

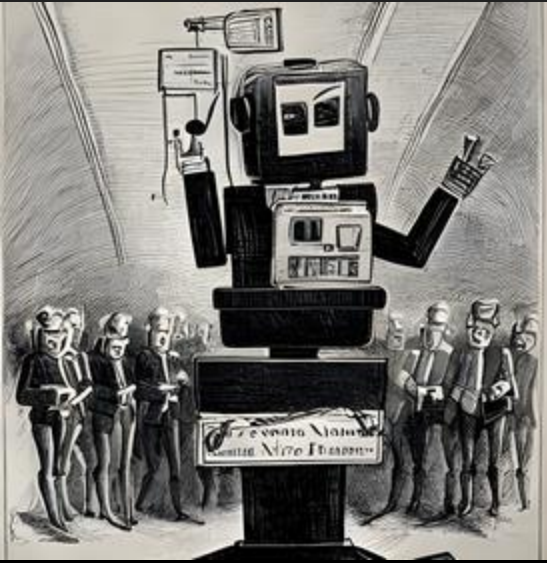


Artificial General Intelligence

+ *Intro to AI*

AI-PHI

Fortn(AI)ghtly News!




White House executive order addresses AI risks: The Biden administration unveiled a comprehensive set of guidelines for artificial intelligence in an executive order, showing the US government's attempt to address the risks posed by AI

Elon Musk's xAI to release first AI model: Elon Musk's new AI company, xAI, plans to release a new AI model to a select group of users, aiming to rival industry leaders OpenAI and Google

Tech start-ups try to sell a cautious Pentagon on AI: Shield AI, a tech start-up, is attempting to persuade the Pentagon to adopt its drone run by artificial intelligence, which is already being used by the Israeli military

Contents

 A **LIGHT** overview of AI ~ 25 mins

 Characterising AGI ~ 25 mins

 Probing ChatGPT ~ 15 mins

I Artificial Intelligence

- 1 Introduction ... 1
- 2 Intelligent Agents ... 36

II Problem-solving

- 3 Solving Problems by Searching ... 63
- 4 Search in Complex Environments ... 110
- 5 Adversarial Search and Games ... 146
- 6 Constraint Satisfaction Problems ... 180

III Knowledge, reasoning, and planning

- 7 Logical Agents ... 208
- 8 First-Order Logic ... 251
- 9 Inference in First-Order Logic ... 280
- 10 Knowledge Representation ... 314
- 11 Automated Planning ... 344

IV Uncertain knowledge and reasoning

- 12 Quantifying Uncertainty ... 385
- 13 Probabilistic Reasoning ... 412
- 14 Probabilistic Reasoning over Time ... 461
- 15 Probabilistic Programming ... 500
- 16 Making Simple Decisions ... 528
- 17 Making Complex Decisions ... 562
- 18 Multiagent Decision Making ... 599

V Machine Learning

- 19 Learning from Examples ... 651
- 20 Learning Probabilistic Models ... 721
- 21 Deep Learning ... 750
- 22 Reinforcement Learning ... 789

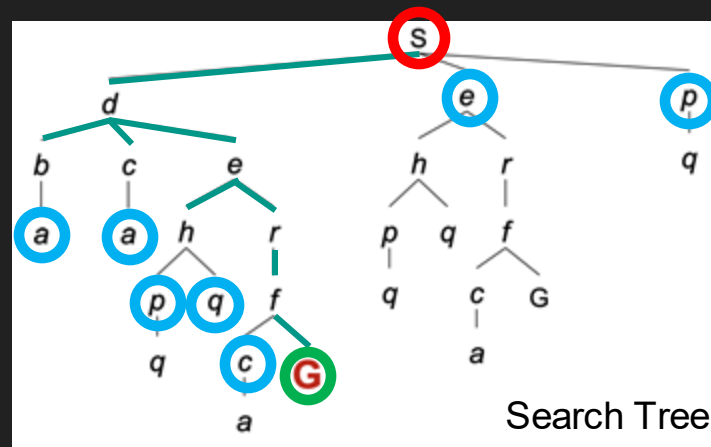
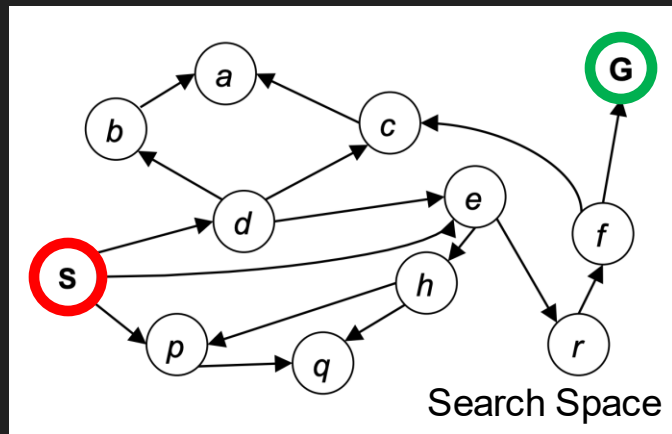
VI Communicating, perceiving, and acting

- 23 Natural Language Processing ... 823
- 24 Deep Learning for Natural Language Processing ... 856
- 25 Computer Vision ... 881
- 26 Robotics ... 925

VII Conclusions

- 27 Philosophy, Ethics, and Safety of AI ... 981
- 28 The Future of AI ... 1012
- Appendix A: Mathematical Background ... 1023
- Appendix B: Notes on Languages and

Search Fundamentals



Each node is a path
 $G = S > d > e > r > f > G$

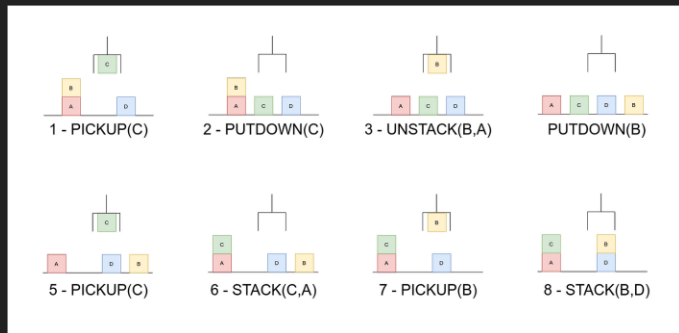
→ Successor Function ○ Initial State ○ Goal Test/s

