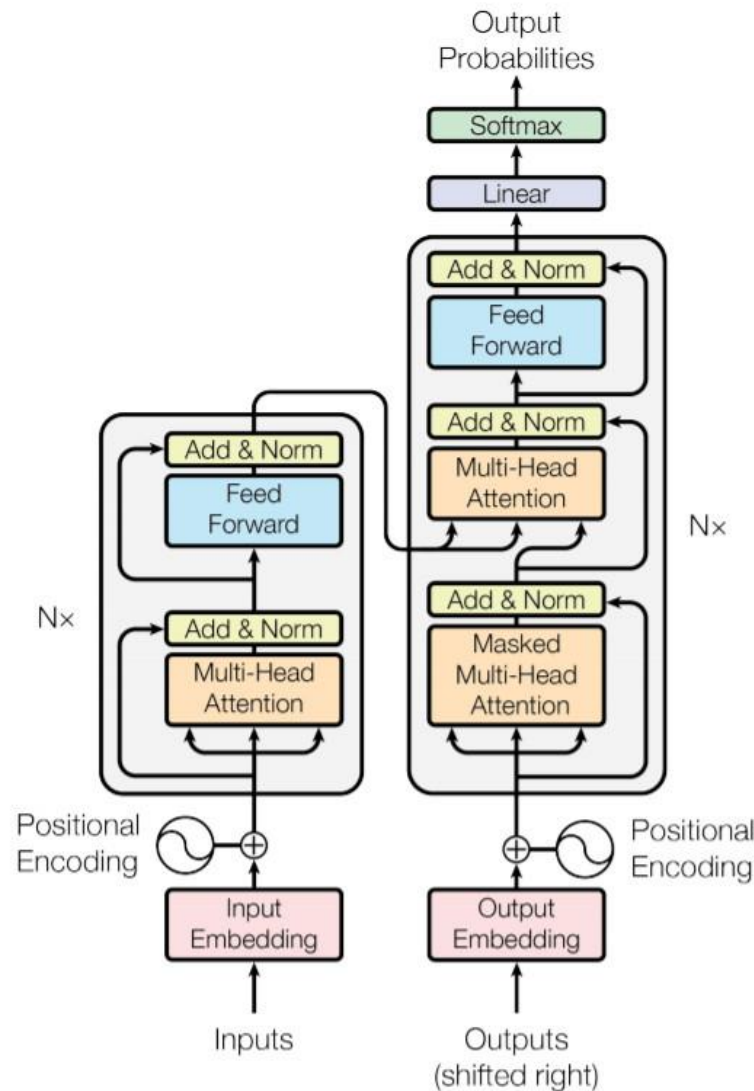
(Manuscript Received Sept. 15, 1950)

A new method of estimating the entropy and redundancy of a language is described. This method exploits the knowledge of the language statistics possessed by those who speak the language, and depends on experimental results in prediction of the next letter when the preceding text is known. Results of experiments in prediction are given, and some properties of an ideal predictor are developed.

- Simple models simply generate words (or characters) according to superficial text statistics, e.g. the frequency of words.
- Language models have been used for decades, and take many forms: rule-based (e.g, ELIZA), grammar-based, statistical, deep-learning based, ...
- They can be used to generate text that is in accordance with the model. Simple models lead to simple behavior, complex models lead to (possibly very) complex behavior

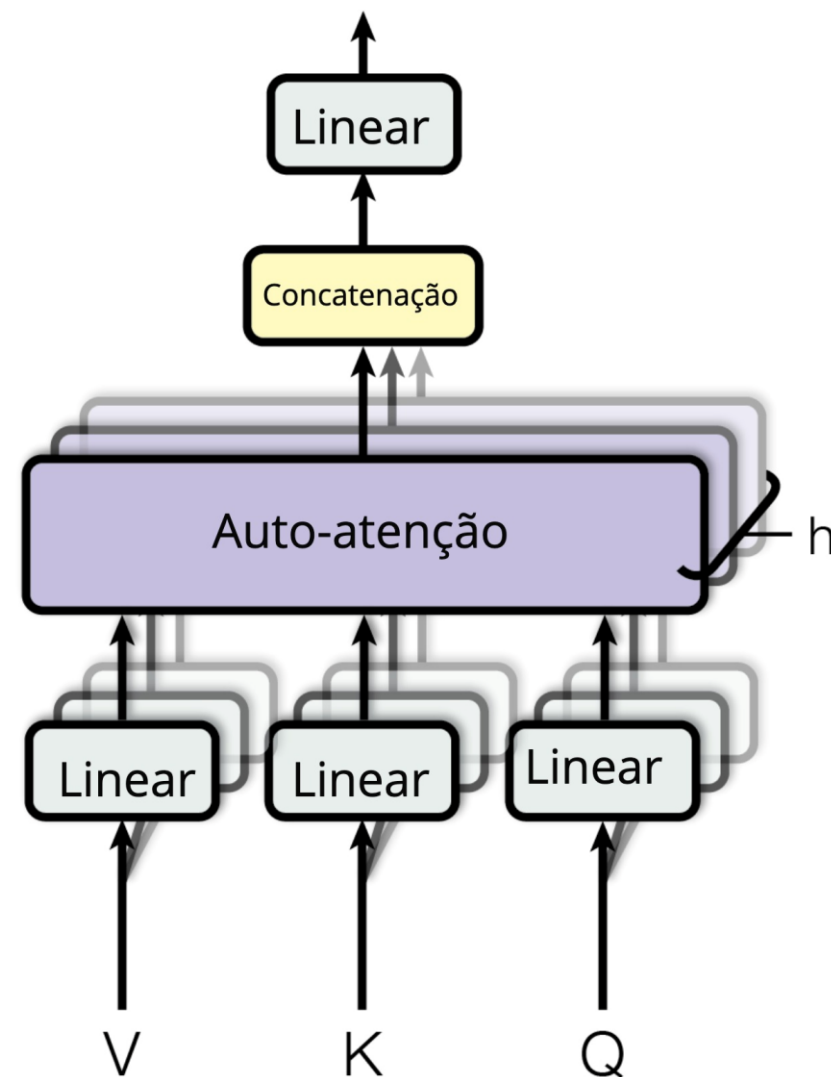
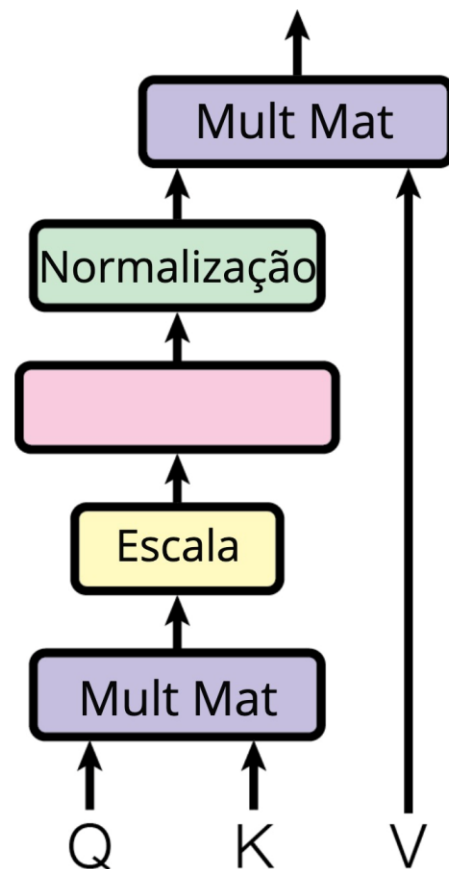
Large Language Models

Training and Inference

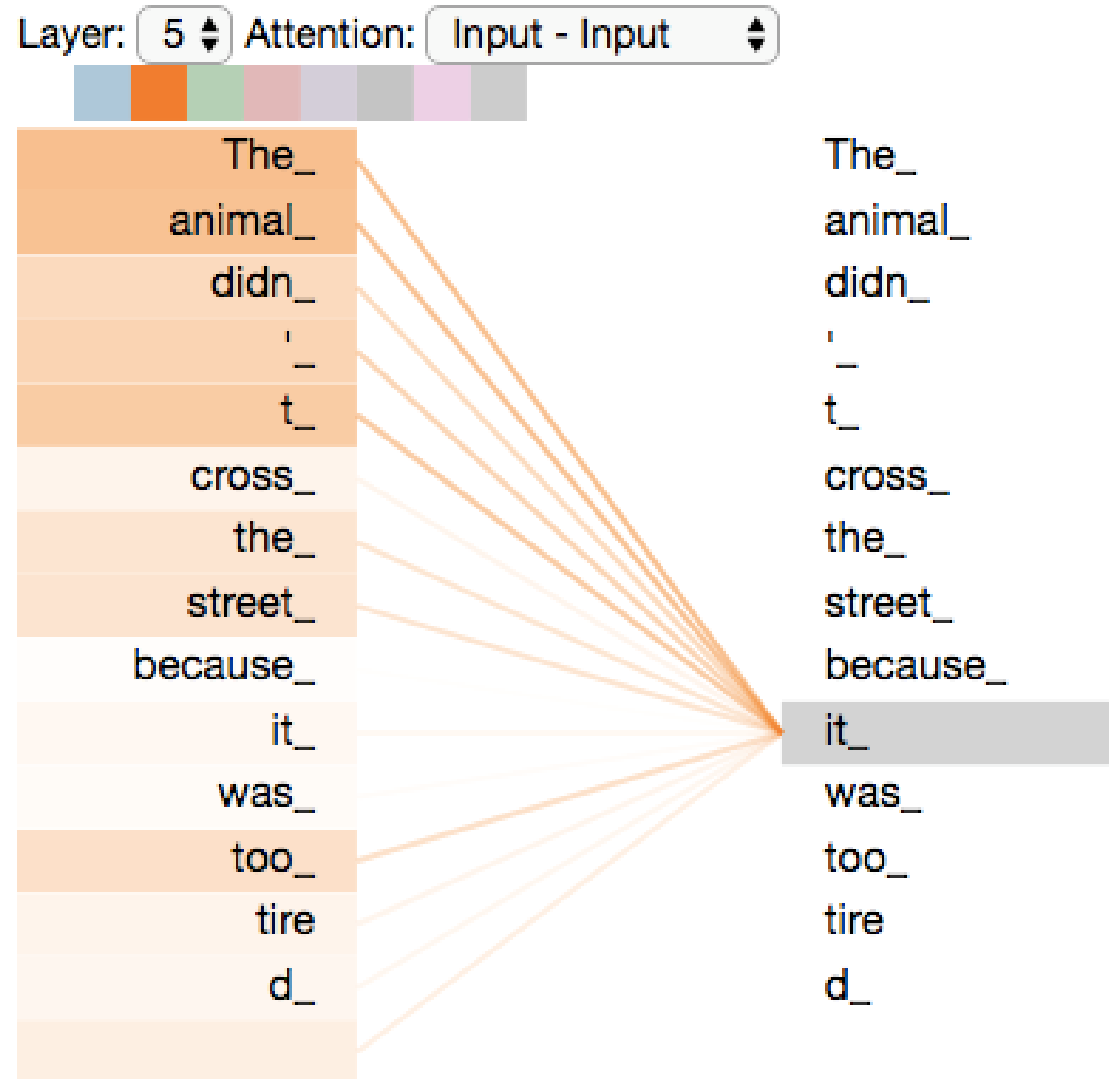


- Use the popular transformer architecture, introduced in 2017
- Words (tokens) are embedded into a “meaning” space
- Model is trained to predict the next word
- Text can be generated according to predicted word probabilities
- Creativity can be controlled by a temperature parameter
- Different choices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

