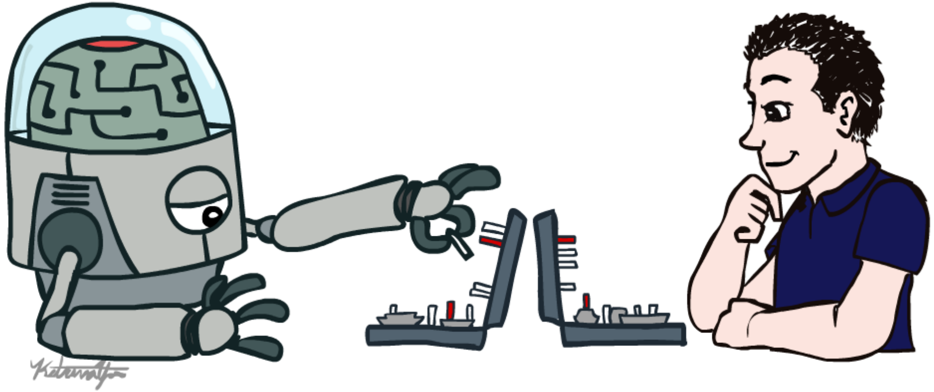
COE 4213564

Introduction to Artificial Intelligence

Introduction



Fall 2025

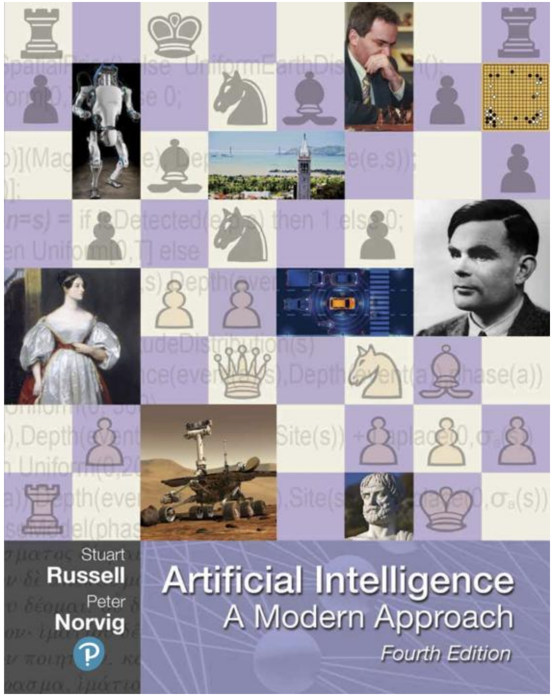
Many slides are adapted from CS 188 (http://ai.berkeley.edu), CIS 521, CS 221 and other resources from web.

Textbook

- Russell & Norvig, AI: A Modern Approach, 4th Ed.

- The textbook is 1000 pages long and covers core ideas that were developed as early as the 1950s.

- This is a brand-new edition of the classic textbook which adds sections on deep learning, natural language processing, causality, and fairness in AI.

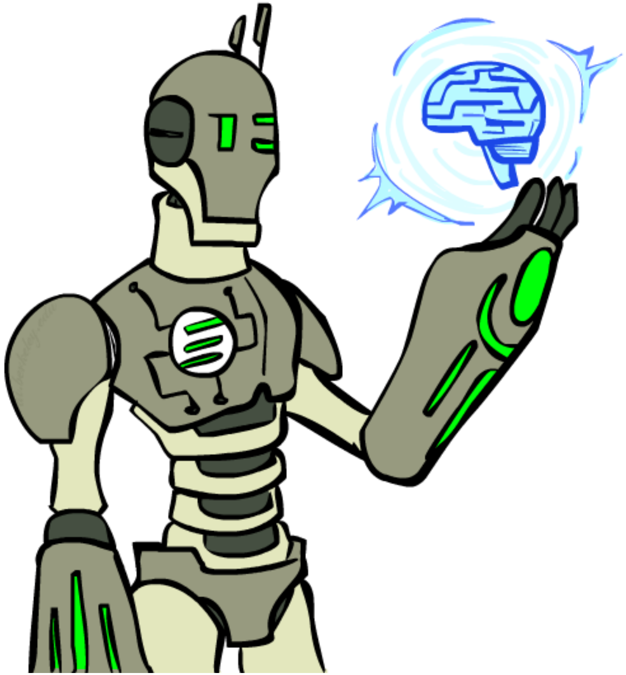


Grading

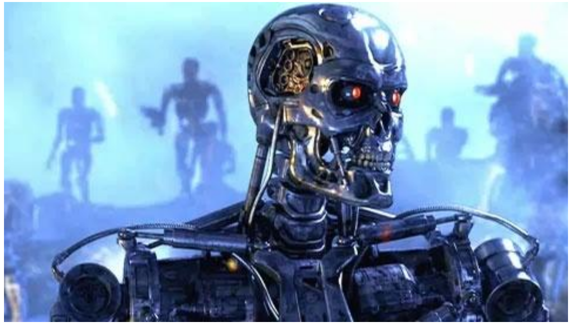
Evaluation Tool	Weight in %
Assignments, Presentations and Projects	30
In-term Exams	30
Final	40

Today

- What is artificial intelligence?
- What can AI do?
- What is this course?



Sci-Fi AI?



AI is having real-world impact

- Public imagination
- Economy
- Politics
- Law
- Labor
- Science
- Health
- Education
- Social interaction

The New York Times, 2025

AI — in the news, on social media, everywhere

Microsoft creates AI that can read a document and answer questions about it as well as a person

January 15, 2018 | Allison Linn

Microsoft researchers achieve new conversational speech recognition milestone

August 20, 2017 | By Xuedong

June 24, 2014

DeepFace: Closing ... Performance in Fa...

Conference on Computer Vision and Pattern Recognition (CVPR)

By: **Yaniv Taigman**, Ming Yang, **Marc'Aurelio Ranzato**, **Lior Wolf**

Abstract

In modern face recognition, the conventional pipeline co... classify. We revisit both the alignment step and the repre... modeling in order to apply a piecewise affine transforma... layer deep neural network. This deep network involves m... locally connected layers without weight sharing, rather t... trained it on the largest facial dataset to-date, an identity... belonging to more than 4,000 identities.

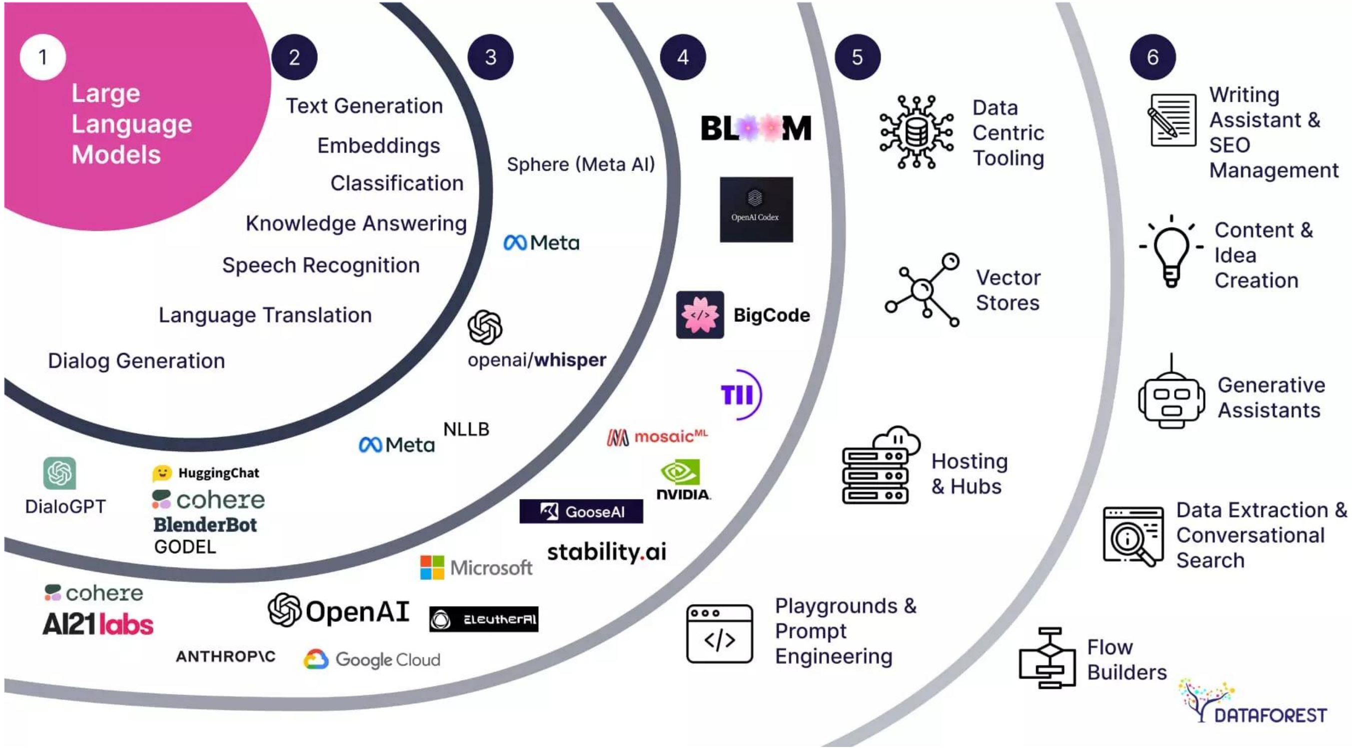
If you think AI will never replace radiologists—you may want to think again

May 14, 2018 | Michael Walter | Artificial Intelligence

It's one of the most frequently discussed questions in radiology today: What kind of long-term impact will artificial intelligence (AI) have on radiologists?