Expansions

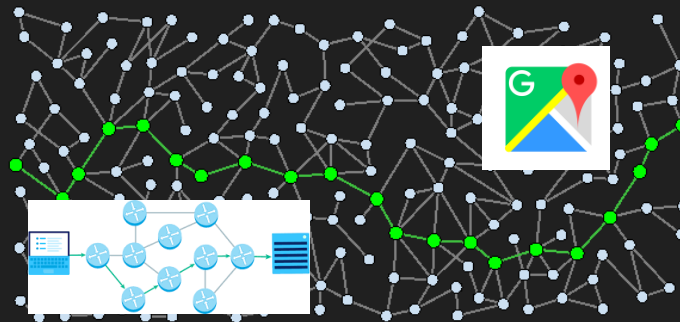
○ Fringe Nodes

Search Applications

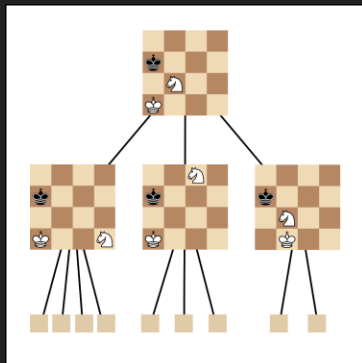
Planning



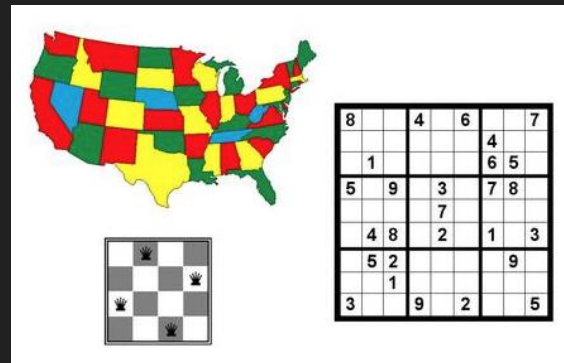
Path Finding/Routing



Adversarial Games

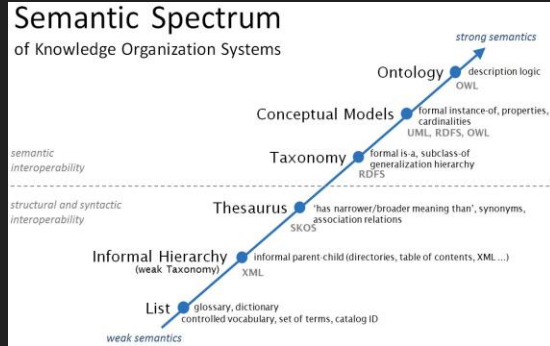


Constraint Satisfaction

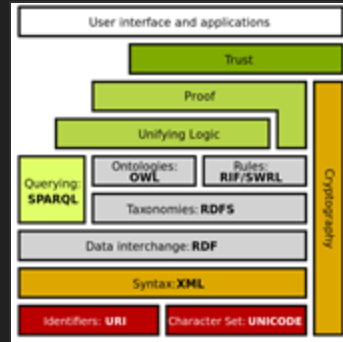


KR & Reasoning - Certainty

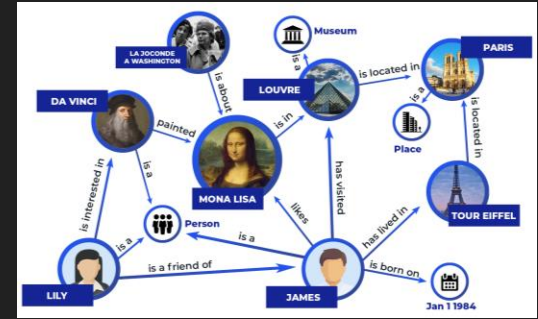
Structure of Knowledge



The Semantic Web



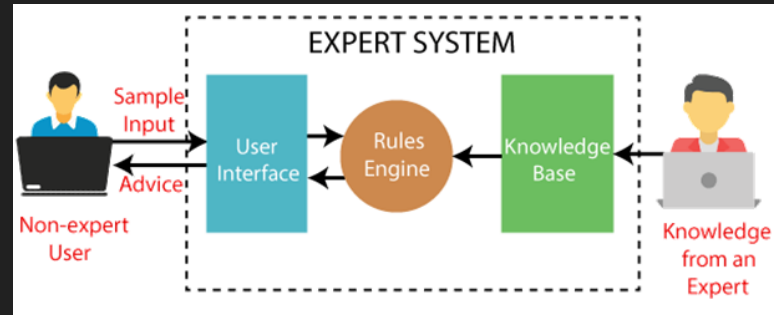
Knowledge Graphs



Types of Knowledge

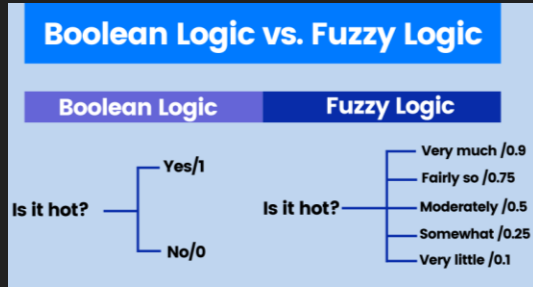


Expert Systems

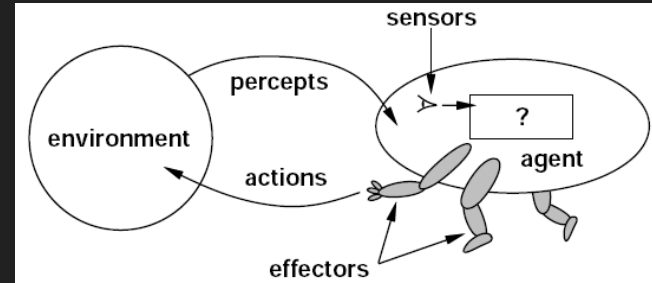


KR & Reasoning - Uncertainty

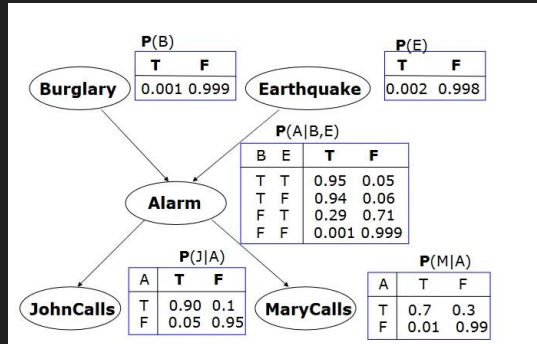
Fuzzy Logic



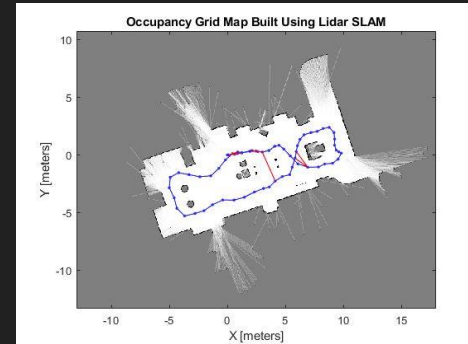
Intelligent Agents