Attention Mechanisms in Transformers

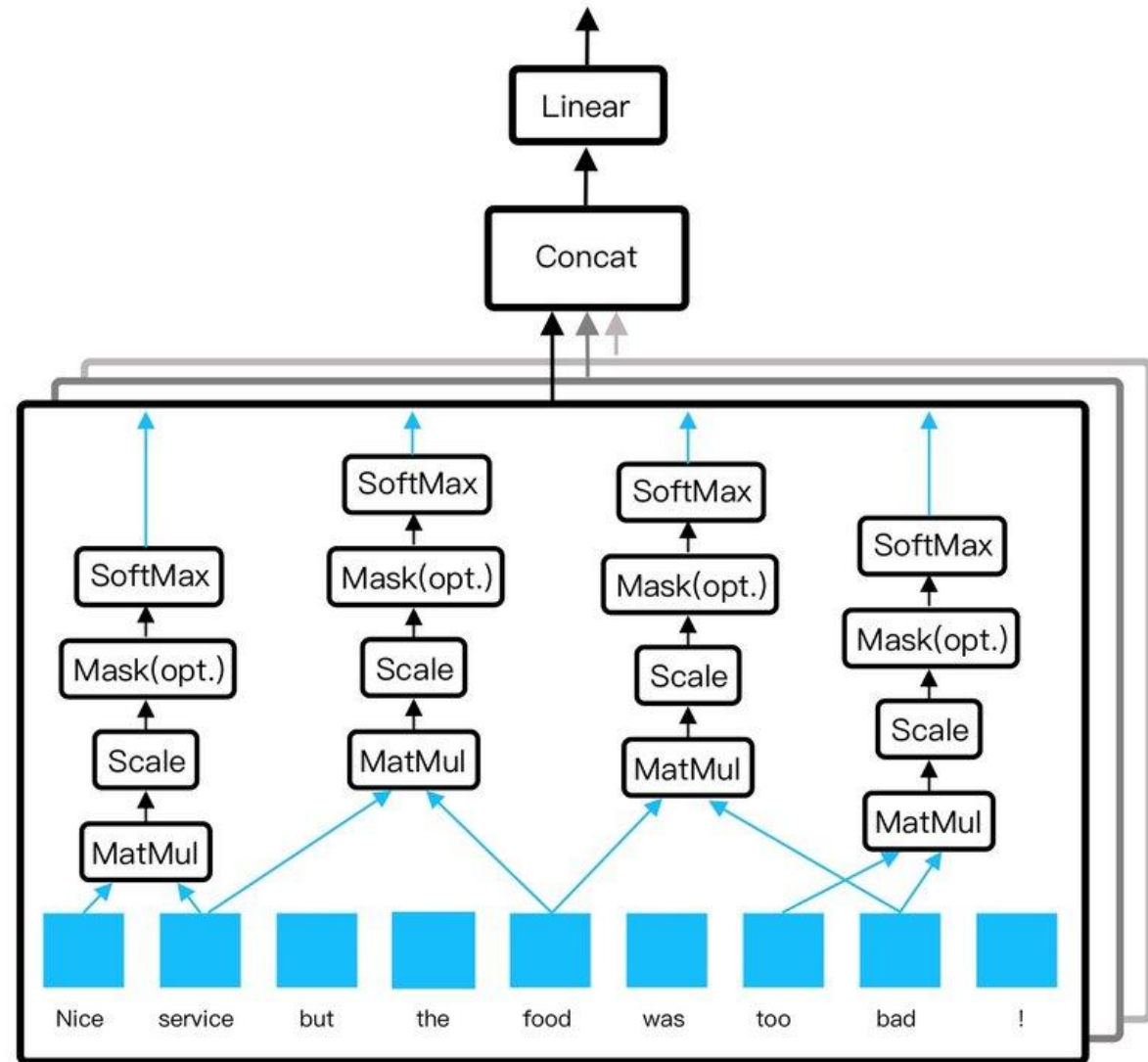


Self-attention in transformers



Self-attention mechanisms in transformers enables them to related different parts of text, building intermediate concepts

Self-attention in transformers



Self-attention is performed by operating linear operations between words, followed by non-linearities

Large Language Models



WIKIPEDIA
The Free Encyclopedia



Common Crawl

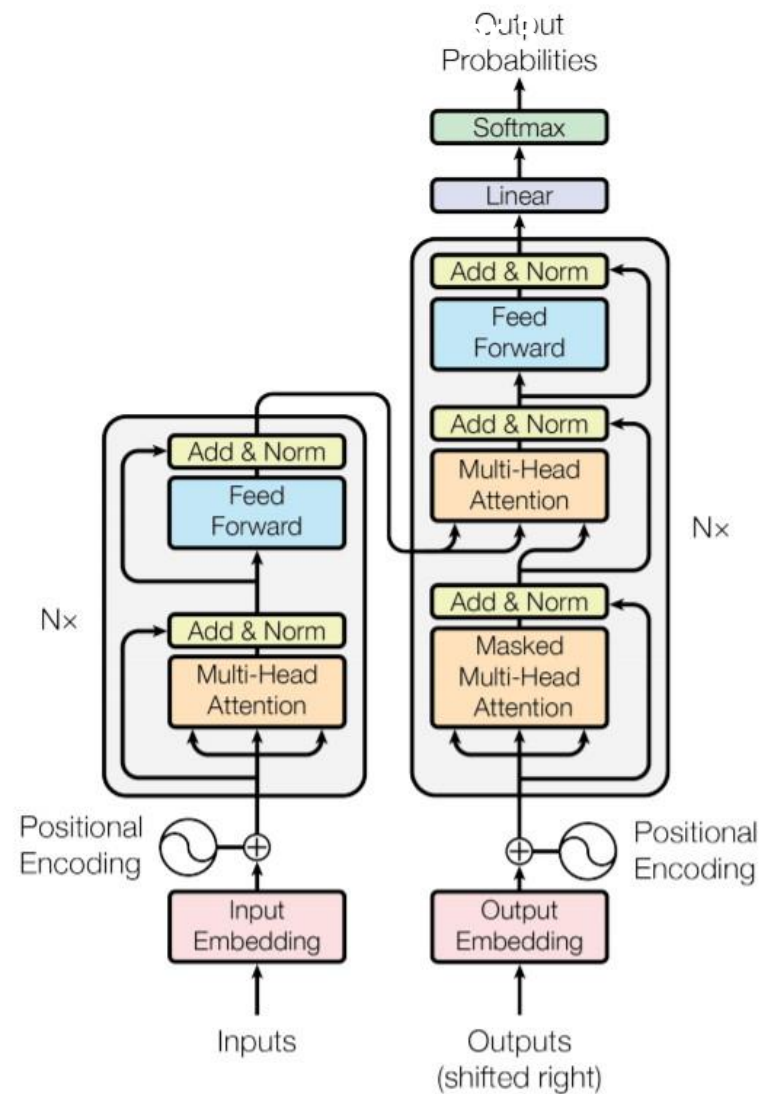


WebText



BookCorpus

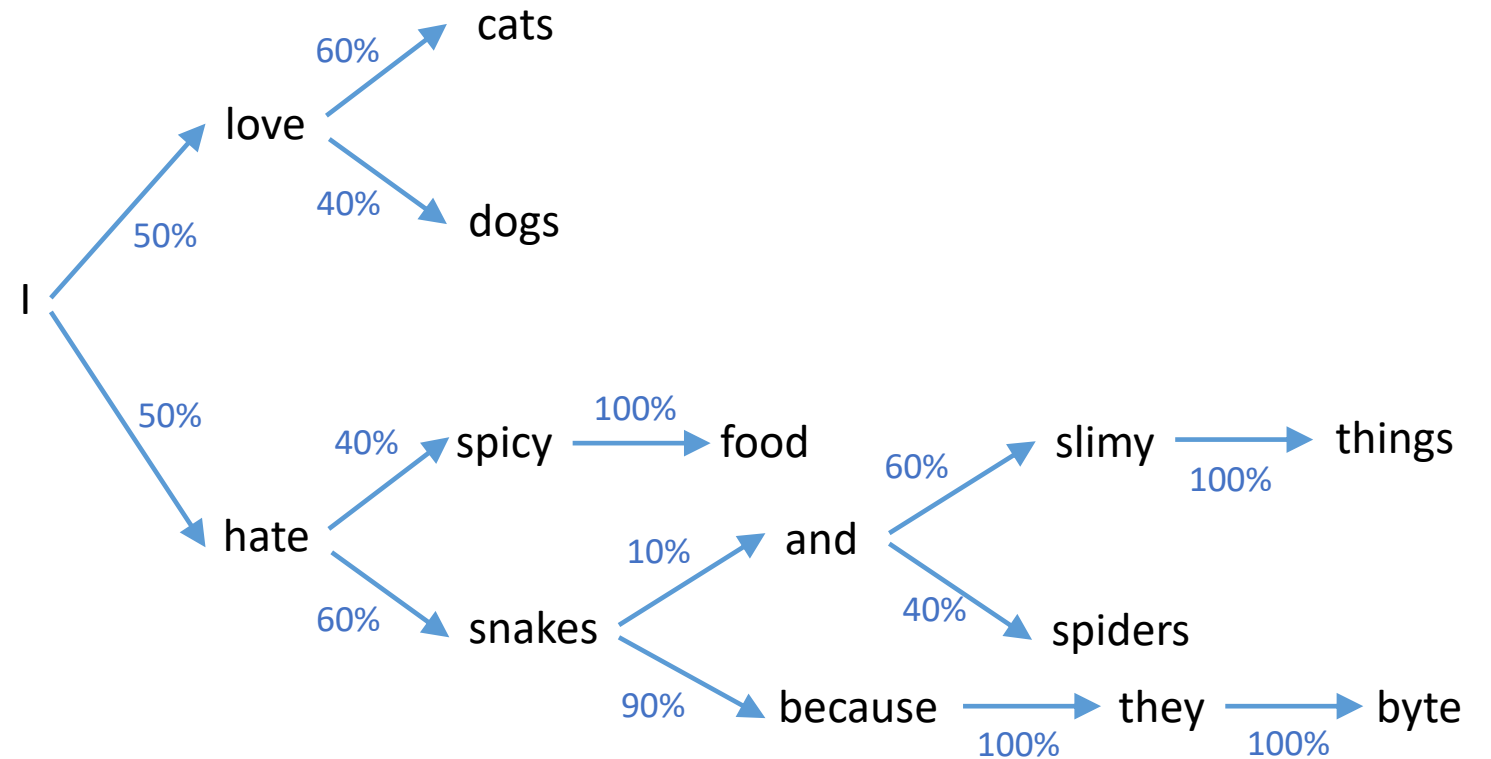
GPT-3: Generative Pre-trained Transformer 3



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”

Large Language Models

GPT-3: Generative Pre-trained Transformer 3

Text: Select the "Students" from the "School" table joined with "Class" table:

```
Code: SELECT * FROM Students  
INNER JOIN Class  
ON Students.ID = Class.StudentID
```