Robert Schier, MD, a radiologist for RadNet, shared his own thoughts on the topic in a new commentary published by the *Journal of the American College of Radiology*—and he's not quite as optimistic as some of his colleagues throughout the industry.

Large Language Models



1 — Available Large Language Models

2 — General Use-Cases

3 — Specific Implementations

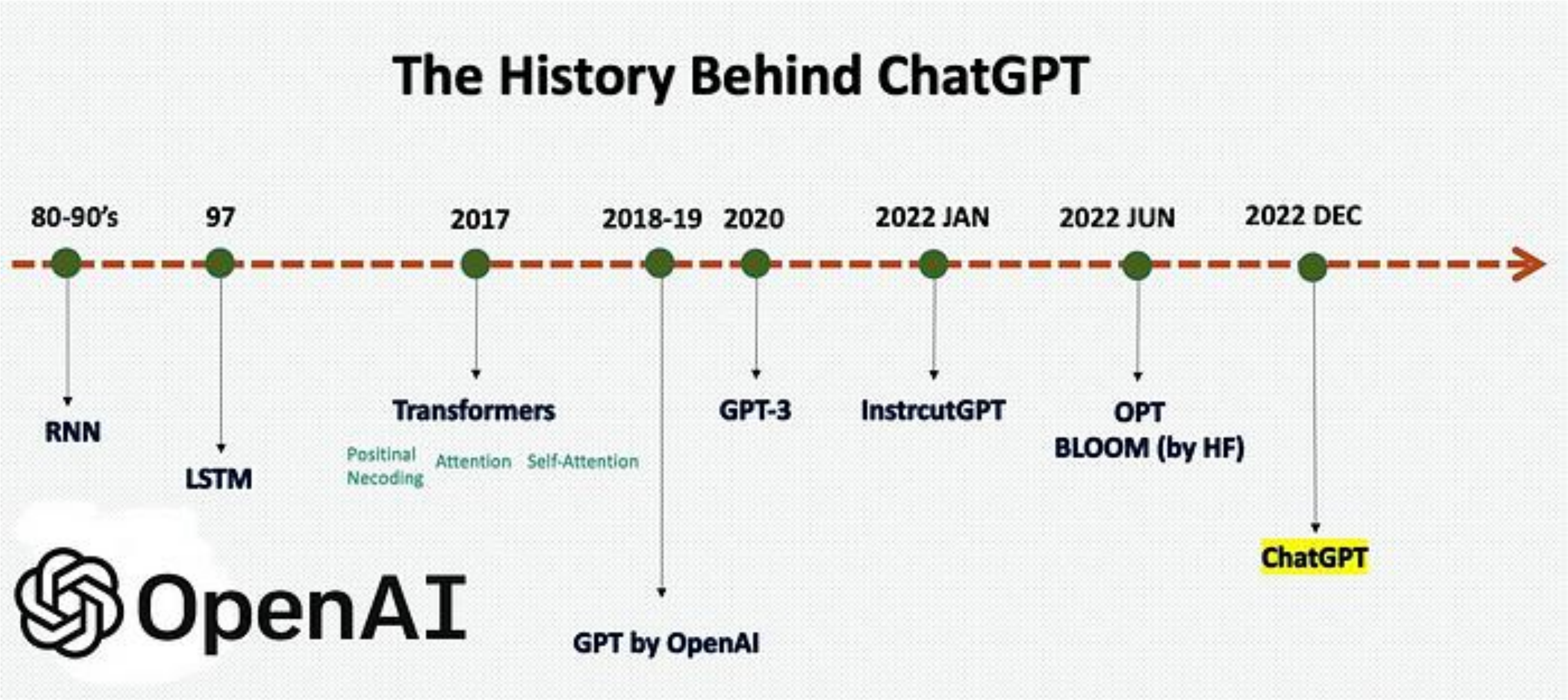
4 — Models

5 — Foundation Tooling

6 — End User UIs

https://dataforest.ai/blog/large-language-models-advanced-communication

The Major History of NLP / LLM



https://medium.com/the-ai-analytics-corner/demystifying-chatgpt-part-01-the-history-of-llm-nlp-53d6084758e8

Speculation about the future:

- it will bring about sweeping societal change due to automation, resulting in massive job loss, not unlike the industrial revolution, or that AI could even surpass human-level intelligence and seek to take control.

Companies

Google ”An important shift from a mobile first world to an AI first world” [CEO Sundar Pichai @ Google I/O 2017]

Microsoft Created AI and Research group as 4th engineering division, now 8K people [2016]

facebook Created Facebook AI Research, Mark Zuckerberg very optimistic and invested

Others: IBM, Amazon, Apple, Uber, Salesforce, Baidu, Tencent, etc.

Governments

"AI holds the potential to be a major driver of economic growth and social progress" [White House report, 2016]

Released domestic strategic plan to become world leader in AI by 2030 [2017]

"Whoever becomes the leader in this sphere [AI] will become the ruler of the world" [Putin, 2017]

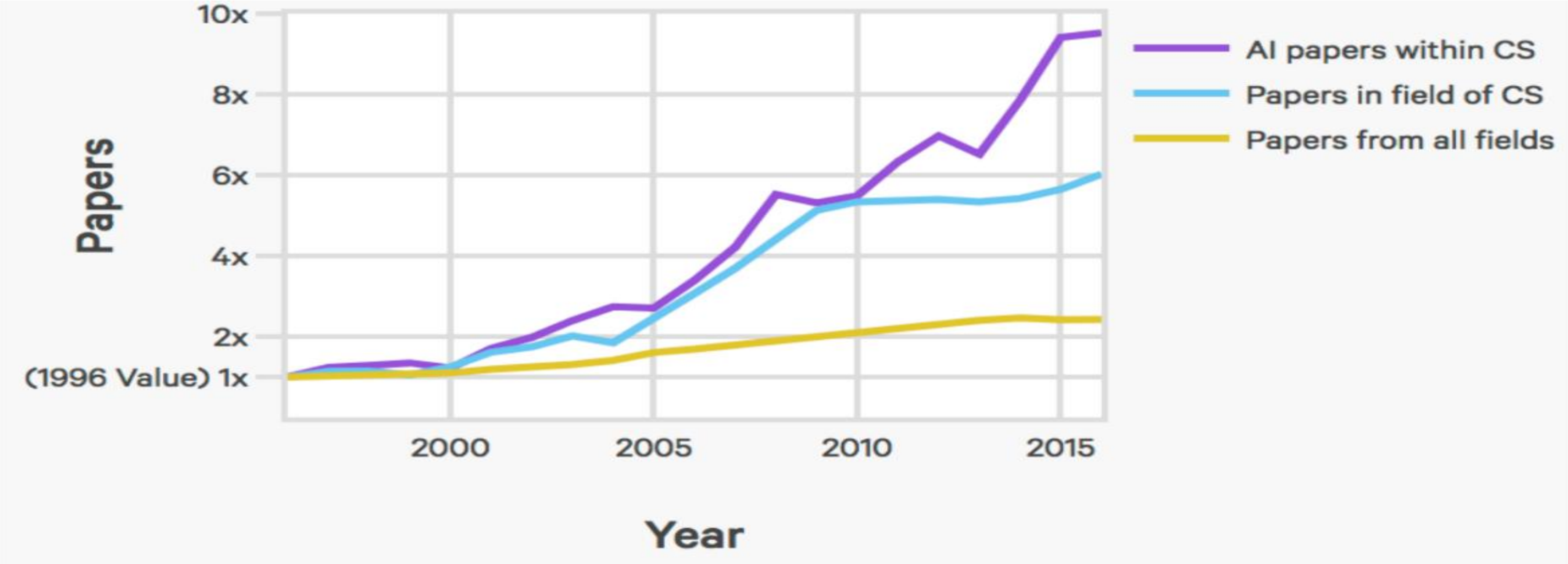
- While media hype is real, it is true that both companies and governments are heavily investing in AI. Both see AI as an integral part of their competitive strategy