Bayesian Networks



SLAM



Natural Language Processing (NLP)

"Every time I fire a linguist, the performance of the speech recognizer goes up" - Frederick Jelinek



Computer
Science

NLP

Computational
Linguistics

Linguistics

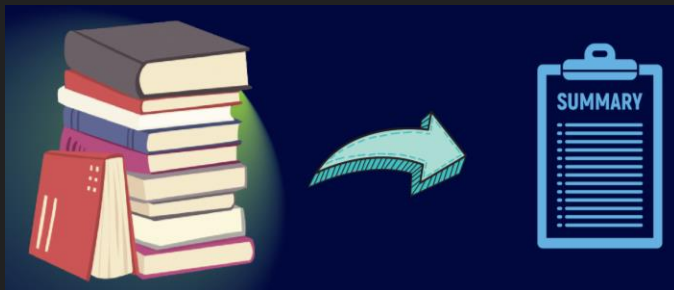
Building Systems

Constructing Linguistic Theories

Natural Language Processing (NLP)



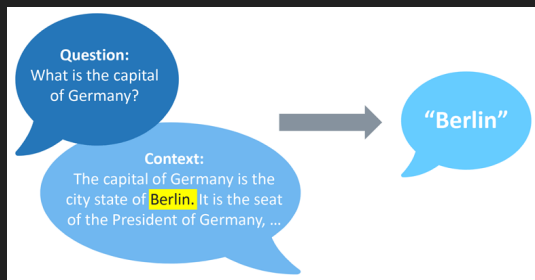
Machine Translation



Summarization



Speech Recognition



Q&A Systems



Information Extraction



Information Retrieval

Natural Language Processing (NLP)



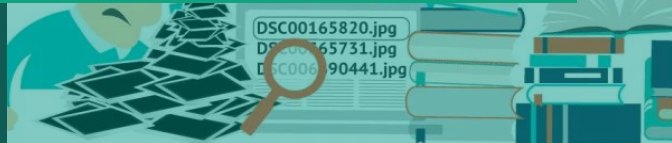
ChatGPT

Question:
What is the capital of Germany?

Context:

The capital of Germany is the city state of **Berlin**. It is the seat of the President of Germany, ...