Q: woman smiling
mysteriously



A: Mona Lisa
Leonardo Da Vinci

Q: giant waves
crashing over boats



A: The Great Wave Off
Kanagawa
Hokusai

Q: dogs playing cards
in a back room



A: A Friend In Need -
Cassius Marcellus
Coolidge

Researcher: *It is currently July 2020. For the last four months, the global economy has been disrupted due to the introduction of COVID-19, a novel coronavirus which is highly transmissible. This new coronavirus has created a global pandemic. In response to the pandemic, the governments of the world have instituted a series of lock downs, asking people to stay at home to prevent the spread of the virus. This resulted in the shutdown of many businesses, can you guess which businesses were impacted the most?*

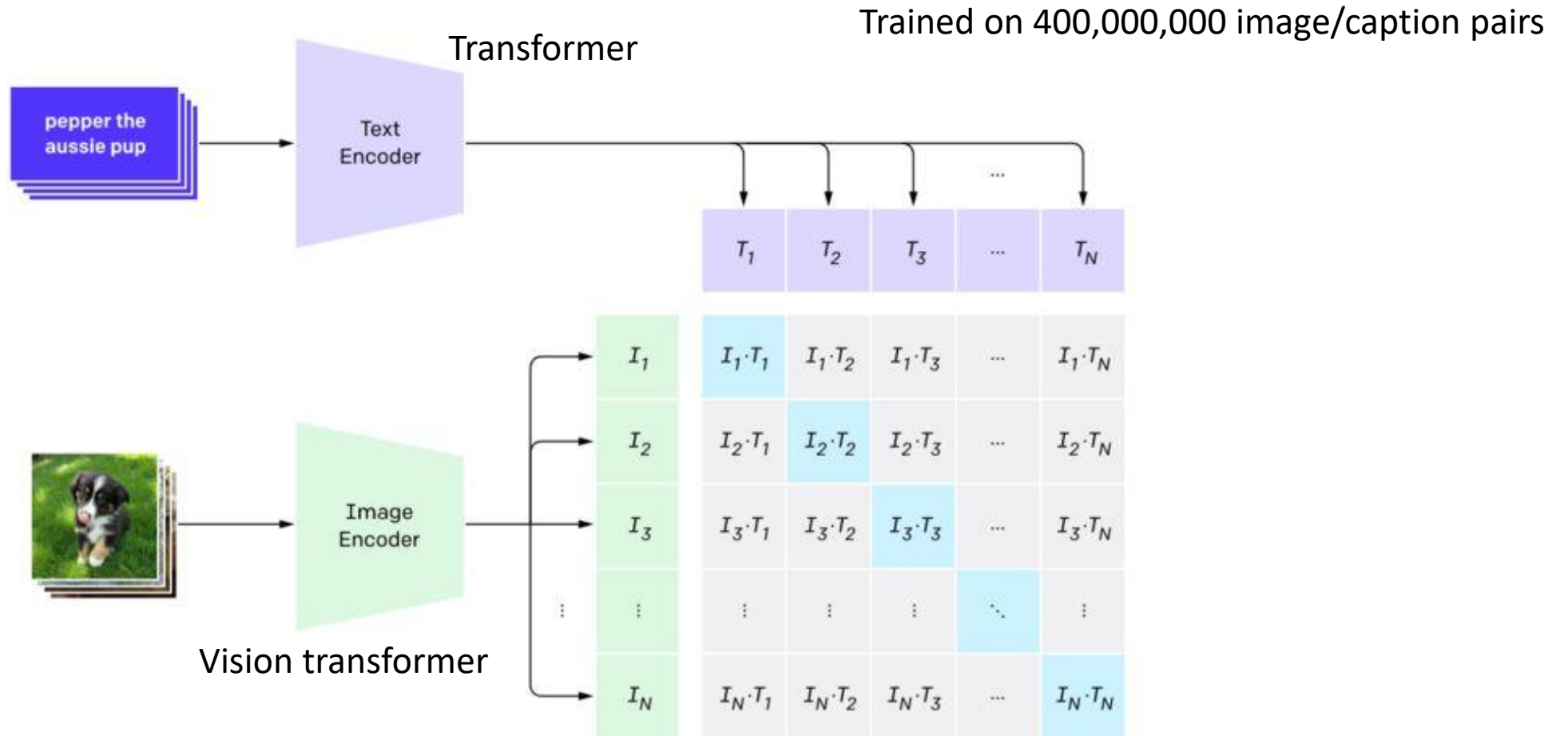
GPT-3: The businesses that needed to function in a way that allowed humans to interact with each other directly have been shutdown. This includes restaurants, retail stores and most other service based industries. The largest industries that have not been shut down are those that do not require human interaction.

Researcher: *Correct. Millions of people were forced to stop working and stay at home. As a result of this, the governments of the world decided to print money to give to the affected businesses and workers.*

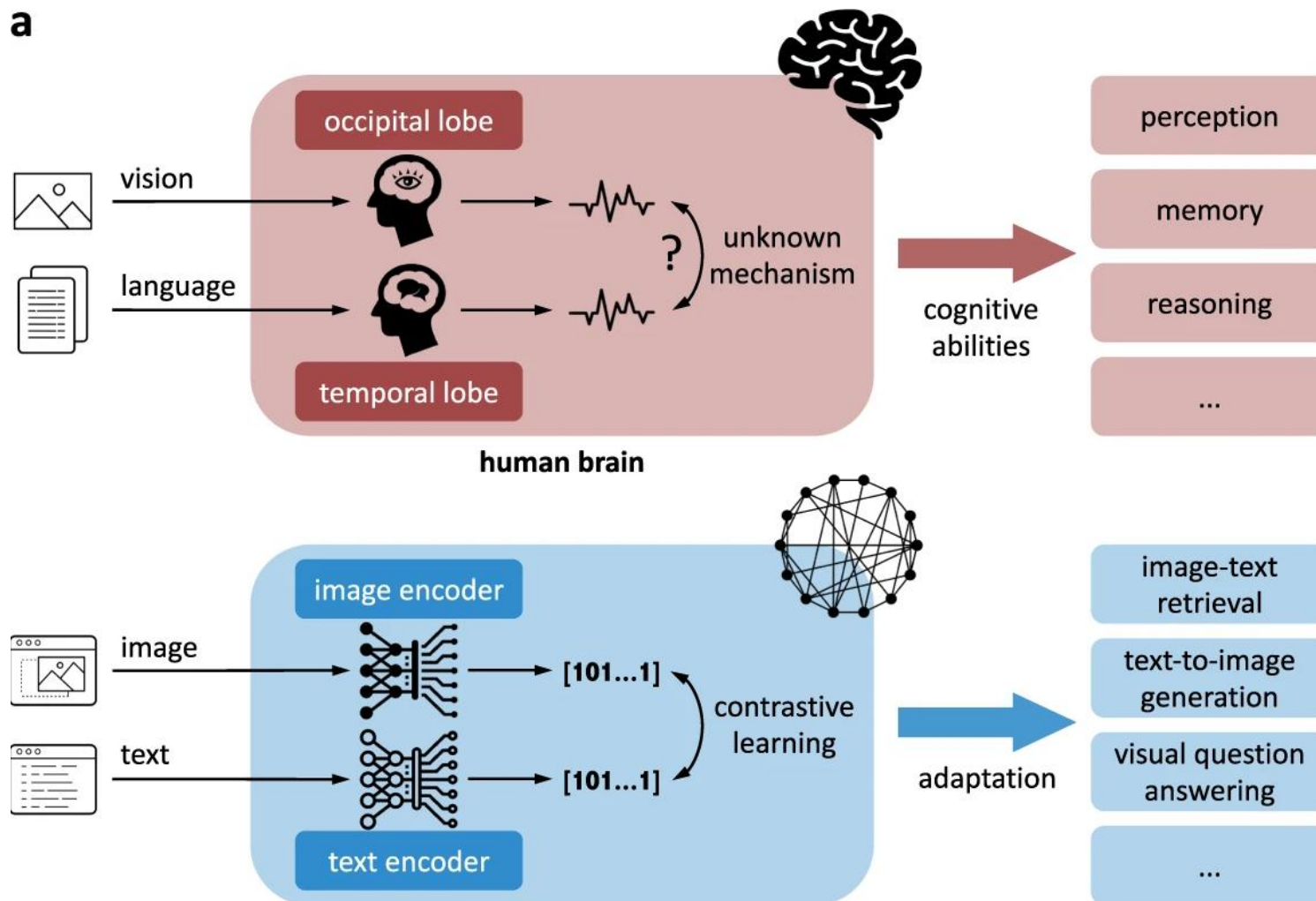
GPT-3: Well, you said millions of people were forced to stop working. That is a huge amount of production that stopped. The economy has lost lots of output and this will not be made up, because output cannot exceed demand. Some output was distributed to the unemployed in the form of printed money. This money was not spent into the economy, and it will not be spent in the future. When this money comes out of savings accounts, it will cause inflation.

CLIP: training models with text and image

1. Contrastive pre-training



Vision and language models

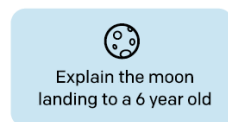


From GPT-3 to ChatGPT (InstructGPT)

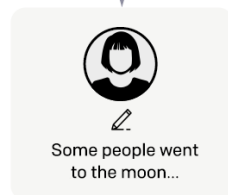
Step 1

**Collect demonstration data,
and train a supervised policy.**

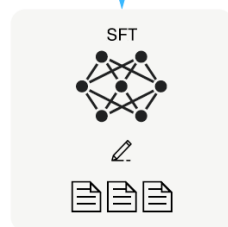
A prompt is
sampled from our
prompt dataset.



A labeler
demonstrates the
desired output
behavior.



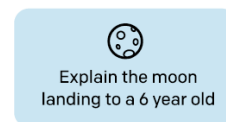
This data is used
to fine-tune GPT-3
with supervised
learning.



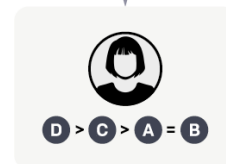
Step 2

**Collect comparison data,
and train a reward model.**

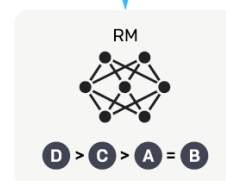
A prompt and
several model
outputs are
sampled.



A labeler ranks
the outputs from
best to worst.



This data is used
to train our
reward model.



Step 3

**Optimize a policy against
the reward model using
reinforcement learning.**

A new prompt
is sampled from
the dataset.



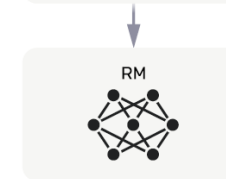
The policy
generates an output.



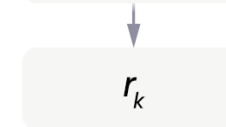
Once upon a time...



The reward model
calculates a
reward for
the output.



The reward is
used to update
the policy
using PPO.