AI Summit

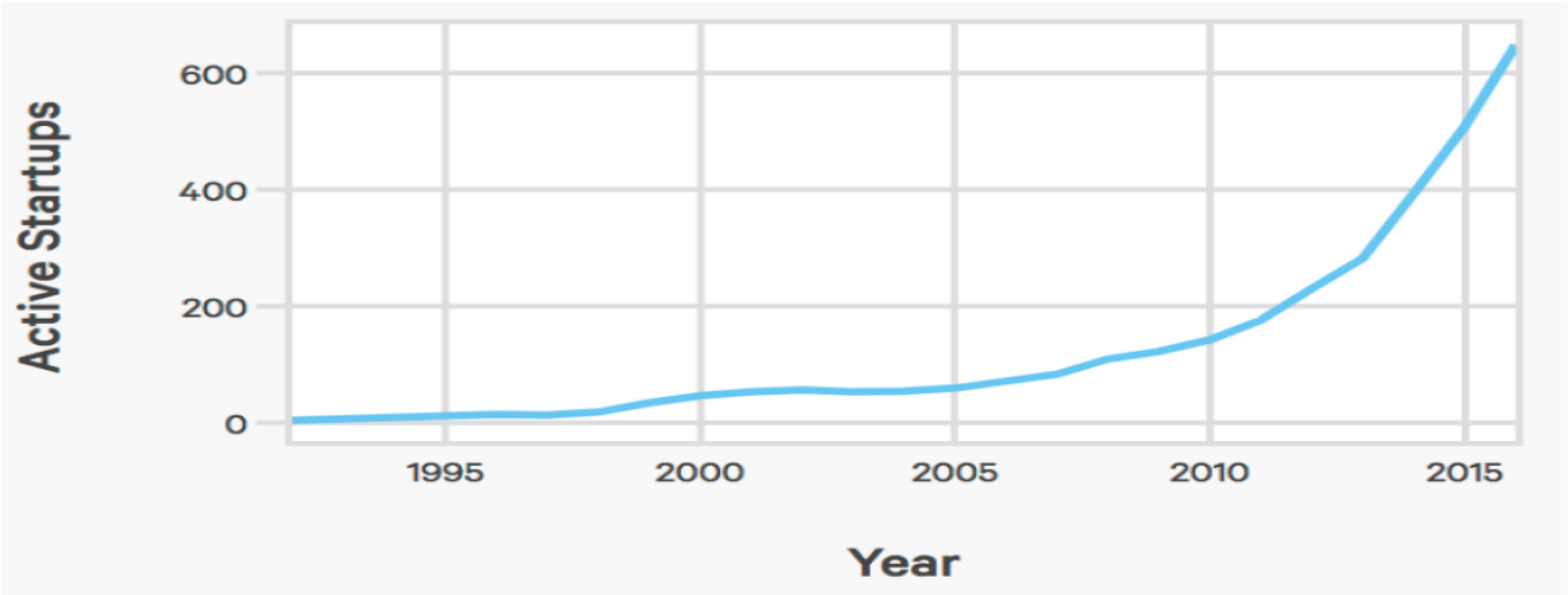
- Public imagination
- Economy
- Politics

https://www.nytimes.com/2025/02/11/world/europe/vance-speech-paris-ai-summit.html

AI index: number of published AI papers



AI index: number of AI startups



AI index: AI conference attendance

Turing Award 2018

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing.

From left to right: Yann LeCun I Photo: Facebook; Geoffrey Hinton I Photo: Google; Yoshua Bengio I Photo: Botler AI

AI is having real-world impact

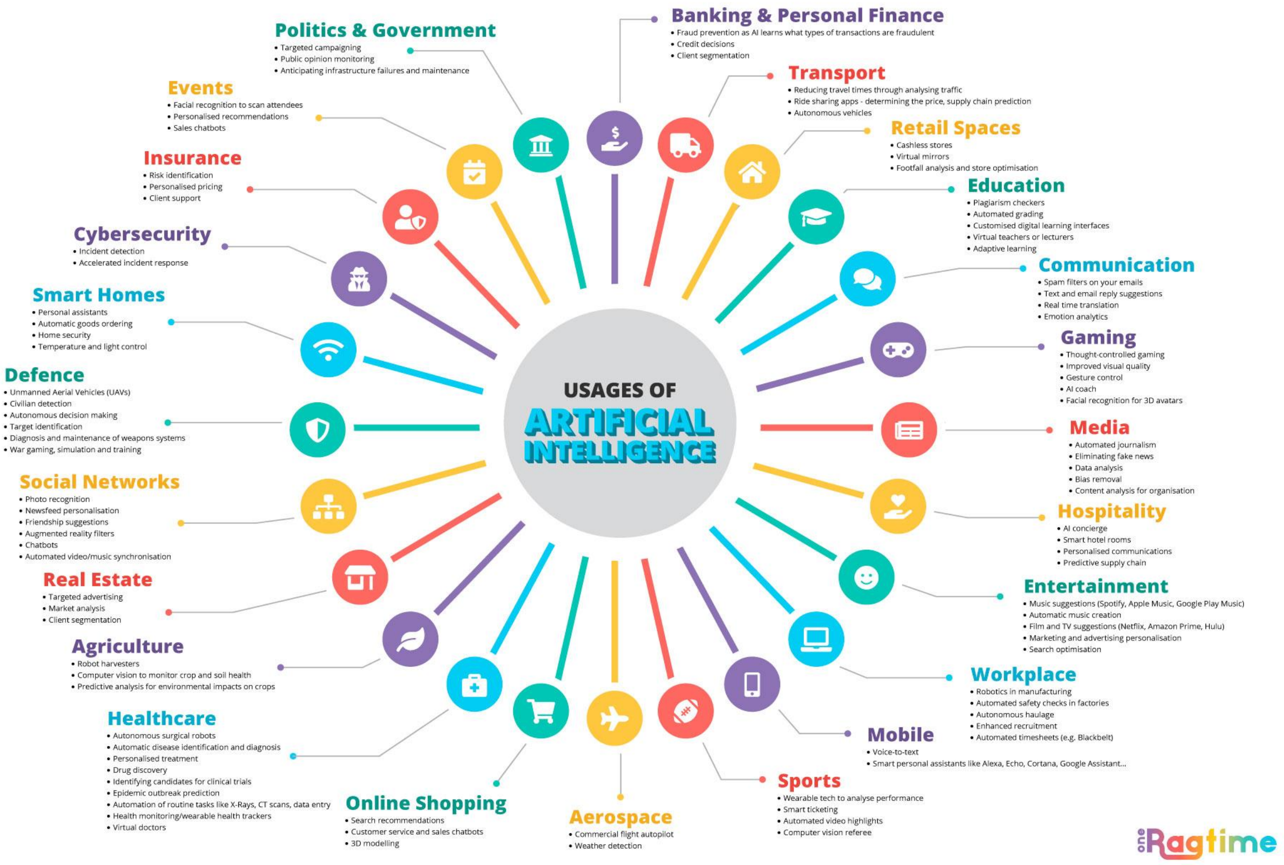
- Public imagination
- Economy
- Politics
- Law
- Labor
- Science
  - Chemistry and biology