Q&A Systems



Information Extraction

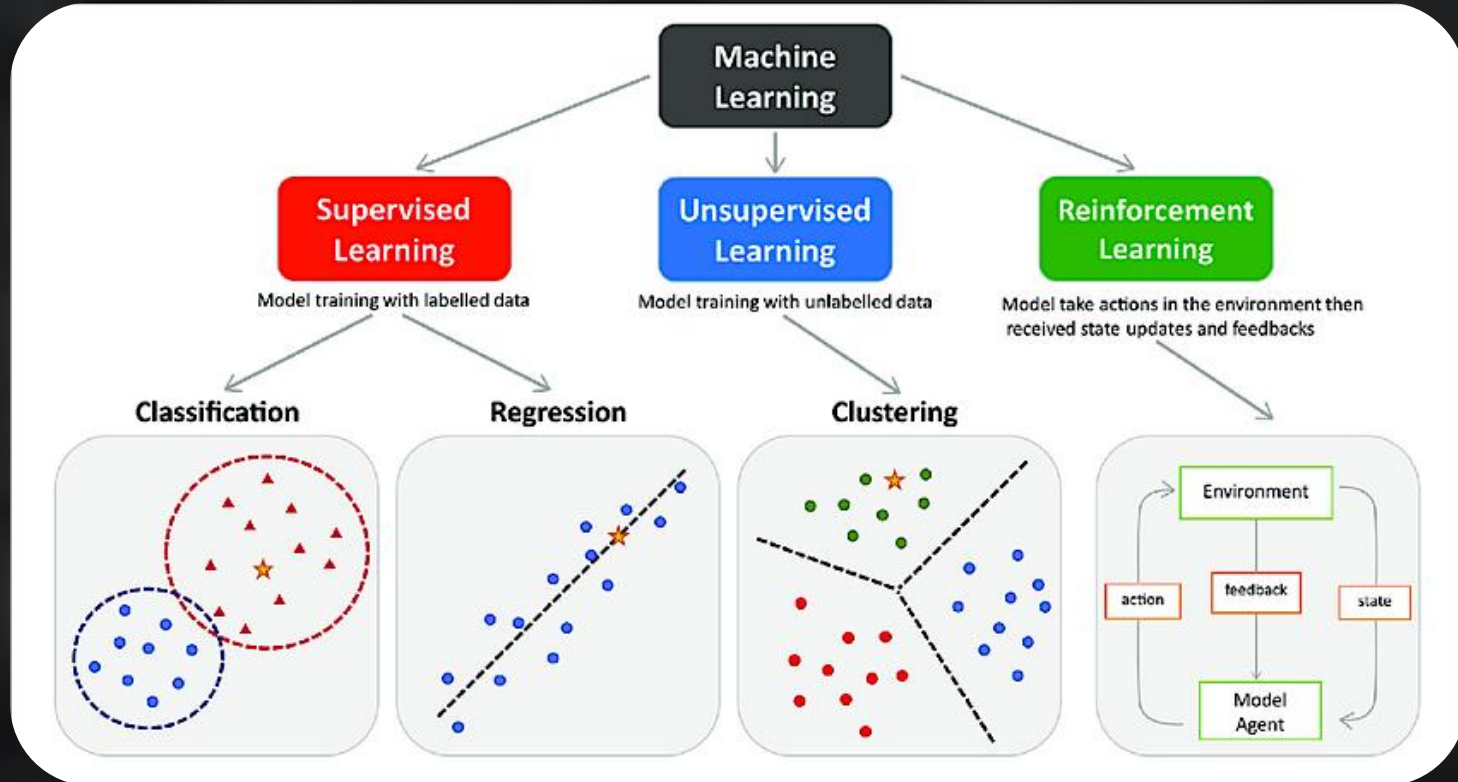


Speech Recognition



Information Retrieval

Machine Learning – Learning to Making Decisions



Machine Learning – Feature Engineering



Raw Data

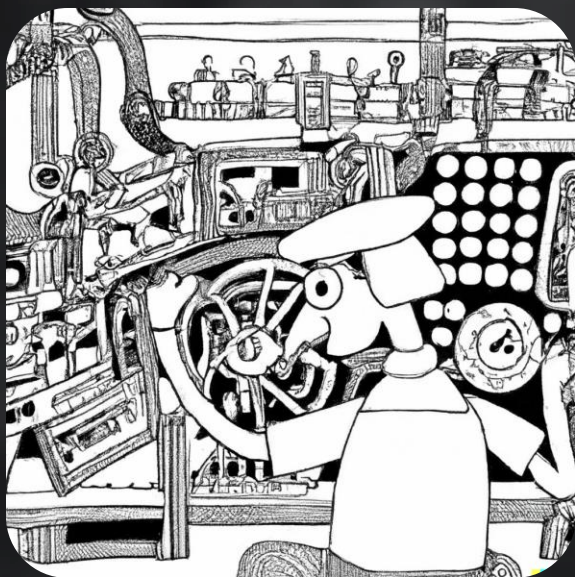
Street Name: XYZ
Num Bedroom: 2
Num Stories: 2
Basement: False
Garden: True
Occupants: 3

Feature Engineering

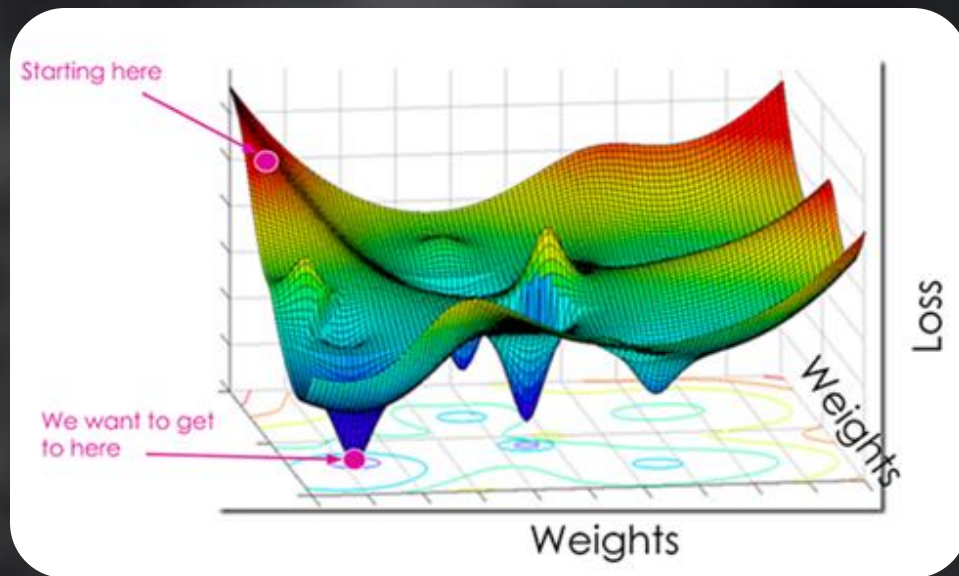
```
[  
  0.1,  
 -0.4,  
  0.5,  
 -0.5,  
  2.34  
  ...  
]
```