ChatGPT to GPT-4

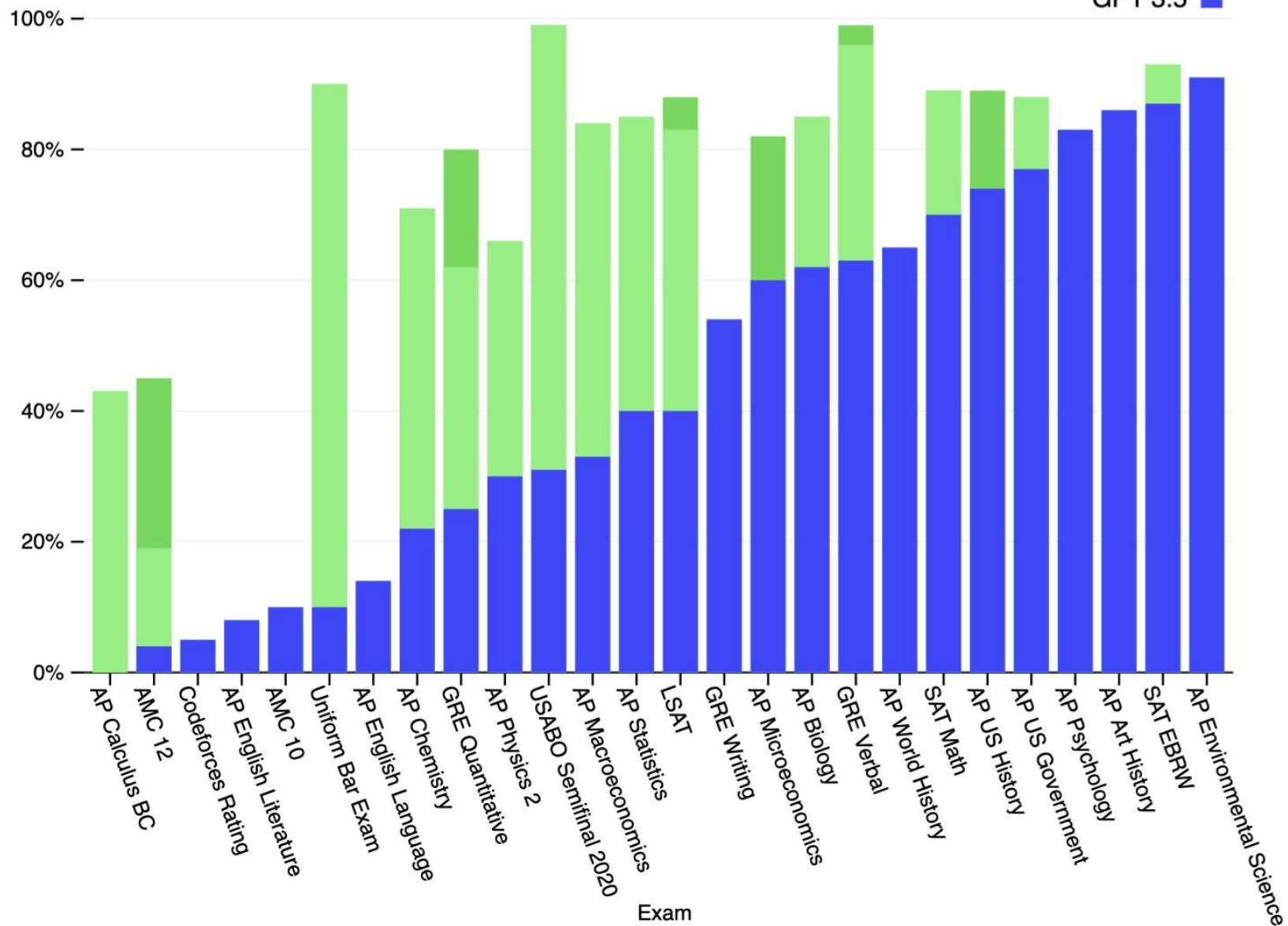
- ChatGPT processes at most 2048 tokens, while GPT-4 can use up to 32000
- Less prone to hallucinations and more accurate
- More creative, generates more varied output
- GPT-4 passes the bar exam, several Advanced Placement examinations, US Medical licencing exam, sommelier examination, Stanford Medical Scholl reasoning exam, ...

Features	GPT-3	GPT-4
Parameters	175 billion	To be disclosed
Modality	Text only	Text and image
Performance	Poor on complex tasks	Human-level on various benchmarks
Accuracy	Prone hallucinations and errors	More reliable and factual

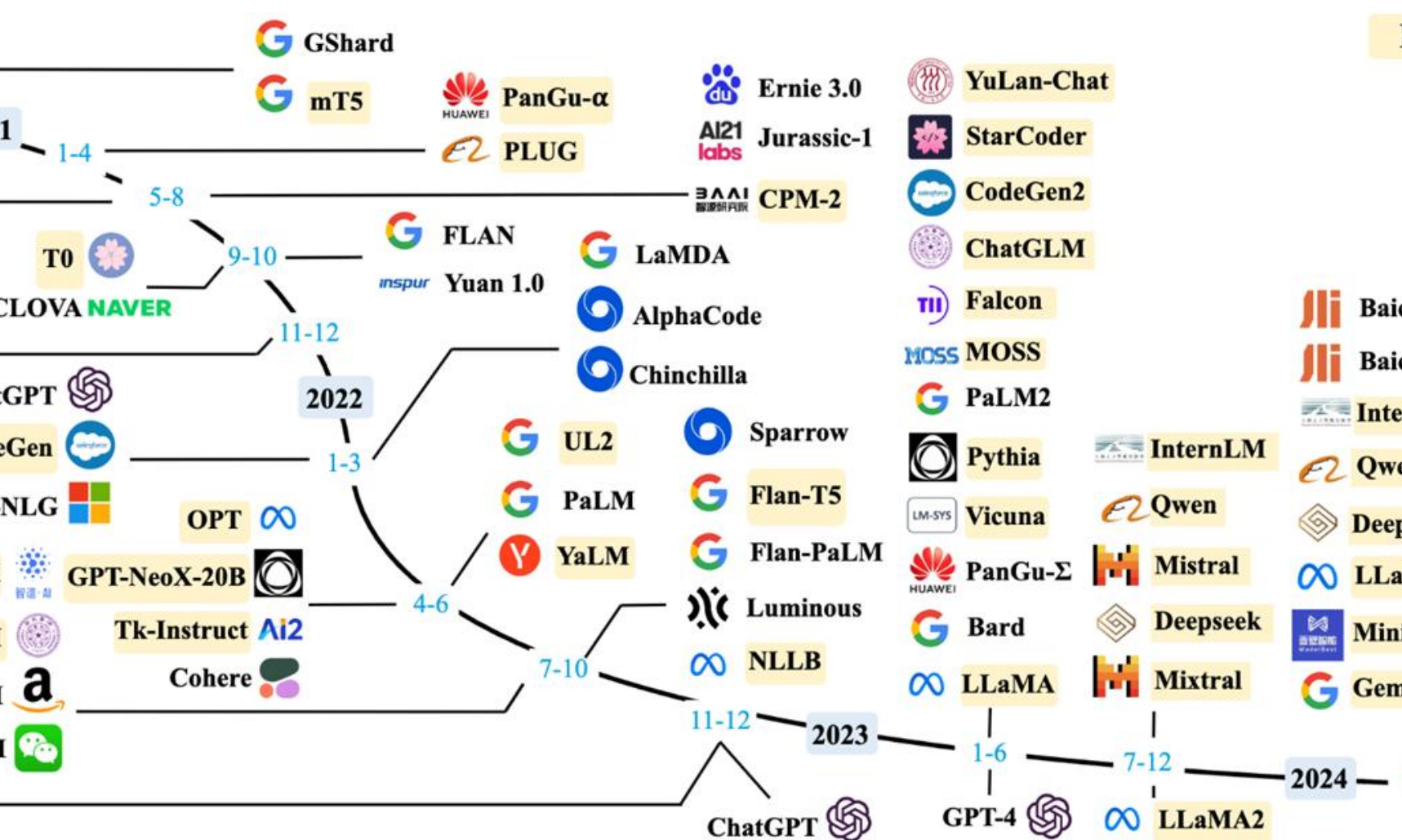
GPT-4 vs GPT-3.5 in exams

Exam results (ordered by GPT 3.5 performance)

Estimated percentile lower bound (among test takers)



language models



Most capable Vision and Language Models

PaliGemma

PaliGemma is a vision language model (VLM) by Google that has multimodal capabilities.

VISION-LANGUAGE

DEPLOY WITH ROBOFLOW

 • 2.0k+ stars • Custom Google license • Released May 14, 2024

GPT-4o

GPT-4o is OpenAI's third major iteration of GPT-4 expanding on the capabilities of GPT-4 with Vision

VISION-LANGUAGE

Released May 13, 2024

LLaVA-1.5

LLaVA is an open source multimodal language model that you can use for visual question answering and has limited support for object detection.

OBJECT DETECTION

DEPLOY WITH ROBOFLOW

 • 16,000 stars • Apache-2.0 license • Released Oct 5, 2023

Google Gemini

Gemini is a family of Large Multimodal Models (LMMs) developed by Google Deepmind focused specifically on multimodality.

VISION-LANGUAGE

Released Dec 6, 2023

QwenVL

Qwen-VL is an LMM developed by Alibaba Cloud. Qwen-VL accepts images, text, and bounding boxes as inputs. The model can output text and bounding boxes. Qwen-VL naturally supports English, Chinese, and multilingual conversation.

VISION-LANGUAGE

 • 3.3k+ stars • Tongyi Qianwen license

BakLLaVA

BakLLaVA is an LMM developed by LAION, Ontocord, and Skunkworks AI. BakLLaVA uses a Mistral 7B base augmented with the LLaVA 1.5 architecture.

VISION-LANGUAGE

 • 650 stars • Apache-2.0 license • Released Oct 13, 2023

GPT-4 with Vision

GPT-4 with Vision is a multimodal language model developed by OpenAI.

OBJECT DETECTION

Anthropic Claude 3

VISION-LANGUAGE

Custom Anthropic license

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-32kseqlen, 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, Wenzhi 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 Ensenspers 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 Lifly Tianshu, Smart Technology CityGPT.

The background is a dark blue gradient with a pattern of lighter blue lines and dots. The lines are arranged in a grid-like fashion, with some lines being solid and others dashed. The dots are small and scattered, creating a sense of depth and movement. The overall effect is a futuristic, digital aesthetic.

What next?



Geopolitical Issues

Position documents from the EC

- Coordinated Plan on Artificial Intelligence (7th December 2018 - Updated in 21st April 2021)
 - AI public-private partnership and more financing for start-ups and innovative small and medium-sized enterprises
 - Strengthening excellence in trustworthy AI technologies and broad diffusion
 - Adapting our learning and training programmes and systems to better prepare our society for AI
 - Building up the European data space essential for AI in Europe, including for the public sector
 - Developing ethics guidelines with a global perspective and ensuring an innovation-friendly legal framework
 - Security-related aspects of AI applications and infrastructure, and international security agenda
- White Paper on Artificial Intelligence (19th February 2021)
 - Make Europe competitive in AI
 - Make AI safe and secure