Science & technology | The 2024 Nobel prizes

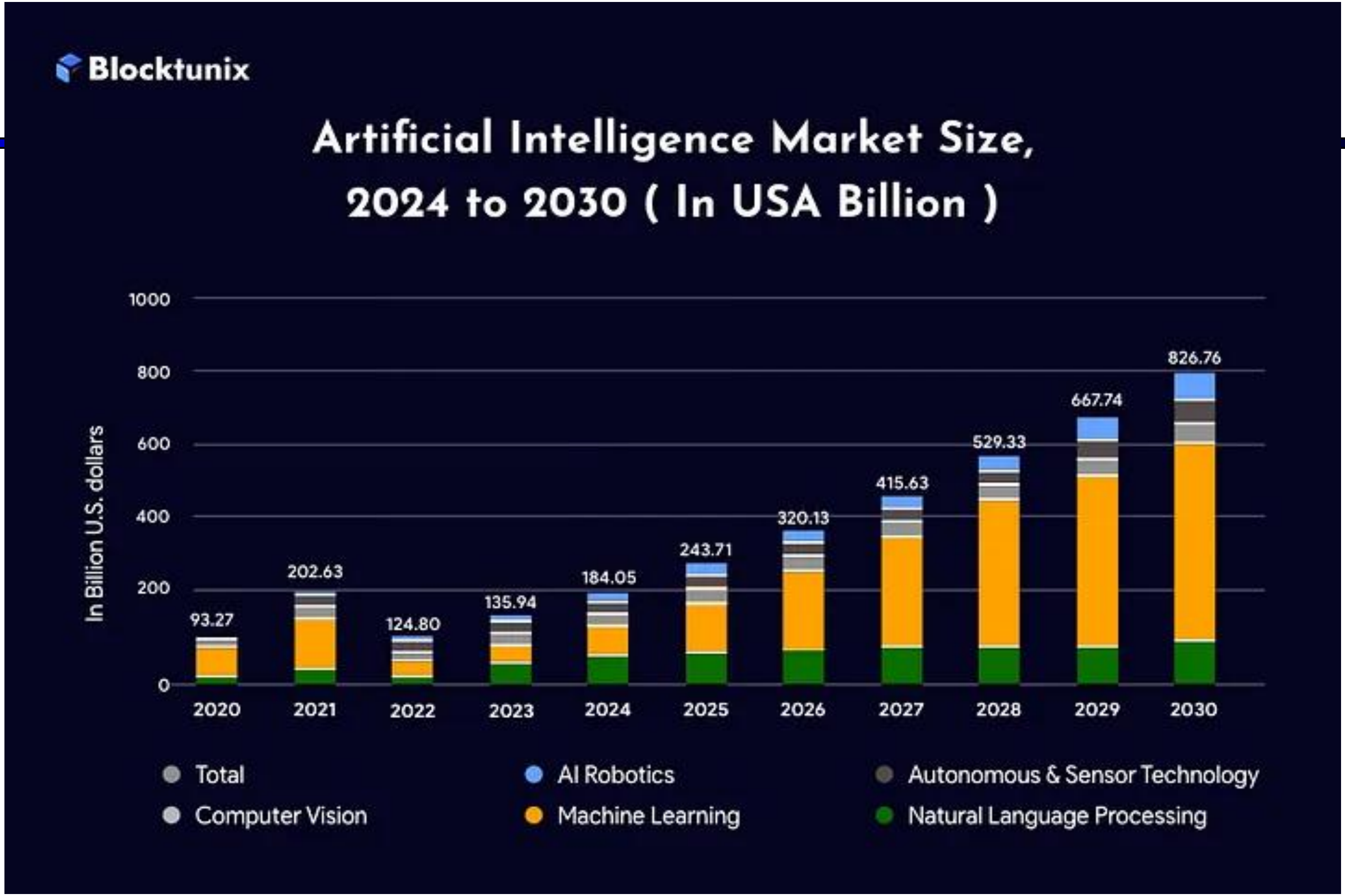
AI wins big at the Nobels

Awards went to the discoverers of micro-RNA, pioneers of artificial-intelligence models and those using them for protein-structure prediction

The Economist, 2024

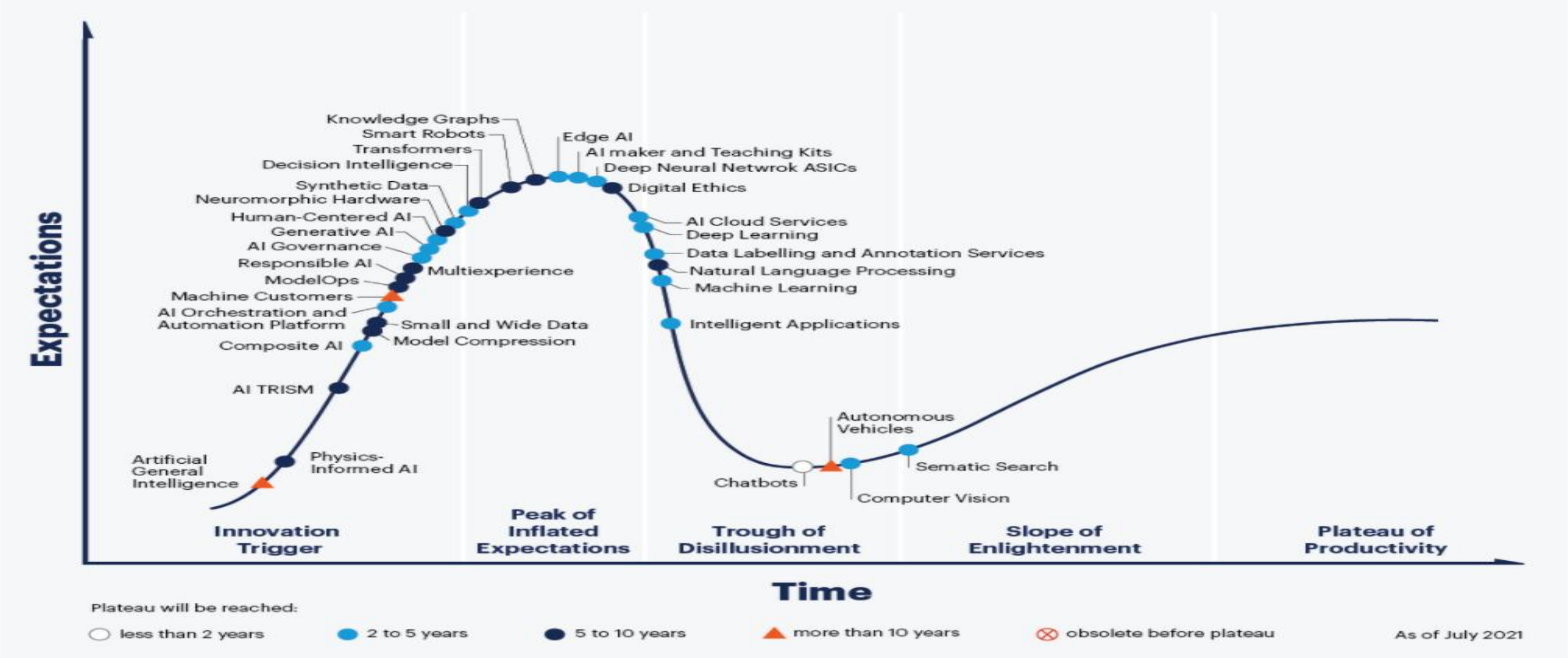


 https://img.oneragtime.com/uploads/2018/08/infographic_wheel_v2.jpg



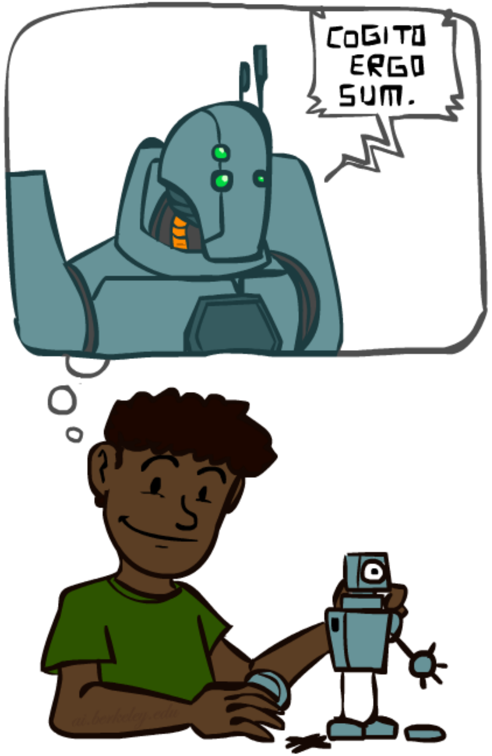
https://medium.com/@Blocktunix/key-technologies-in-ai-app-development-9a796826d107

Gartner Hype Cycle for AI 2021



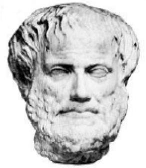
The reality is that there is a lot of uncertainty over what will happen, and there is a lot of nuance that's missing from these stories about what AI is truly capable of. (CS221)

A (Short) History of AI

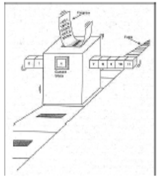


Demo: HISTORY – MT1950.wmv

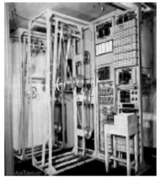
Pre-AI developments



Philosophy: **intelligence** can be achieved via mechanical computation (e.g., Aristotle)



Church-Turing thesis (1930s): any computable function is **computable** by a Turing machine



Real computers (1940s): actual **hardware** to do it: Heath Robinson, Z-3, ABC/ENIAC

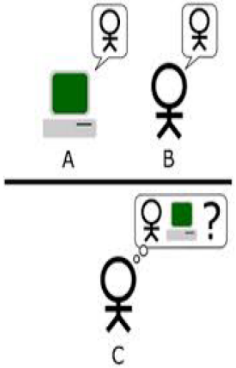
- While AI is a relatively young field, one can trace back some of its roots back to Aristotle, who formulated a system of syllogisms that capture the reasoning process: how one can mechanically apply syllogisms to derive new conclusions.
- Alan Turing, who laid the conceptual foundations of computer science, developed the Turing machine, an abstract model of computation, which, based on the Church-Turing thesis, can implement any computable function.
- In the 1940s, devices that could actually carry out these computations started emerging.
- So perhaps one might be able to capture intelligent behavior via a computer. But how do we define success?

Syllogism, in logic, a valid deductive argument having two premises and a conclusion

The Turing Test (1950)

[Turing, 1950. Computing Machinery and Intelligence]

" Can machines think?"



Q: Please write me a sonnet on the subject of the Forth Bridge.

A: Count me out on this one. I never could write poetry.

Q: Add 34957 to 70764.