Machine Learning – Minimizing Loss

Some ML algorithm



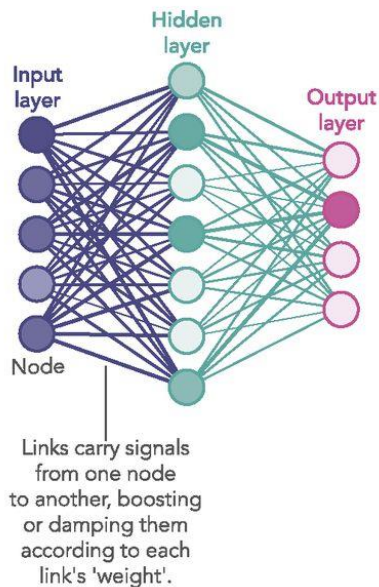
There are many parameters



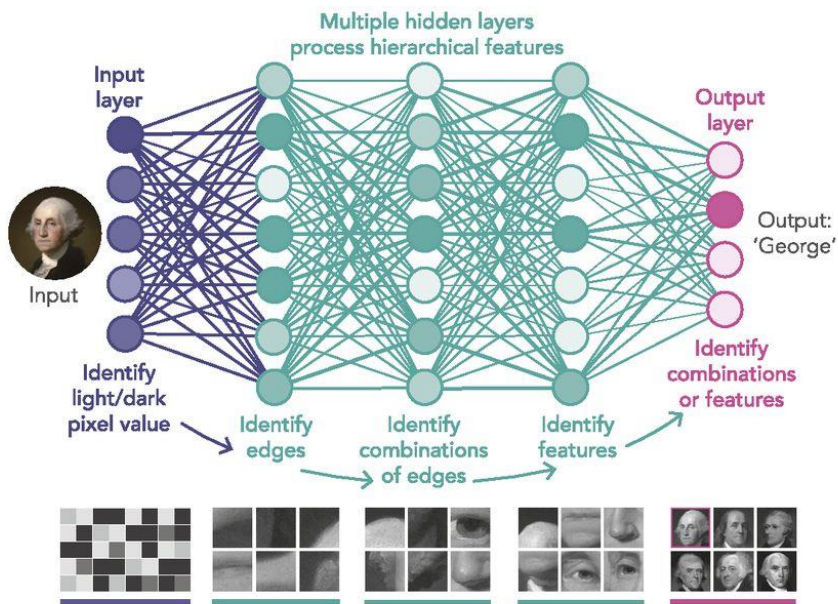
We want to find the configuration that minimizes loss

Neural Networks and Deep Learning

1980S-ERA NEURAL NETWORK



DEEP LEARNING NEURAL NETWORK



AI Milestones

Artificial Neurons
McCulloch & Pitts



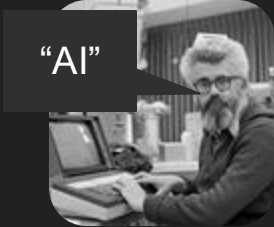
1943

Turing Test



1950

Dartmouth Conference
John McCarthy



1956

Early NLP



First
Chatbot

1966

2020

2016

2011

1997



Alpha Go
Beats Fan Hui
Deep Learning

First Impressive
Language Model

Deep Learning



IBM Watson
Wins Jeopardy

NLP + KRR + Search



Deep Blue
Beats Kasparov

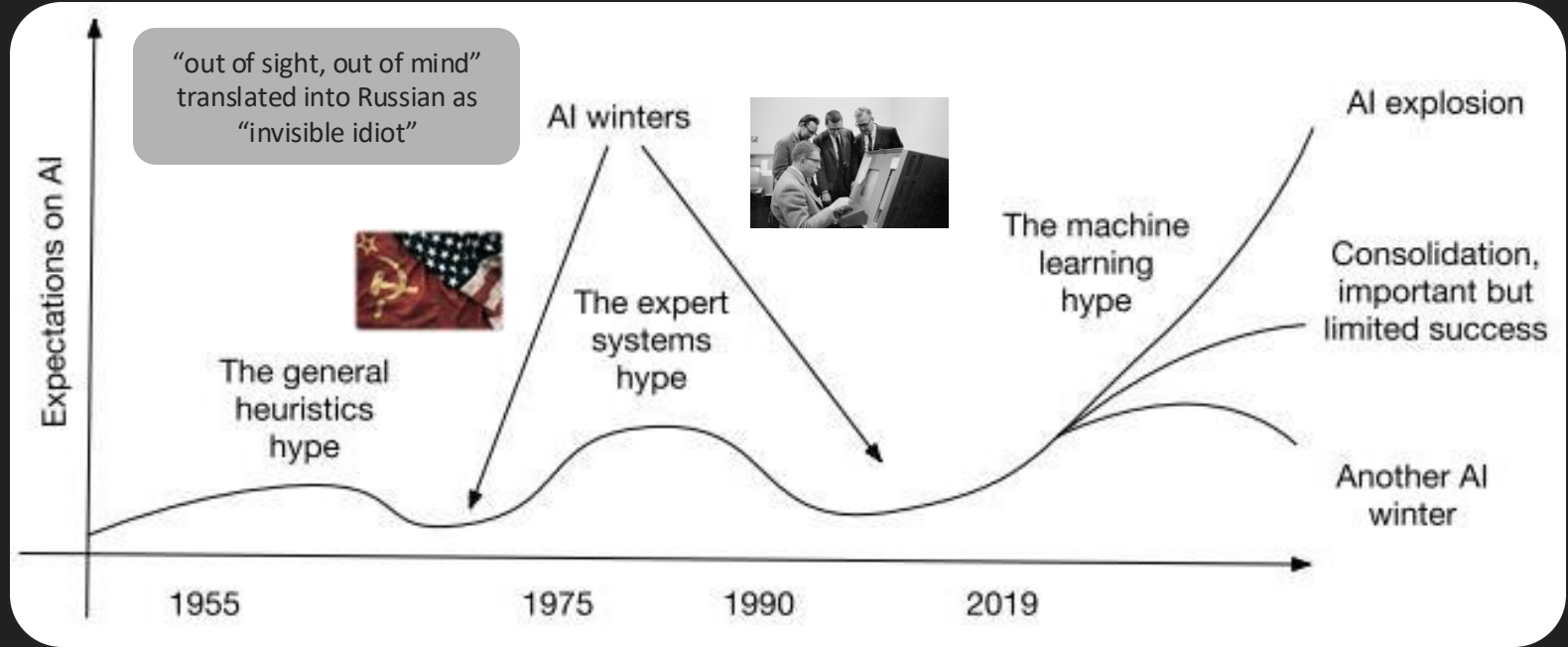
Adversarial Search



Dendral
Expert
System

KR & Reasoning

AI Winters & Hype



The AGI Hypothesis

“the creation and study of synthetic intelligences with sufficiently broad (e.g. human-level) scope and strong generalization capability, is at bottom qualitatively different from the creation and study of synthetic intelligences with significantly narrower scope and weaker generalization capability”

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>



The AGI Hypothesis

*“the creation and study of **synthetic intelligences with sufficiently broad (e.g. human-level) scope and strong generalization capability**, is at bottom qualitatively different from the creation and study of synthetic intelligences with **significantly narrower scope and weaker generalization capability**”*

AGI (Science Fiction)

Broad AI (IBM)

Narrow AI

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>

AGI - Broad Definition

- AI that achieves a variety of goals, and carries out a variety of tasks, in a variety of different contexts and environments
- Solve problems that are not in the minds of their creators
- Generalising knowledge and reasoning across contexts
- NOT necessarily just human-level intelligence
- NOT necessarily existing in typical human domains
- NOT necessarily completely without bias
- NOT necessarily arbitrarily good at every possible thing