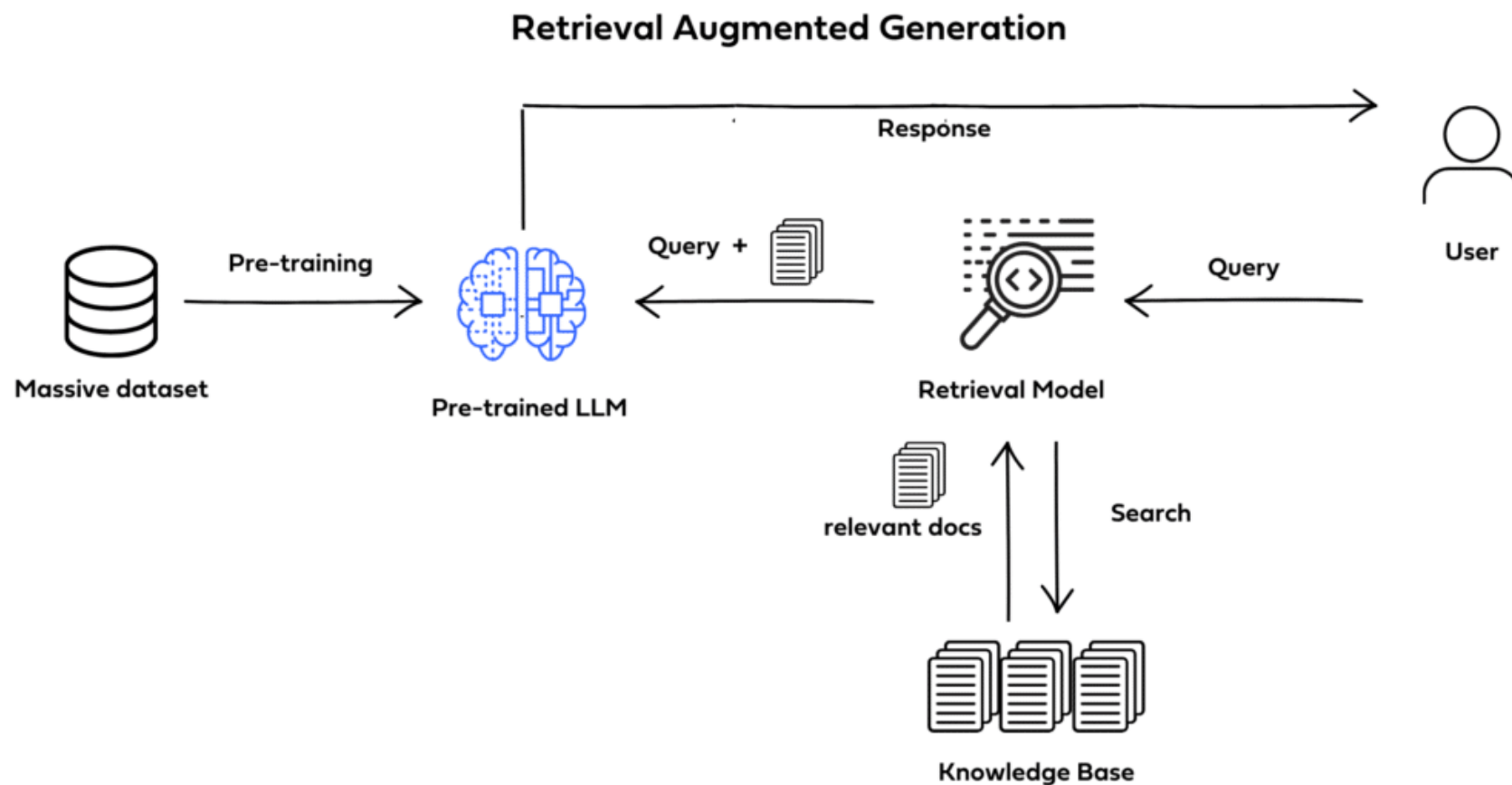
Position documents from the EC

- Artificial Intelligence Act (21st April 2021, modified and approved by the Eur. Parliament 20th March 2024)
 - Define prohibited practices
 - Define levels of risk: unacceptable risk, high risk, limited risk, and minimal or no risk
 - Define transparency for AI systems
 - Support for innovation
 - Governance and Implementation
- Digital Markets Act (approved December 15th, 2021)
 - Aims at ensuring a level playing field for all digital companies, regardless of their size, e.g.:
 - Lays down clear rules for big platforms which aim to stop them from imposing unfair conditions on businesses and consumers:
 - ranking services and products offered by the gatekeeper itself higher than similar services or products offered by third parties
 - not giving users the possibility of uninstalling any preinstalled software or app.
- Digital Services Act (approved Jan 20th, 2022)
 - Aims at creating a safer digital space for digital users and companies, by protecting fundamental rights online.
 - Regulate the trade and exchange of illegal goods, services and content online
 - Regulate algorithmic systems amplifying the spread of disinformation.
 - Give people more control over what they see online, deciding whether to allow targeted advertising or not
 - Have clear information over why specific content is recommended to them.

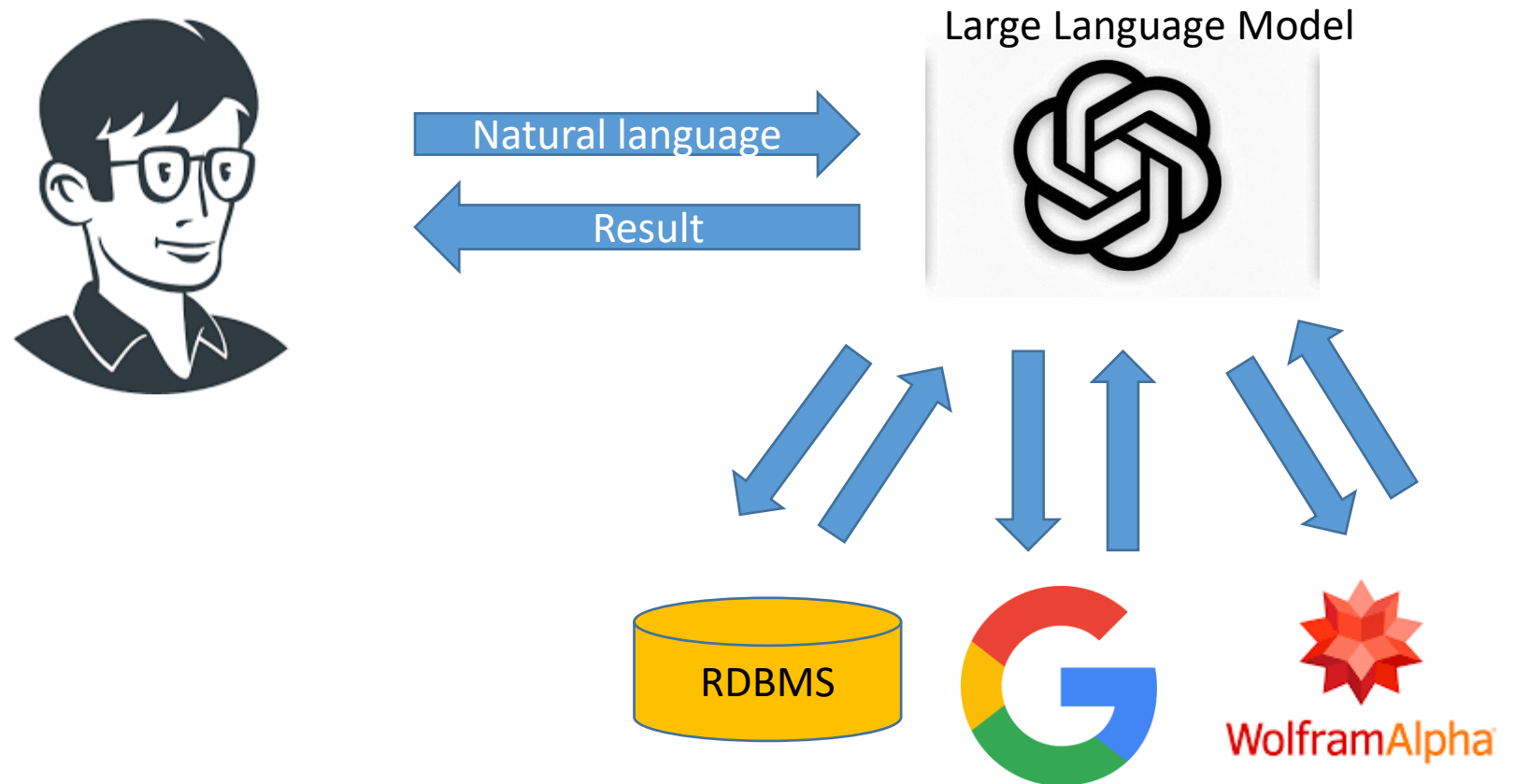
The background of the image is a dark blue, almost black, space filled with a complex network of glowing red nodes and thin, light red lines connecting them. The nodes are small, bright red spheres of varying sizes, some appearing as single points and others as small clusters. The lines are thin and create a web-like structure that fills the entire frame, giving the impression of a digital or molecular network. The overall effect is one of depth and connectivity.

From models to systems

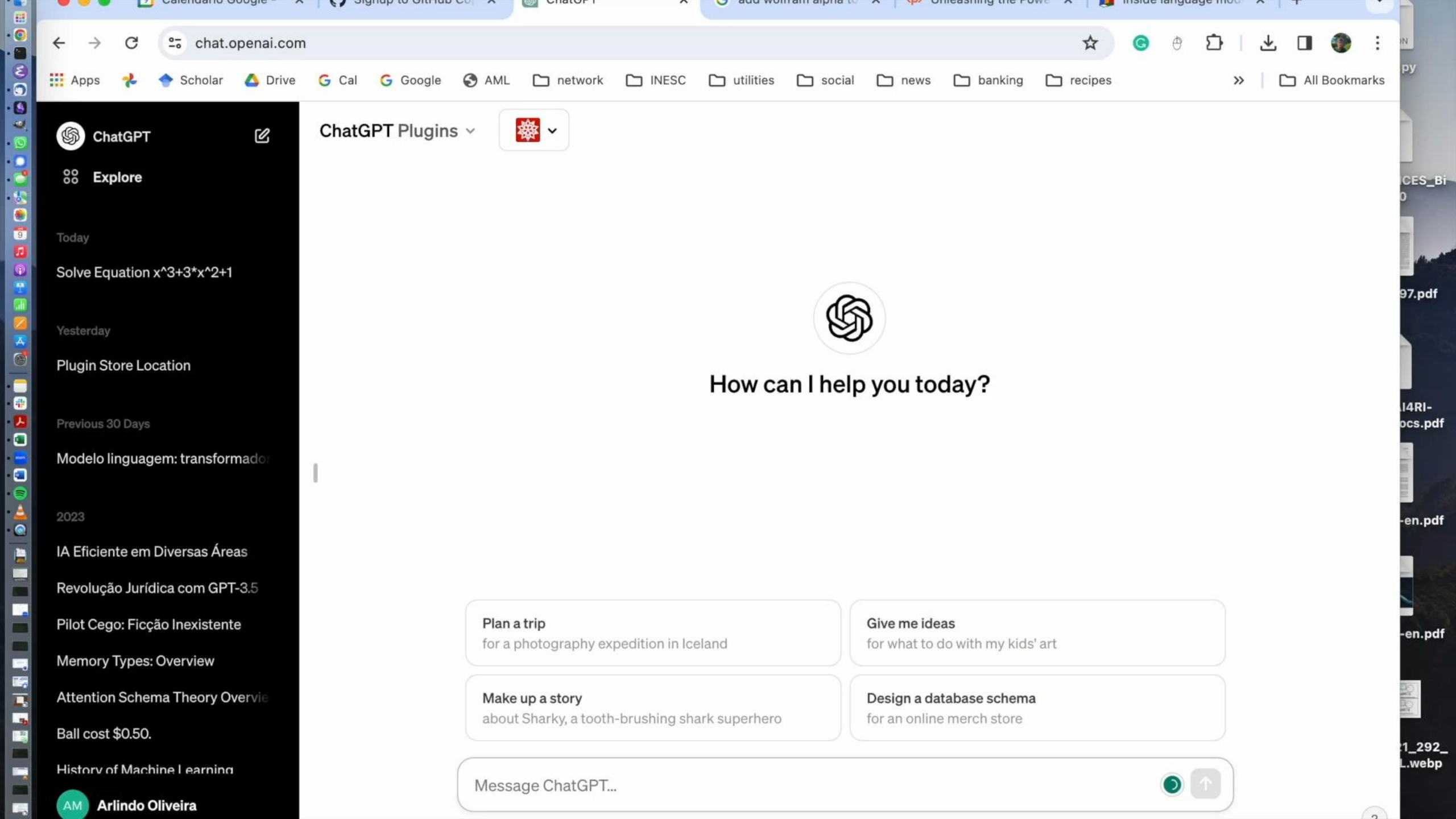
Retrieval augmented generation



The power of LLMs plus symbolic data processing systems (RAG)