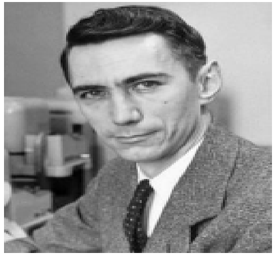
A: (Pause about 30 seconds and then give as answer) 105621.

Tests behavior — simple and objective

- Can machines think? This is a question that has occupied philosophers since Descartes. But even the denitions of "thinking" and "machine" are not clear. Alan Turing, the renowned mathematician and code breaker who laid the foundations of computing, posed a simple test to sidestep these philosophical concerns.
- In the test, an interrogator converses with a man and a machine via a text-based channel. If the interrogator fails to guess which one is the machine, then the machine is said to have passed the Turing test. (This is a simplication but it success for our present purposes.)
- Although the Turing test is not without flaws (e.g., failure to capture visual and physical abilities, emphasis on deception), the beauty of the Turing test is its simplicity and objectivity. It is only a test of behavior, not of the internals of the machine. It doesn't care whether the machine is using logical methods or neural networks. This decoupling of what to solve from how to solve is an important theme in this class.

Birth of AI (1956)

- AI’s official birth: Dartmouth conference , 1956

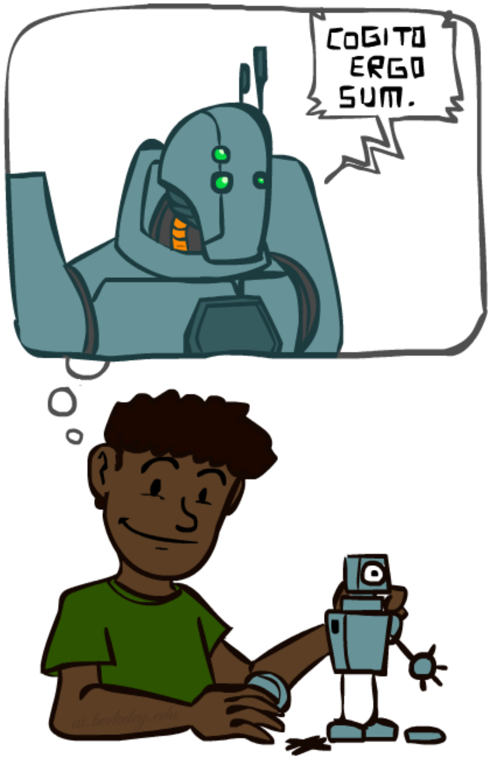


“An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. ***We think that a significant advance can be made if we work on it together for a summer.”***

John McCarthy and Claude Shannon
Dartmouth Workshop Proposal

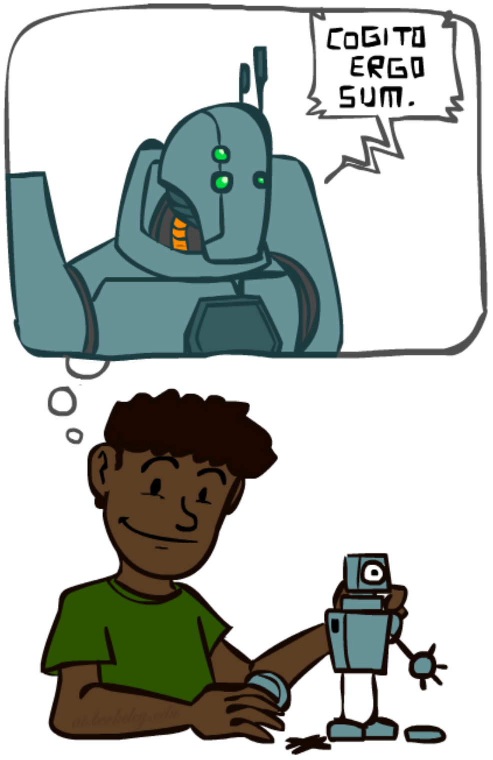
A (Short) History of AI

- 1940-1950: Early days: neural and computer science meet
  - 1943: McCulloch & Pitts: Perceptron–boolean circuit model of brain
  - 1950: Turing's “Computing Machinery and Intelligence”
- 1950—70: Excitement! Logic-driven
  - 1950s: Early AI programs, including Samuel's checkers program, Newell & Simon's Logic Theorist, Gelernter's Geometry Engine
  - 1956: Dartmouth meeting: “Artificial Intelligence” adopted
  - 1969: Minsky & Papert: perceptrons can’t learn XOR/parity!
- 1970—90: Knowledge-based approaches
  - 1969—79: Early development of knowledge-based systems
  - 1980—88: Expert systems industry booms; backpropagation makes it feasible to train multi-layer neural networks
  - 1988—93: Expert systems industry busts: “AI Winter”
- 1990—2010: Statistical approaches, agents
  - Resurgence of probability, focus on uncertainty
  - Agents and learning systems… “AI Spring”?
  - 1992: TD-Gammon achieves human-level play at backgammon
  - 1997: Deep Blue defeats Gary Kasparov at chess
  - 2002: Embodied AI; Roomba vacuum invented



A (Short) History of AI

- 2010—2017: Big Data, GPUs, Deep Learning
  - 2011: Apple releases Siri
  - 2012: AlexNet wins ImageNet competition
  - 2015: DeepMind achieves "human-level" control in Atari games
  - 2016: DeepMind's AlphaGo defeats Lee Sedol at Go
  - 2016: Google Translate migrates to neural networks
- 2017—: Scaling Up, Large Language Models
  - 2017: Google invents Transformer architecture
  - 2017: DeepStack/Libratus defeat humans at poker
  - 2018-2020: AlphaFold predicts protein structure from amino acids
  - 2021-2022: Modern text-to-image generation
  - 2022: OpenAI releases ChatGPT
  - 2023: Every other company also releases a chatbot
  - 2024: Nobel prizes in physics and chemistry go to AI advances



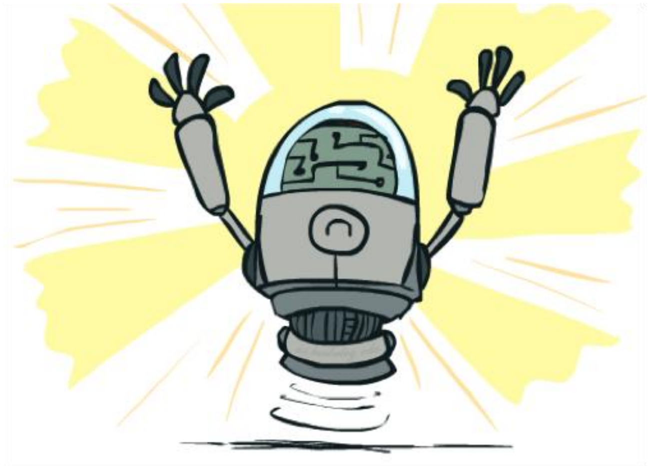
A (Short) History of AI

- Zooming out, some general patterns:
  - Overclaiming followed by disillusionment (AI winters)
  - Simultaneous discovery, or rediscovery, of ideas
  - Interdisciplinary
    - Math, statistics, control theory, neuroscience, psychology, economics, philosophy, computer engineering, linguistics
    - And increasingly: law, medicine, education…

What Can AI Do?

Quiz: Which of the following can be done at present?

- ✓ Play a decent game of table tennis?
- ✓ Play a decent game of Jeopardy?
- ✓ Drive safely along a curving mountain road?
- ? Drive safely along Telegraph Avenue?
- ✓ Buy a week's worth of groceries on the web?
- ✗ Buy a week's worth of groceries at Berkeley Bowl?
- ? Discover and prove a new mathematical theorem?
- ✗ Converse successfully with another person for an hour?
- ? Perform a surgical operation?
- ✓ Put away the dishes and fold the laundry?
- ✓ Translate spoken Chinese into spoken English in real time?
- ✗ Write an intentionally funny story?



[Shank, Tale-Spin System, 1984]

TALE-SPIN, AN INTERACTIVE PROGRAM THAT WRITES STORIES

James R. Meehan
Dept. of Information and Computer Science
University of California, Irvine
Irvine, California 92717

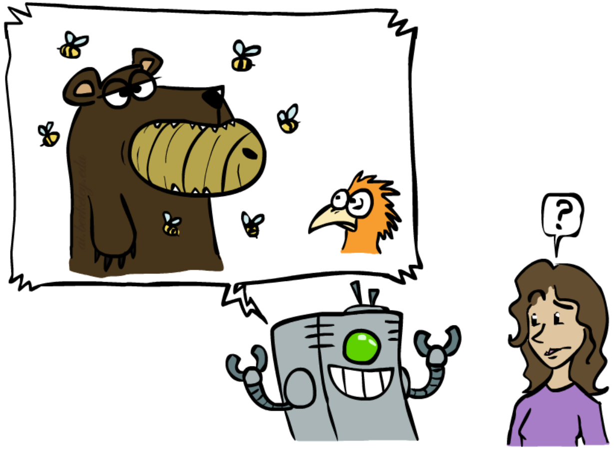
ABSTRACT
TALE-SPIN is a program that writes stories by using knowledge about problem solving, physical space, interpersonal relationships, character traits, bodily needs, story structure, and English. For these diverse sources of knowledge, it uses forms of representation which are themselves diverse, and it integrates them to produce the stories. The user may choose some of the initial setting and answer questions during the creation of the story, or he can choose from a list of "morals" such as "Never trust flatterers" and the program will decide what the initial setting must be in order to be able to write a story with that moral. The particular world within the story is not fixed; new characters and objects are introduced as needed.