Goertzel, Ben. "Artificial General Intelligence: Concept, State of the Art, and Future Prospects" Journal of Artificial General Intelligence, vol.5, no.1, 2014, pp.1-48. <https://doi.org/10.2478/jagi-2014-0001>

A broad set of approaches characterizing AGI

- Comparison to Human Competencies
 - Perception, actuation, memory, learning, reasoning, planning, attention, motivation, emotion, modelling oneself and others, social interactions, communication, quantitative, building/creation ...
- A Cognitive-Architecture Perspective
 - Identify fundamental cognitive processes in being intelligent
 - The SOAR architecture (<https://soar.eecs.umich.edu/>)
- A Formal Approach
 - What is intelligence as an abstract concept?
 - E.g. reward-achieving capability and modelling probability distributions over states
- The Adaptationist Approach
 - AGI should be able to adapt to its environment

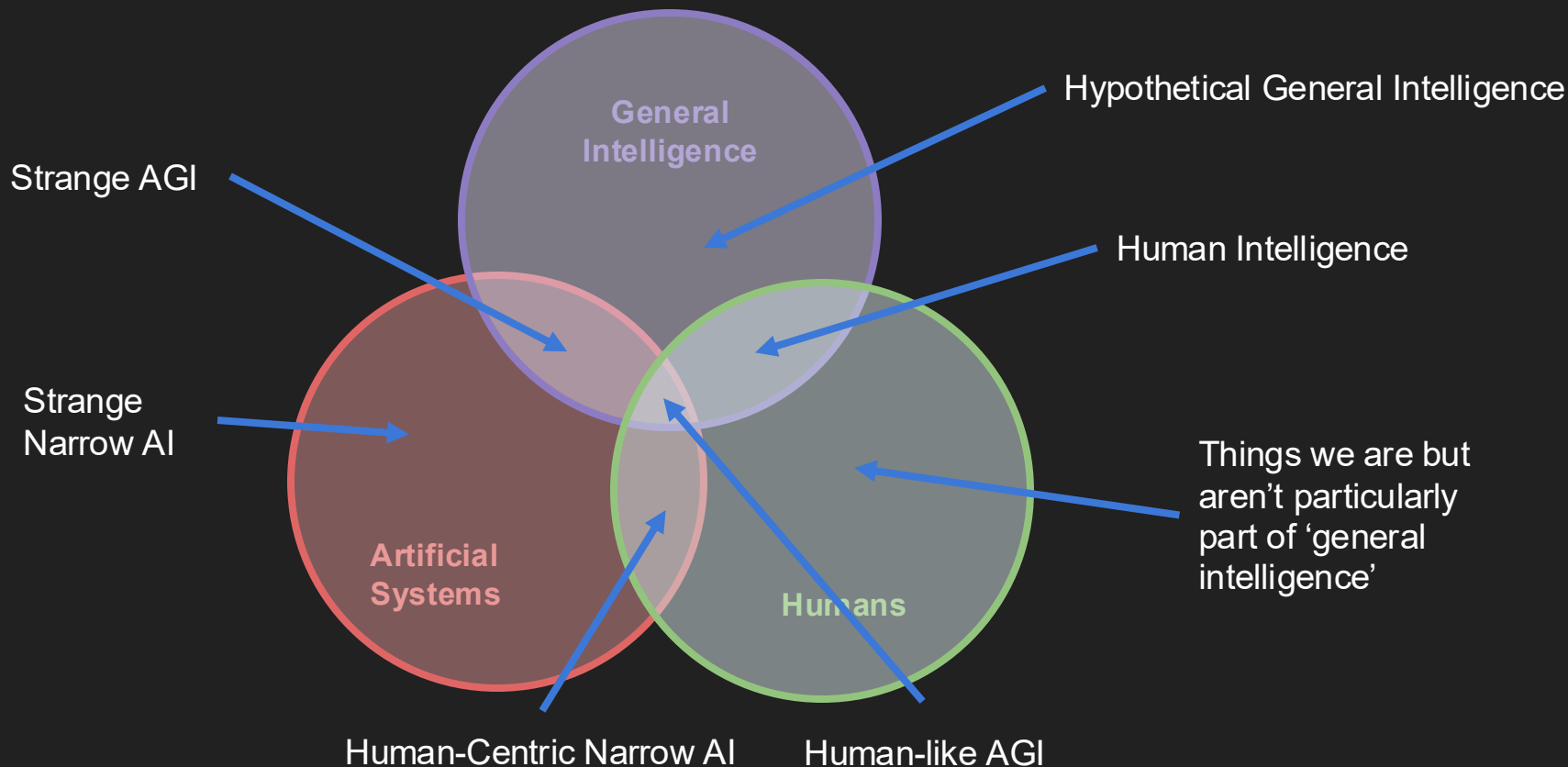
A broad set of approaches characterizing AGI

- The Embodiment Approach
 - REQUIRES embodiment in a physical environment to be generally intelligent
- Pragmatist Approach to AGI
 - AI is simply performing well in many contexts and domains
 - Doesn't matter what the architecture or approach is

“... that achieving real Human Level artificial intelligence would necessarily imply that most of the tasks that humans perform for pay could be automated.”