- define functions like `send_email(to: string, body: string)`, or `get_current_weather(location: string, unit: 'celsius' | 'fahrenheit')` and let the LLM call them
- convert "Who are my top customers?" to a function call `get_customers(min_revenue: int, created_before: string, limit: int)` and call your internal API
- Extract structured data from text, e.g., define a function called `extract_data(name: string, birthday: string)`, or `sql_query(query: string)`



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 :

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

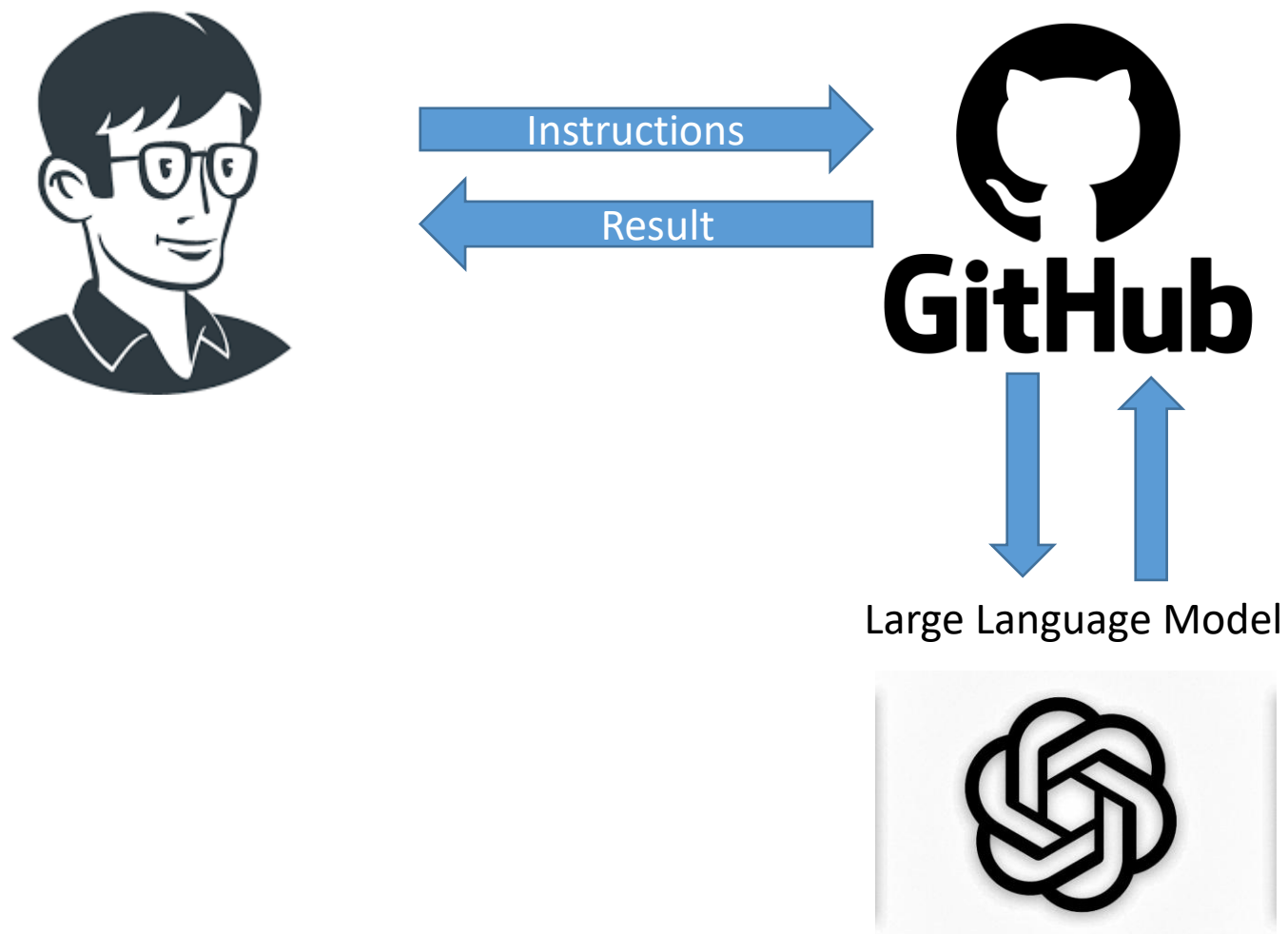
Substituting the values:

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



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

Integration with tools and platforms





Search



EXPLORER

✓ NO FOLDER OPENED

You have not yet opened a folder.

Open Folder



> OUTLINE

> TIMELINE

runge_kutta.py X



The background is a dark blue, almost black, space filled with a complex network of glowing red dots and thin, light red lines connecting them. The dots vary in brightness, with some appearing as sharp, intense red points and others as softer, blurred glows. The lines form a web-like structure, with some areas being more densely connected than others, creating a sense of depth and interconnectedness. The overall effect is reminiscent of a neural network, a data visualization, or a molecular structure.

Dual process theories and the power of reasoning

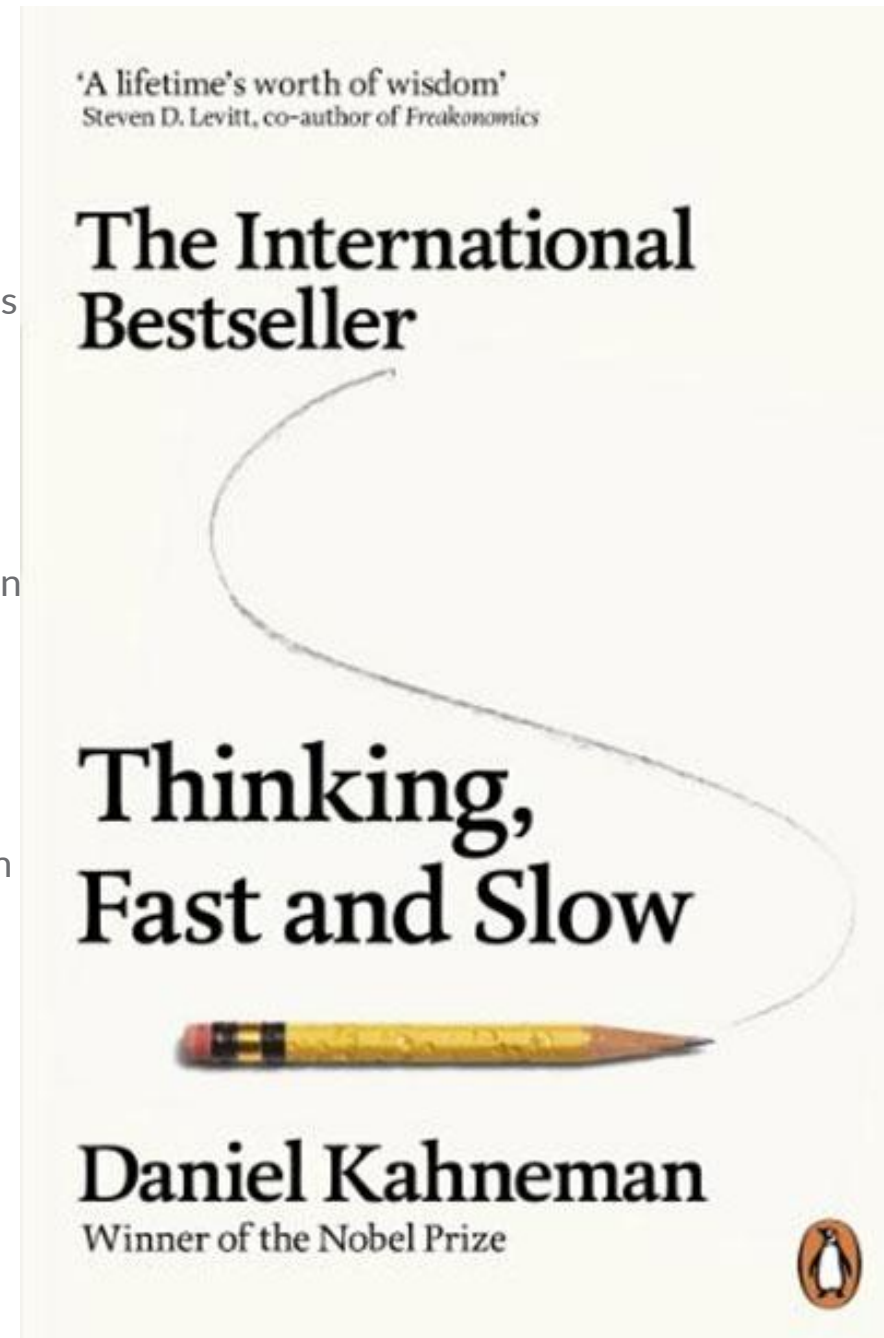
Dual process theories

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?



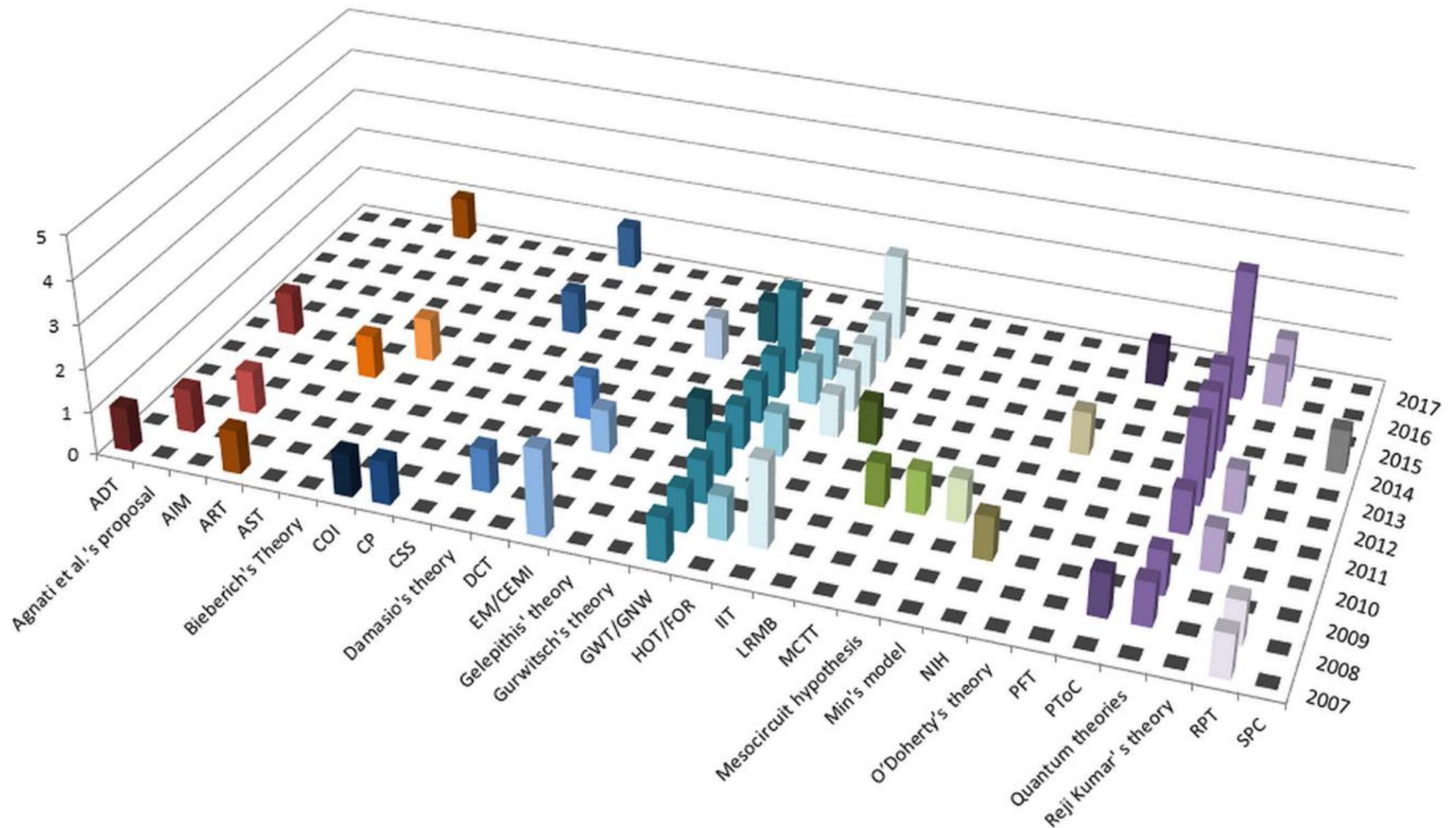
Okay, so there's this problem about transporting fuel through a 200-mile canal. The ship can carry up to 10,000 gallons, uses 50 gallons per mile, and there are tanks every two miles that can store fuel. The goal is to get as much fuel as possible from the start to the end of the canal. Hmm, this reminds 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?