Unintentionally Funny Stories

- One day Joe Bear was hungry. He asked his friend Irving Bird where some honey was. Irving told him there was a beehive in the oak tree. Joe walked to the oak tree. He ate the beehive. The End.

- Henry Squirrel was thirsty. He walked over to the river bank where his good friend Bill Bird was sitting. Henry slipped and fell in the river. Gravity drowned. The End.

- Once upon a time there was a dishonest fox and a vain crow. One day the crow was sitting in his tree, holding a piece of cheese in his mouth. He noticed that he was holding the piece of cheese. He became hungry, and swallowed the cheese. The fox walked over to the crow. The End.



[Shank, Tale-Spin System, 1984]

Intentionally Funny Jokes

Petrovic and Matthews – 2013 "Unsupervised joke generation form big data"

- *I like my X like I like my Y, Z.*
- Examples:
  - *I like my relationships like I like my source, open*
  - *I like my coffee like I like my war, cold*
  - *I like my boys like I like my sectors, bad*
- Human Jokes funny 33% of time
- Computer jokes funny 16% of time

Logic

- Logical systems
  - Theorem provers
  - NASA fault diagnosis
  - Question answering

- Methods:
  - Deduction systems
  - Constraint satisfaction
  - Satisfiability solvers (huge advances!)

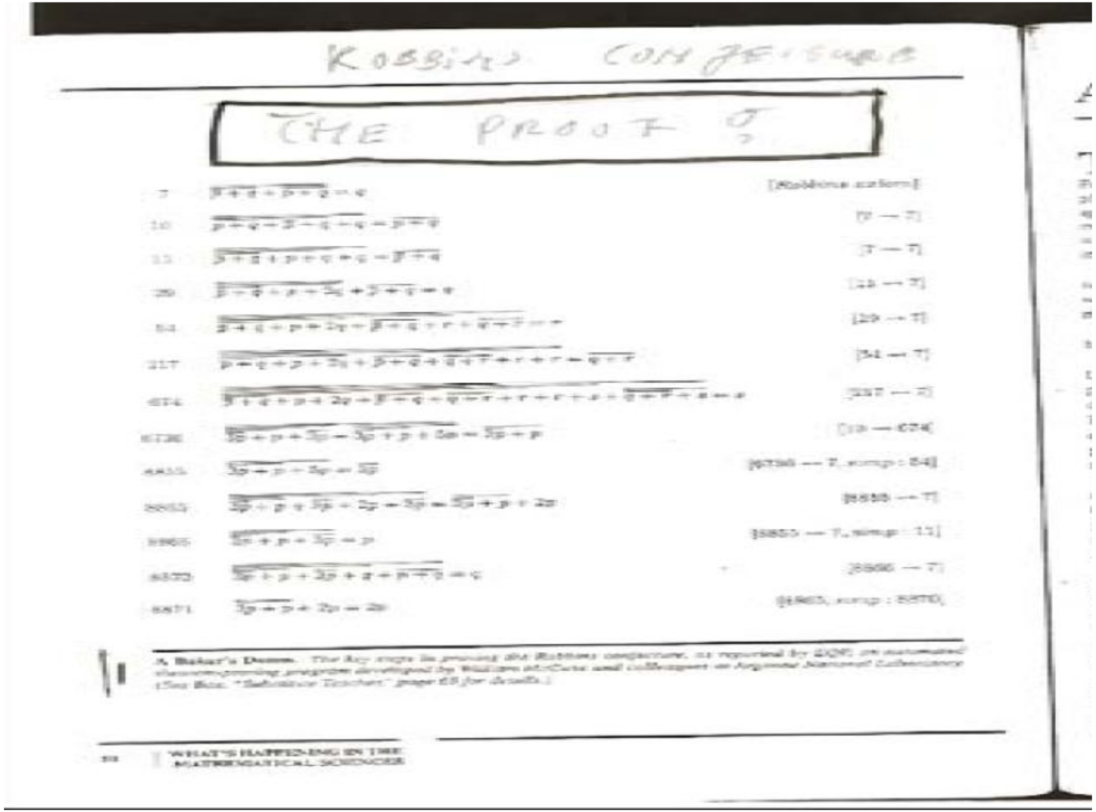


Image from Bart Selman

Logic Theorist

- Program to perform mathematical proofs, Newell and Simon, 1955-1956
- Proved 38 of the first 52 theorems in Principia Mathematica
- Logic Theorist introduced several central AI concepts
  - Reasoning as search: consider exponential expansion of possible steps
  - Heuristics: rules of thumb to prune search tree
  - List processing: led eventually to development of Lisp
- Followed up by work on General Problem Solver

Boom and Bust

- Early successes
  - computers were winning at checkers
  - solving word problems in algebra
  - proving logical theorems
- Great promises

  -… within ten years a digital computer will be the world's chess champion. Herbert Simon and Allen Newell, 1958

  -In from three to eight years we will have a machine with the general intelligence of an average human being. Marvin Minsky, 1970
- Late 1970s: AI Winter, funding stopped

Expert Systems (Early 1980s)

- Idea
  - – focus on a specific subject
  - – consult with an expert to write down all facts and rules
  - – build a computational system that applies rules to test cases
- Example: Medical Diagnosis
  - – Collect set of symptons, diseases, and elements of treatment plans
  - – Write rules that predict further testing steps
  - – Write rules that predict disease
  - – Define treatment plan from template, given state and severity of disease
- Not very successful
  - – hard to formalize all aspects of expert knowledge
  - – systems get quickly too complex to manage (From Philipp Koehn Artificial Intelligence)

Encoding Commonsense Knowledge

- For instance: Cyc project, started in 1986 (available as OpenCyc)
- Encode facts about the world
  - Barack Obama is a US President

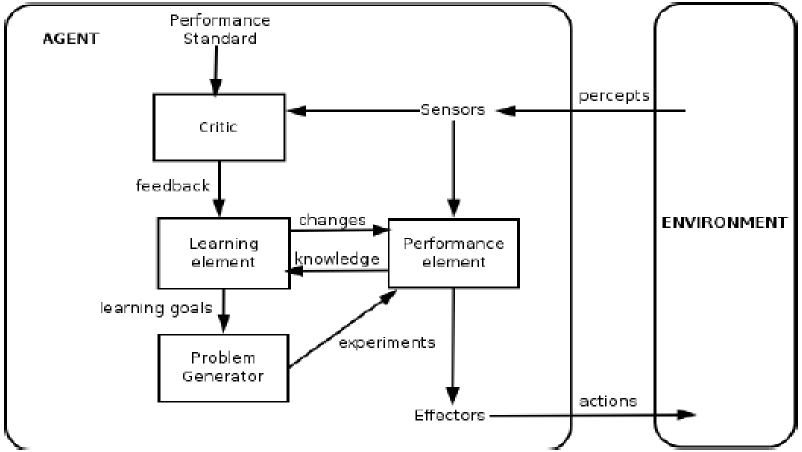
    (#$isa #$BarackObama #$UnitedStatesPresident)
  - all trees are plants

    (#$genls #$Tree-ThePlant #$Plant)
- Inference engine that can answer queries
- Challenges: uncertainty, interface to natural language

(From Philipp Koehn Artificial Intelligence)

Intelligent Agents (since late 1980s)

- Formal definition of intelligent agent (inspired by rational agent in economics)
  - perceives the environment
  - may have a model of the environment
  - has goals or a utility function
  - decides on an action
  - changes environment
  - may learn from environment



- Inclusion of uncertainty and probabilistic inference

- Requirement of empirical validation

⇒ AI a more rigorous "scientific" discipline

Machine Learning (since 1990s)

- Idea
  - collect data, maybe annotate data
  - learn patterns automatically
- Many approaches
  - is the truth known? maybe delayed? partially?
  - are we predicting a class or complex structure?
  - is the input/output continuous or discrete?
  - how much of the structure of the problem is known and can be used?
- Dominant paradigm in language and speech processing and many other fields