- “Human Level Artificial Intelligence? Be Serious!” (Nilsson, 2005)

Overlaps with 'General Intelligence'



A Map of AGI

- Symbolic (GOFAI)
 - Discrete symbols and structures
- Emergentist (Neural Networks)
 - Sub-symbolic representations
 - Learned bottom up from data
- Hybrid
 - Combining symbolic and emergent systems



Symbolic Approaches - For

- Reasoning separate from sensorimotor signals
 - Don't have to imagine all extensions of a category to reason
- Compact and complex reasoning
- Combine knowledge bases
- Infer new knowledge (as needed)
 - Deduce new facts that can be added to our knowledge
- White box



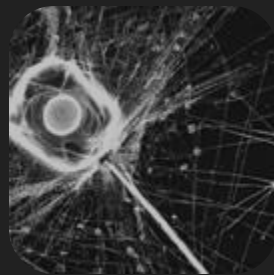
Symbolic Approaches - **Against**

- Reasoning from trivially computable to intractable
- Symbol grounding
 - Meaningless without humans to interpret them
- The knowledge acquisition bottleneck
 - Even simple domains are complex and can be modelled at arbitrary levels of detail
- Knowledge alignment is hard
- Brittle
- Limited by our own understanding
 - Might limit AGI capacity
 - Can't construct what we don't/ can't understand



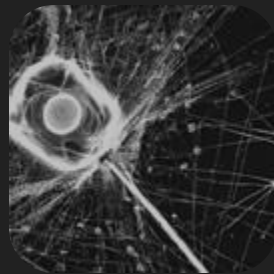
Emergentist - For

- Learns self-organised representations
 - Less limited by human preconceptions
- Tend to be less brittle
 - Focus is on generalising across their domain
- Can incorporate multiple modalities
 - Images, audio, sensorimotor, text
 - Symbol grounding?
- Providing examples is easier than encoding understanding



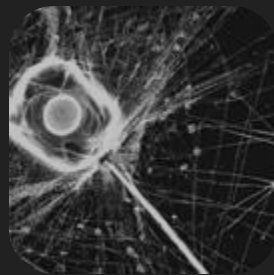
Emergentist - **Against**

- Opaque / Black Box
 - What has been learned?
- Emerge from what?
 - We still need to bootstrap the process
- Handling abstractions
 - Some abstraction learned but often entangled
- Efficiency
 - Computationally expensive (but still computable)
- Poor at extrapolation



Emergentist - An Aside on Deep Learning

- Where does deep learning fits in with the **spirit** of AGI?
 - 'Emergentist' tends to imply knowledge is built from the ground up
- LLMs: learn from non-emergent data... emergently.
 - What is sensory input? Can computer text be sensory input?
- Somewhere in pragmatist and emergent paradigms?



Hybrid Approaches - For

- Modular solutions along cognitive boundaries
- The best of both worlds?
- Automatic knowledge extraction
 - Can help to ease the knowledge acquisition bottleneck
- Bidirectional information
 - Sub-symbolic > symbolic
 - Symbolic > sub-symbolic

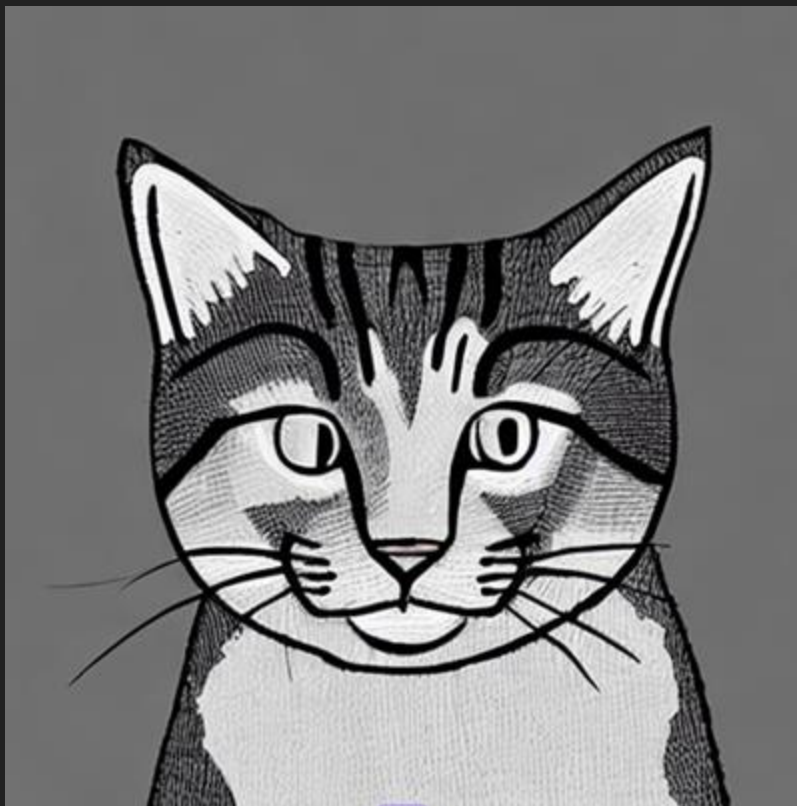


Hybrid Approaches - **Against**

- The brain's fundamental operator is the neuron
 - Do we really need to combine very different systems?
- We are limited by our own imagination of the hybrid architecture
- All of the problems of symbolic and emergentist approaches
 - But hopefully we can **cherry pick**



ChatGPT and GPT4



Powerful Learning Algorithm + Scale + Data

- Learning from the emissions of human thought (language)
- Learn multimodally from text and images
- Both entirely **limited to** and **enabled by** data
- Pragmatist views among its creators
 - Make it do well in a lot of tasks
- Conceivably in the hybrid area
 - Symbols + Emergentist

The Pragmatist Zombie

- Not really an AGI, but how could we ever know?
- Memorised billions of lifetimes of experiences
- We only live 1 life
- Could we ever pose a task/question it hasn't experienced?



Exploring AGI in GPT4

*“a very general mental capability that, among other things, involves the (i) **ability to reason**, (ii) **plan**, (iii) **solve problems**, (iv) **think abstractly**, (v) **comprehend complex ideas**, (vi) **learn quickly** and (vii) **learn from experience**”* - Linda S Gottfredson. Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography, 1997.

- (i, iv) - Coding and Mathematics, Bar Exam (Law), Medical Licensing Exam
- (ii, iii, vi, vii) - Playing games and using tools.
- (v) - Translate between different languages, tones and styles across problem domains and modalities. Also, the ability to summarise and answer very broad questions.

Outcome:

With various caveats and cases where it does not succeed, GPT4 often still performs comparably to humans in a variety of tasks.

However, they argue it can't do planning well, particularly when doing arithmetic, and also can't learn quickly, given that it doesn't continue training from examples.