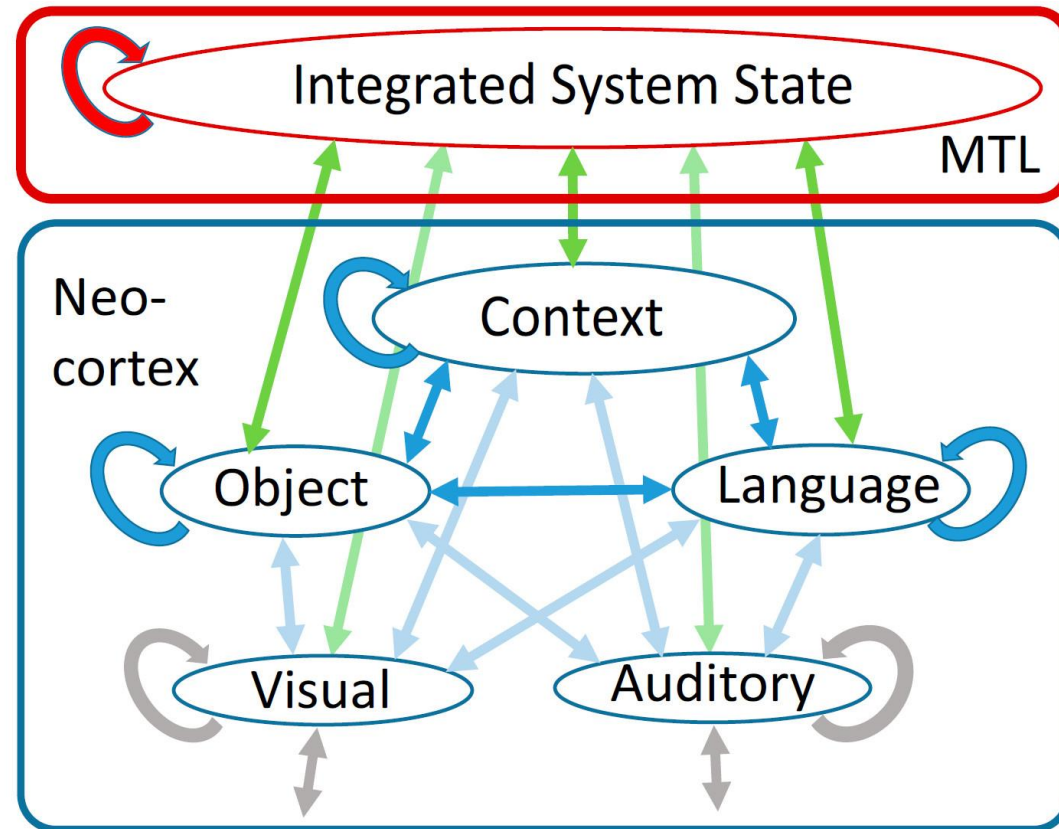
Reasoning in LLMs

- **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.



Sattin, D., Magnani, F. G., Bartesaghi, L., Caputo, M., Fittipaldo, A. V., Cacciato, M., ... & Leonardi, M. (2021). Theoretical Models of Consciousness: A Scoping Review. *Brain sciences*, 11(5), 535.

Future system architectures



A



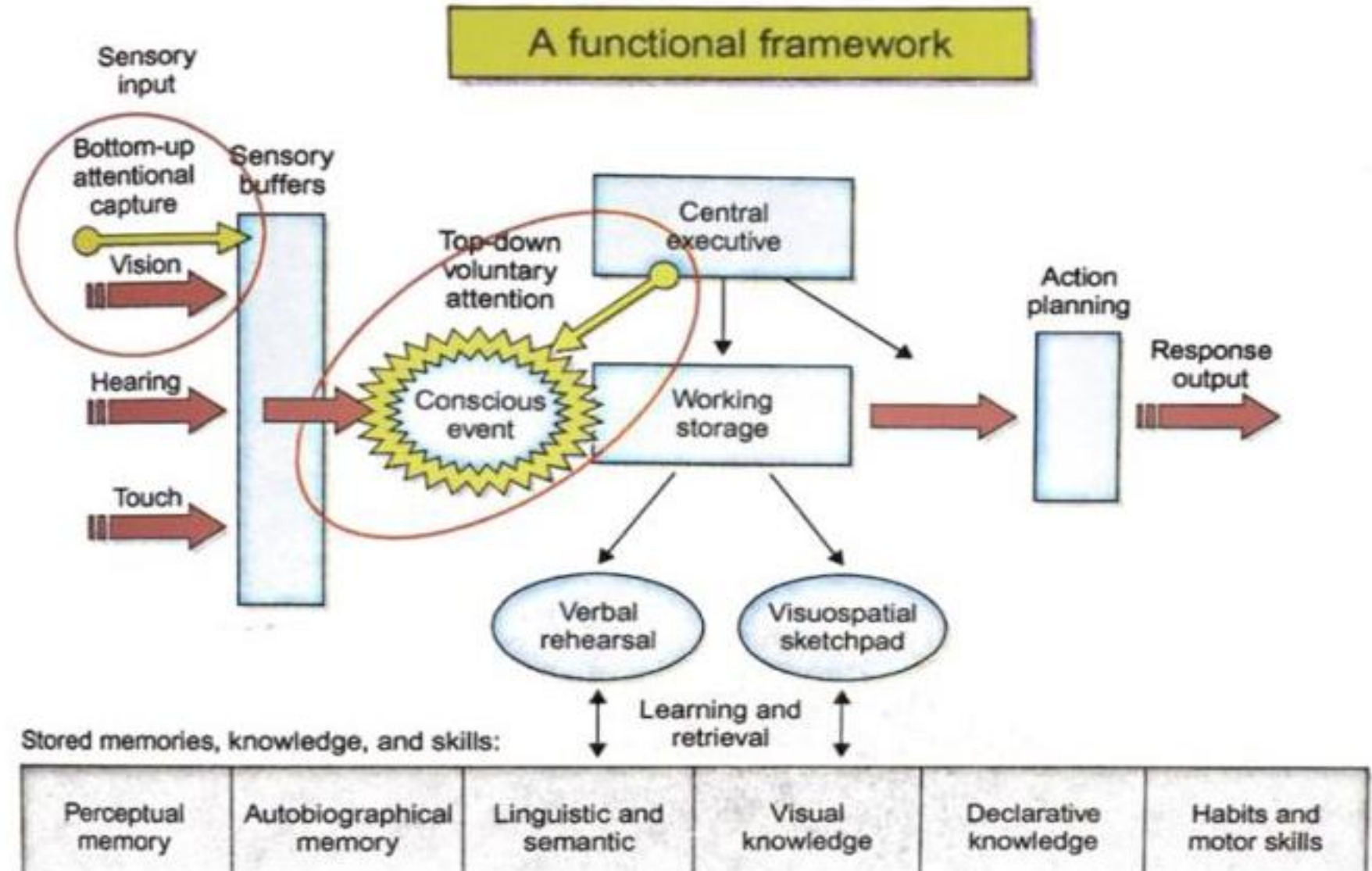
“...the **bat** hit...”

B



“...the **numbat** eats termites...”

Global Workspace Theory (GWT)



Baars, B. J., & Gage, N. M. (2010). Cognition, Brain and Consciousness: Introduction to Cognitive Neuroscience. Burlington, MA, USA: Academic Press.