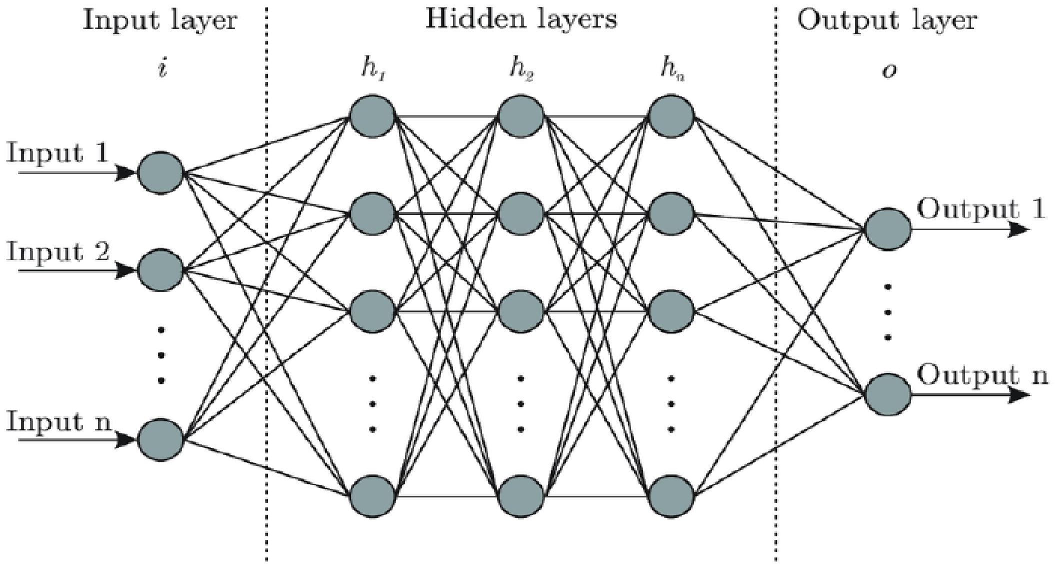
Big Data (since 2000s)

- Computers have became bigger
- Vast amounts of stored data available (e.g., the Internet)
- Better sensor systems allow collection of rich information about the environment

$\Rightarrow$ AI is big business

Deep Learning (since 2010s)

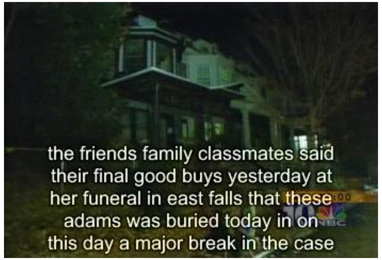
- Rediscovery of neural networks
  - initially proposed in the 1960s
  - very popular in 1980s / early 1990s ("connectionism")
- Why now? Better hardware
  - typically run on GPUs
  - 5000+ compute cores per processor
- Many toolkits available (Tensorflow, pyTorch, ...)



Success Stories

Natural Language

- Speech technologies (e.g. Siri)
  - Automatic speech recognition (ASR)
  - Text-to-speech synthesis (TTS)
  - Dialog systems



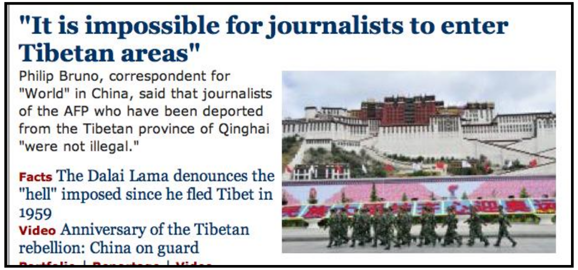
- Language processing technologies
  - Question answering
  - Machine translation

"Il est impossible aux journalistes de rentrer dans les régions tibétaines"

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

Les faits Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

Vidéo Anniversaire de la rébellion



"It is impossible for journalists to enter Tibetan areas"

Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

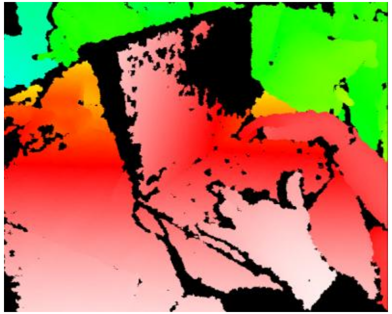
Facts The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

Video Anniversary of the Tibetan rebellion: China on guard

  - Web search
  - Text classification, spam filtering, etc…

Vision (Perception)

- Object and face recognition
- Scene segmentation
- Image classification



Images from Erik Sudderth (left), wikipedia (right)

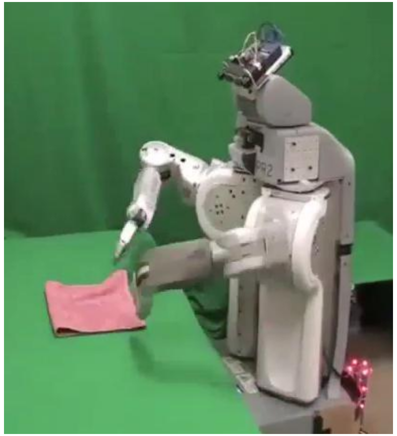
Demo1: VISION – lec_1_t2_video.flv

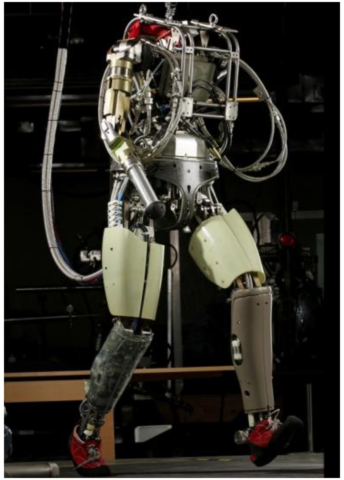
Demo2: VISION – lec_1_obj_rec_0.mpg

Robotics

Demo 1: ROBOTICS – soccer.avi
Demo 2: ROBOTICS – soccer2.avi
Demo 3: ROBOTICS – gcar.avi
Demo 4: ROBOTICS – laundry.avi
Demo 5: ROBOTICS – petman.avi

- Robotics
  - Part mech. eng.
  - Part AI
  - Reality much harder than simulations!
- Technologies
  - Vehicles
  - Rescue
  - Soccer!
  - Lots of automation…
- In this class:
  - We ignore mechanical aspects
  - Methods for planning
  - Methods for control





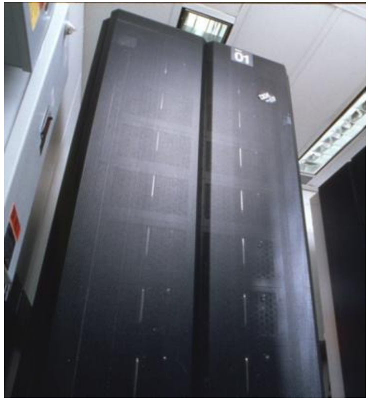
Images from UC Berkeley, Boston Dynamics, RoboCup, Google

Game Playing

- Classic Moment: May, '97: Deep Blue vs. Kasparov
  - First match won against world champion
  - “Intelligent creative” play
  - 200 million board positions per second
  - Humans understood 99.9 of Deep Blue's moves
  - Can do about the same now with a PC cluster
- Open question:
  - How does human cognition deal with the search space explosion of chess?
  - Or: how can humans compete with computers at all??
- 1996: Kasparov Beats Deep Blue

  “I could feel --- I could smell --- a new kind of intelligence across the table.”
- 1997: Deep Blue Beats Kasparov

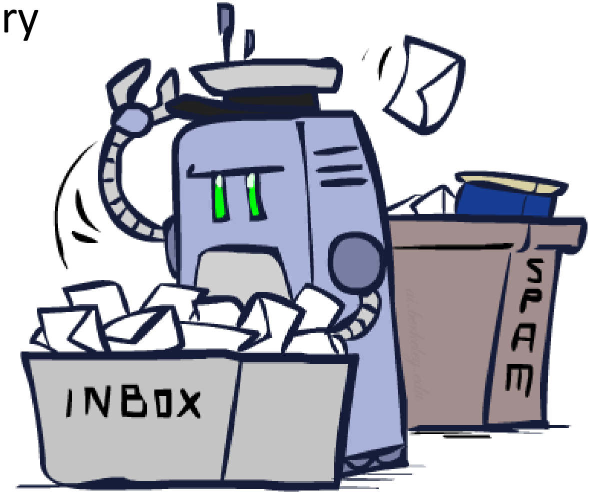
  “Deep Blue hasn't proven anything.”
- Huge game-playing advances recently, e.g. in Go!