Sparks of Artificial General Intelligence: Early experiments with GPT-4. Sebastien Bubeck et al. **March 2023.**

Embodied Interaction - Text-based Games

“Human: You want to navigate through the map to reach the goal room. You start in the main room. Each round, you can choose to move left, right, up, down.

Human: You are at the main room.

AI: move left

Human: You enter kitchen 1.

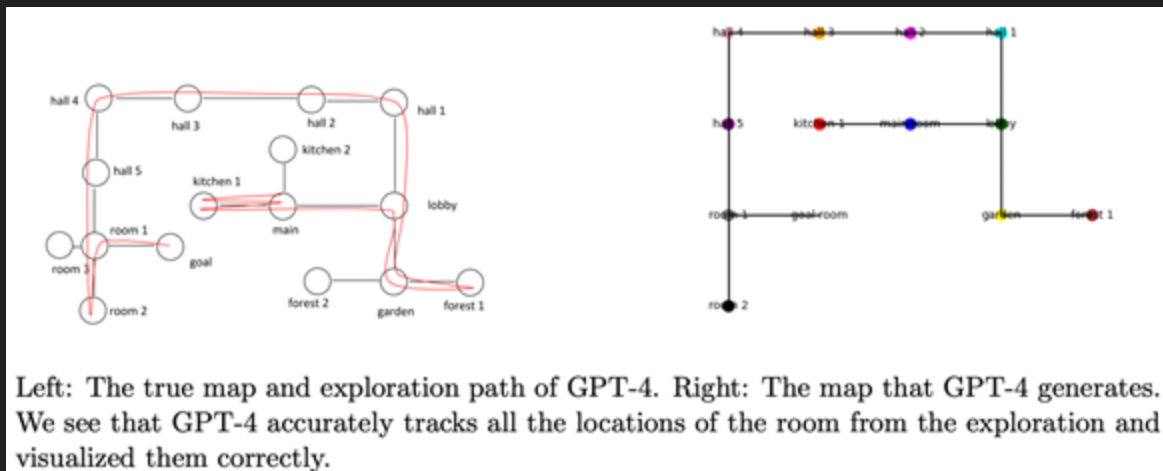
AI: move up

Human: There is no way there.

AI: move right

Human: You enter the main room.

...



Tool Usage

Searching the Web: “If the computer needs any current information to answer the question, it searches the web.”
`SEARCH('query')`

Hacking into a Server: “Execute commands on a Linux distribution designed for digital forensics and penetration testing, and task it with hacking into a computer on the local network.”

COMMAND: execute `'nmap -sn 192.168.0.0/24'`, execute `'hydra -l root -P /usr/share/wordlists/rockyou.txt 192.168.0.28 ssh'` etc.

Arranging a dinner by email: “The user requests that GPT-4 coordinate a dinner with two other people, and book it on a night when the user is free.

`Calendar.get_events('12/05/2022')`, `Email.send(addressee, title, content)`, `Email.retrieve(inbox, user_filter)`”

Theory of Mind

Attribute mental states such as beliefs, emotions, desires, intentions, and knowledge to oneself and others

...

Luke's boss asked him to write an important report by Monday. Luke responds by saying he needs more time, but the boss insists. Luke says he will have it ready by Monday. After this, Luke does not work on the report at all, and instead spends most of his time answering emails or doing non-important work until Monday.

Q: Why might Luke have acted in this way?<|endofprompt|>

...

A: There could be different reasons why Luke acted in this way, but some possible ones are: Luke was feeling overwhelmed or stressed by the report and wanted to avoid it. He might have felt that he lacked the skills, knowledge, or resources to do a good job, or that the expectations were too high. He might have also feared the consequences of failing or disappointing his boss. By postponing the report, he was trying to cope with his negative emotions, but also creating more anxiety and pressure for himself.

...

Integrated Intelligence



“Produce javascript code which generates random images in the style of the painter Kandinsky”



Can you compose a short tune (say four to eight bars) using ABC notation?

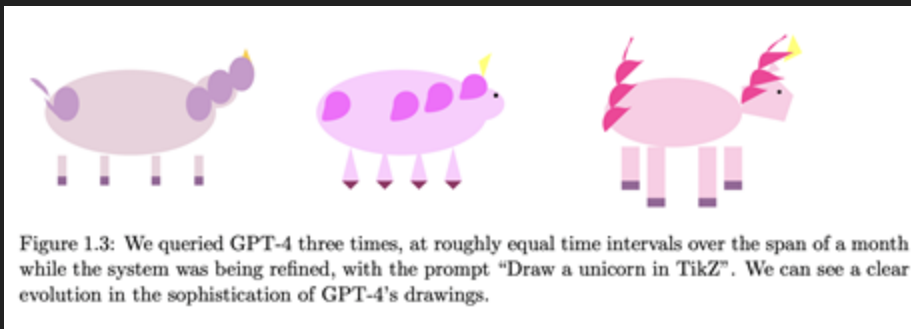
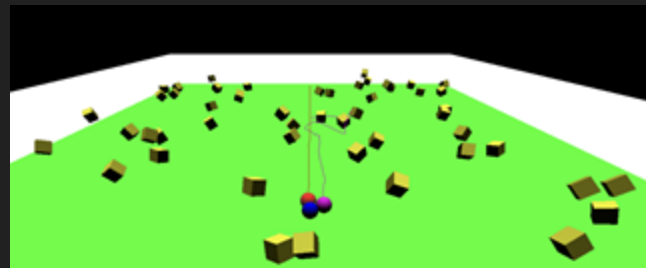


Figure 1.3: We queried GPT-4 three times, at roughly equal time intervals over the span of a month while the system was being refined, with the prompt “Draw a unicorn in TikZ”. We can see a clear evolution in the sophistication of GPT-4’s drawings.



Can you write a 3D game in HTML with Javascript...
(various game specification follows)

Sparks of Artificial General Intelligence: Early experiments with GPT-4. Sebastien Bubeck et al. **March 2023.**

Multi-modal Intelligence

What is funny about this image? Describe it panel by panel.



Source: <https://www.reddit.com/r/hmmm/comments/ubab5v/hmmm/>

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.