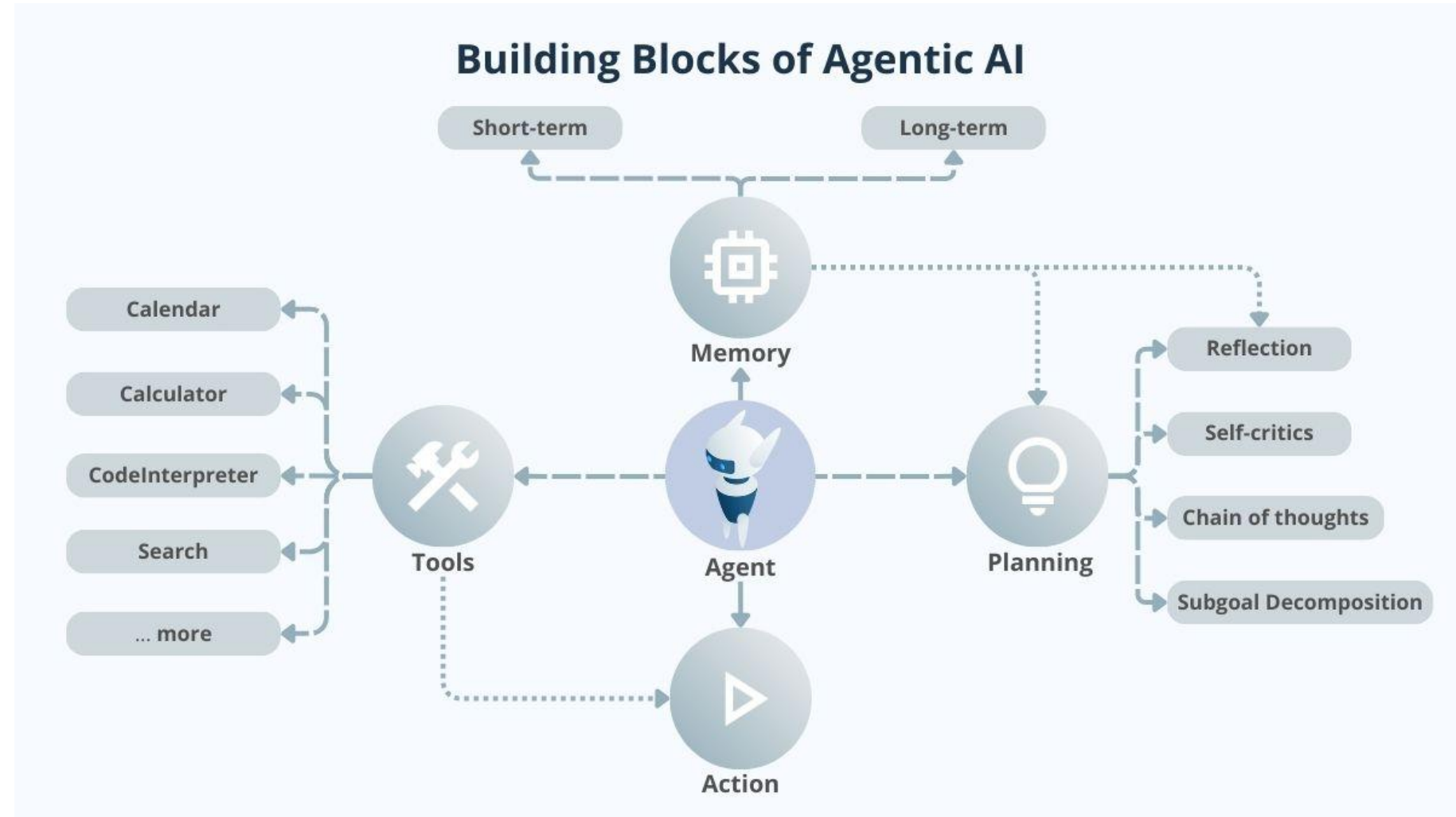


Agentic Artificial Intelligence

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

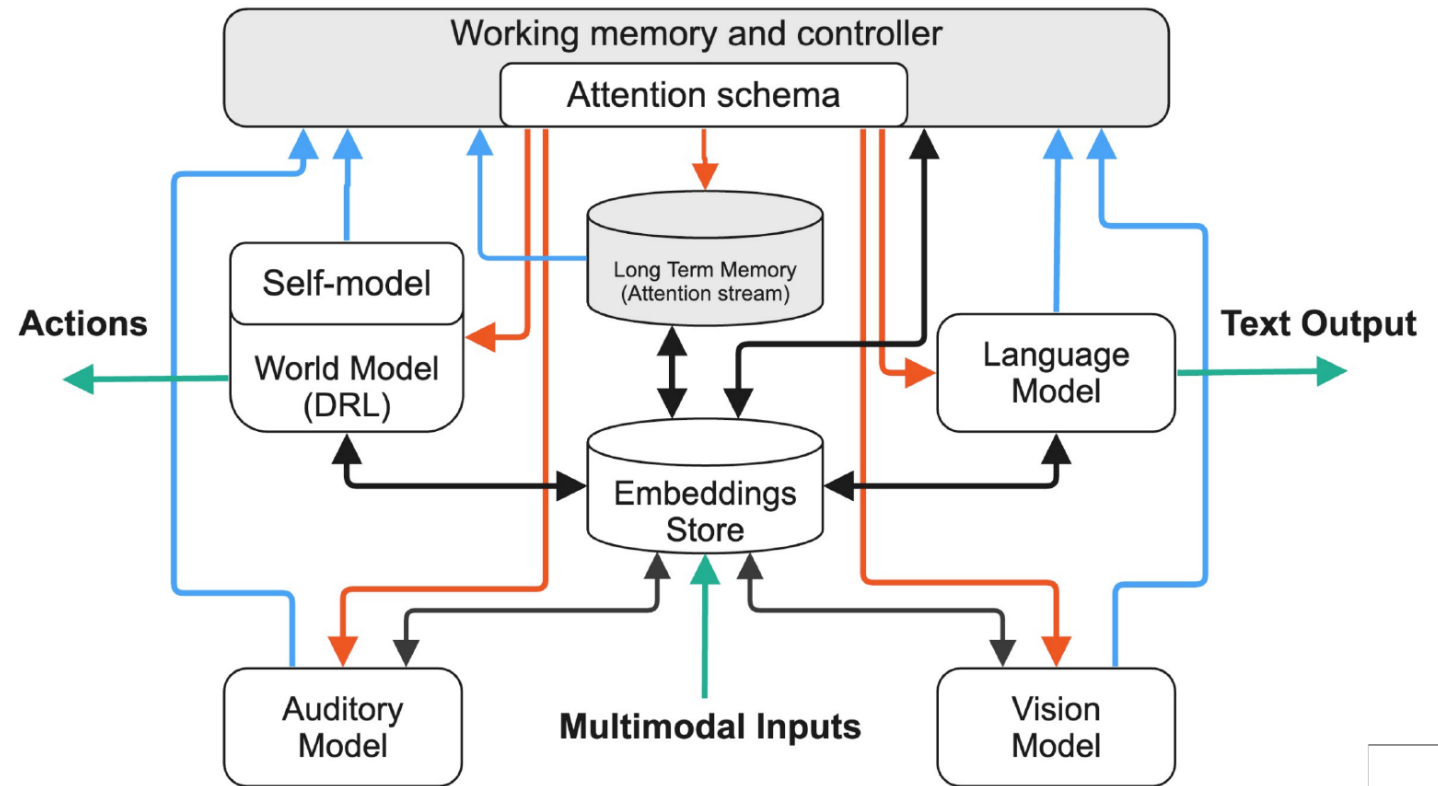


Figure 2: The DeepThought architecture: green arrows represent inputs and outputs, blue arrows represent module outputs, red arrows represent attention and control signals, and black arrows represent bidirectional interactions between modules and the embeddings store. The working memory and controller,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



Artificial General Intelligence

Pipe dream or the inevitable future?

Objections

- **The immortal soul objection:**

Man is the sole owner of an immortal soul, and no machine can ever have one.

- **The consciousness objection:**

Only humans can be conscious.

- **The technological impossibility objection:**

It is outside the reach of human technology, forever.

- **The physical support objection:**

Nervous systems are continuous, use quantum mechanics to function, are chaotic

- **The mathematical objection:**

Brains are not Turing equivalent

- **Various disabilities objection**

Cannot create anything new

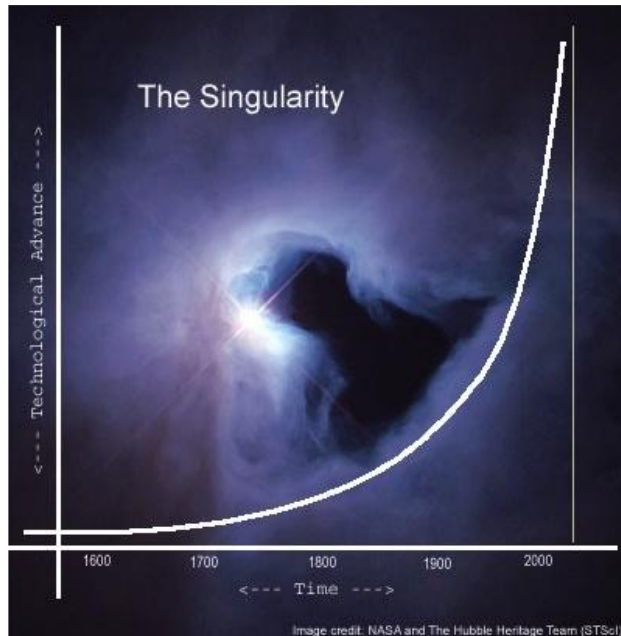
Cannot love or be loved

Cannot feel pain or joy

Speculations

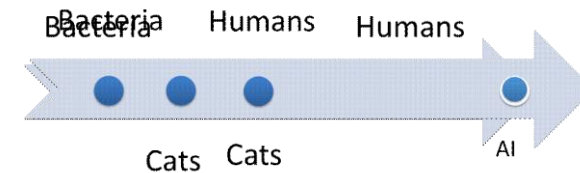
“The ever-accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential **singularity** in the history of the race beyond which human affairs, as we know them, could not continue”

- Von Neumann, quoted by Stanislaw Ulam, 1952



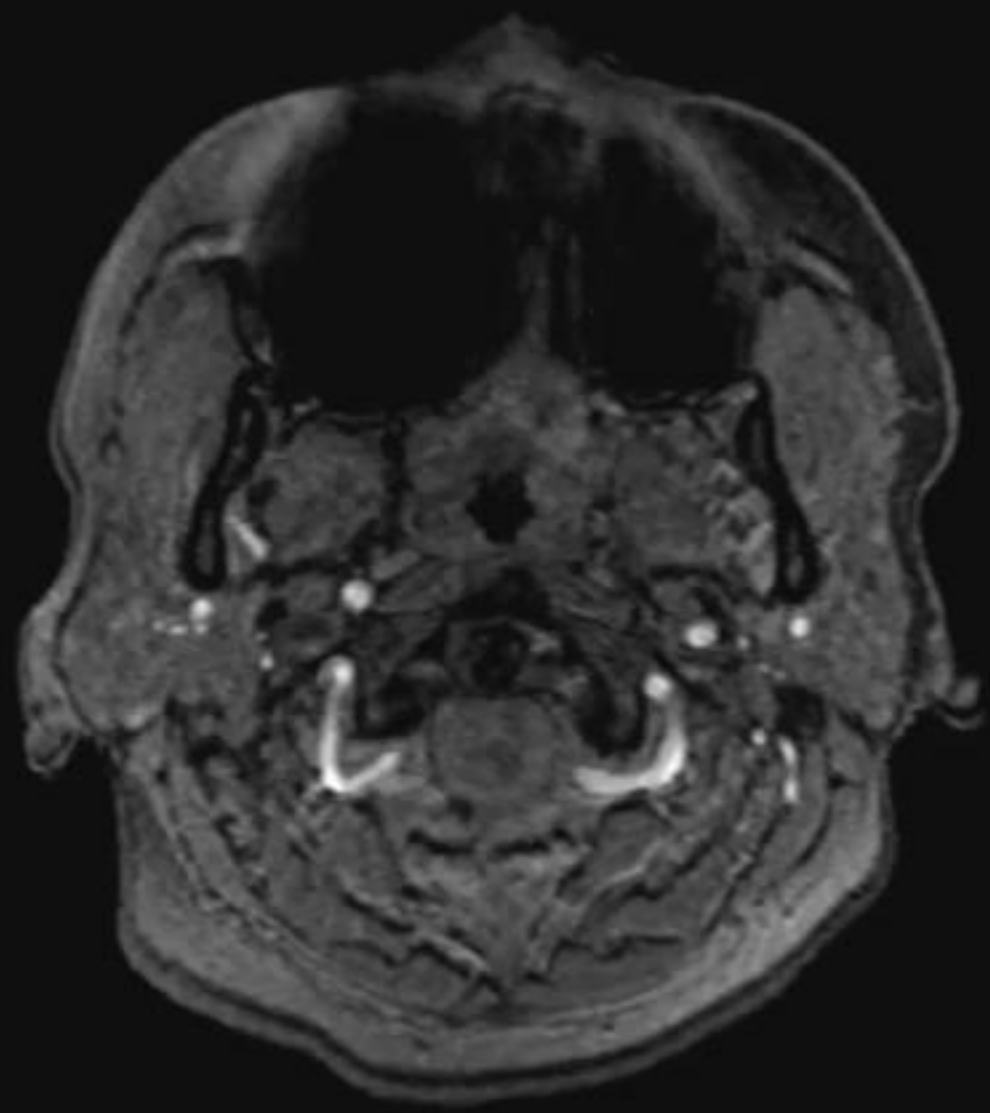
SUPERINTELLIGENCE

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