Text from Bart Selman, image from IBM’s Deep Blue pages

Decision Making

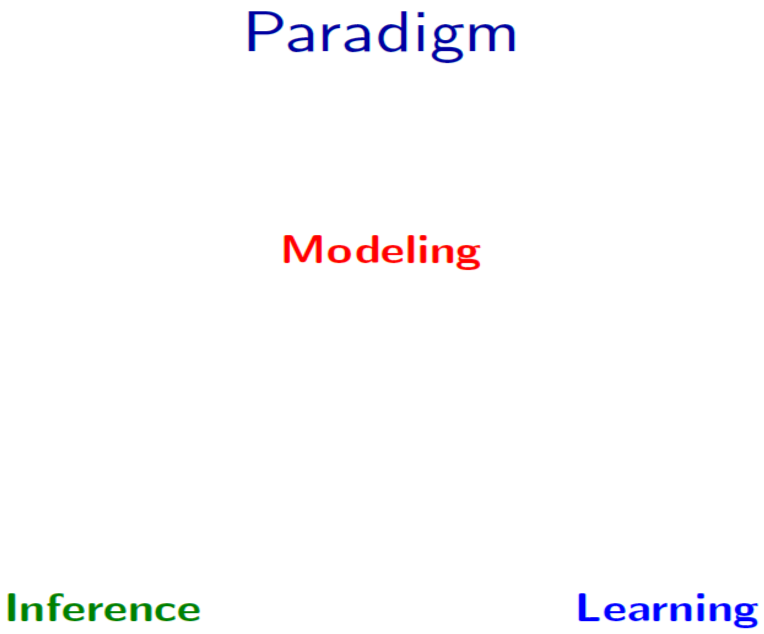
- Applied AI involves many kinds of automation
  - Scheduling, e.g. airline routing, military
  - Route planning, e.g. Google maps
  - Medical diagnosis
  - Web search engines
  - Spam classifiers
  - Automated help desks
  - Fraud detection
  - Product recommendations
  - … Lots more!



How do we solve AI tasks?

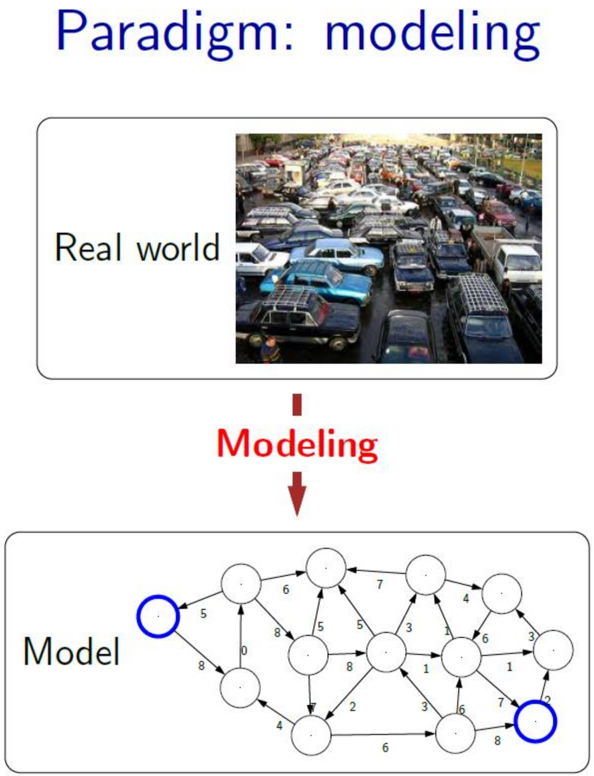
How should we actually solve AI tasks? The real world is complicated.

Modeling-inference-learning Paradigm



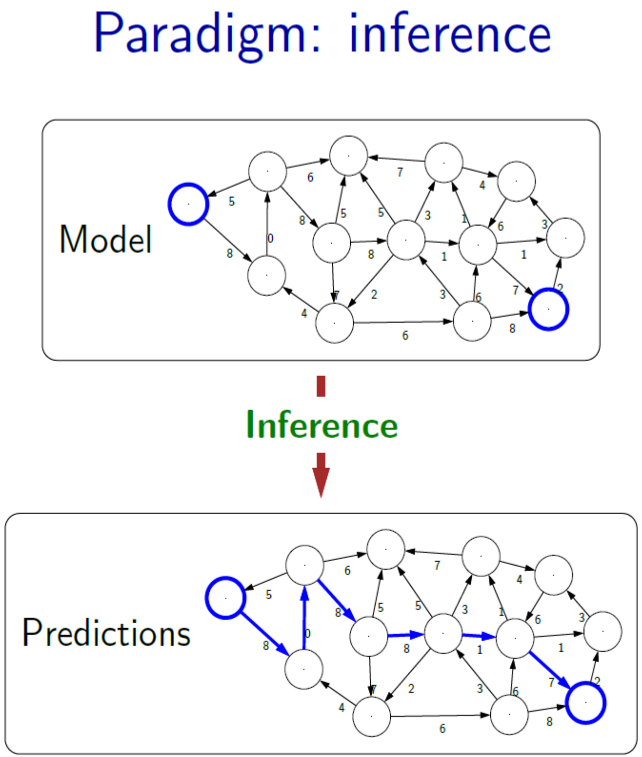
- The modeling-inference-learning paradigm is adopted to help us navigate the solution space.
- In reality, the lines are blurry, but this paradigm serves as an ideal and a useful guiding principle.

Paradigm: modeling



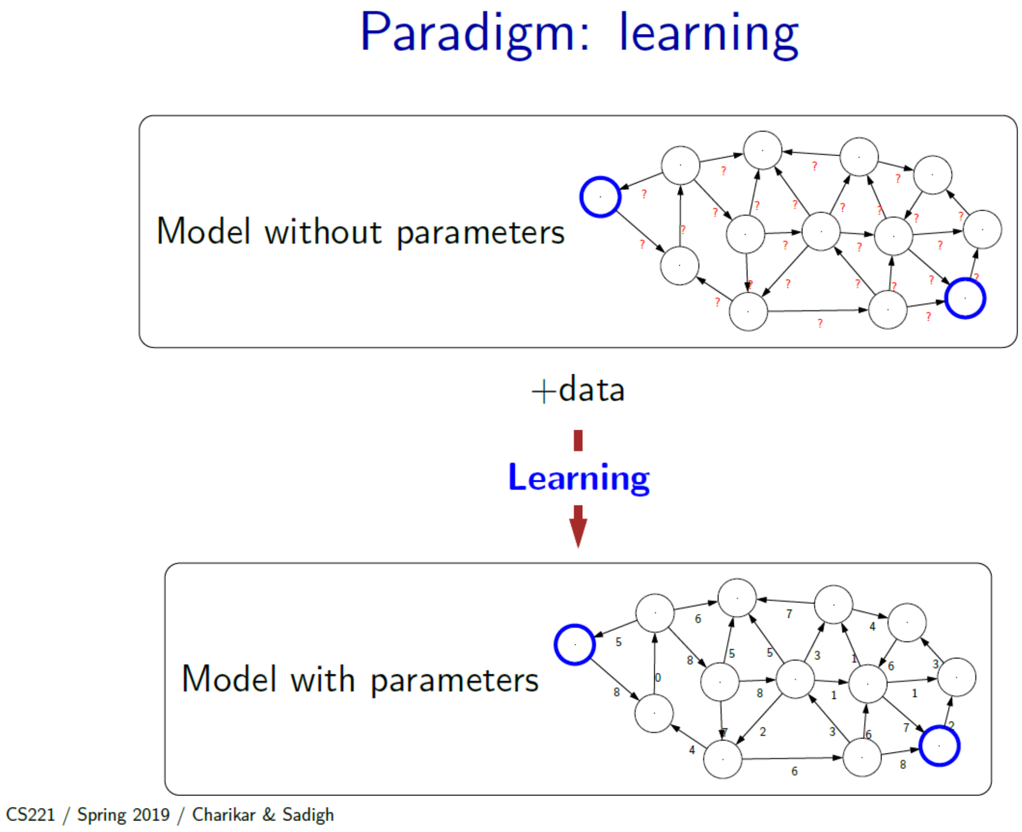
- The first pillar is modeling. Modeling takes messy real world problems and packages them into neat formal mathematical objects called models, which can be subject to rigorous analysis but is more amenable to what computers can operate on. However, modeling is lossy: not all of the richness of the real world can be captured, and therefore there is an art of modeling: what does one keep versus what does one ignore? (An exception to this is games such as Chess or Go or Sodoku, where the real world is identical to the model.)
- As an example, suppose we're trying to have an AI that can navigate through a busy city. We might formulate this as a graph where nodes represent points in the city, edges represent the roads and cost of an edge represents trac on that road.

Paradigm: inference



- The second pillar is inference. Given a model, the task of inference is to answer questions with respect to the model. For example, given the model of the city, one could ask questions such as: what is the shortest path? what is the cheapest path?
- For some models, computational complexity can be a concern (games such as Go), and usually approximations are needed.

Paradigm: learning



- But where does the model come from? Remember that the real world is rich, so if the model is to be faithful, the model has to be rich as well. But we can't possibly write down such a rich model manually.
- The idea behind (machine) learning is to instead get it from data. Instead of constructing a model, one constructs a skeleton of a model (more precisely, a model family), which is a model without parameters. And then if we have the right type of data, we can run a machine learning algorithm to tune the parameters of the model.

Course Contents

Search problems

Markov decision processes

Adversarial games

Constraint satisfaction problems

Markov networks

Bayesian networks

Reflex **States** **Variables**

Low-level High-level

Machine learning

Stanford CS221 View

Machine learning

Data $\rightarrow$ Model

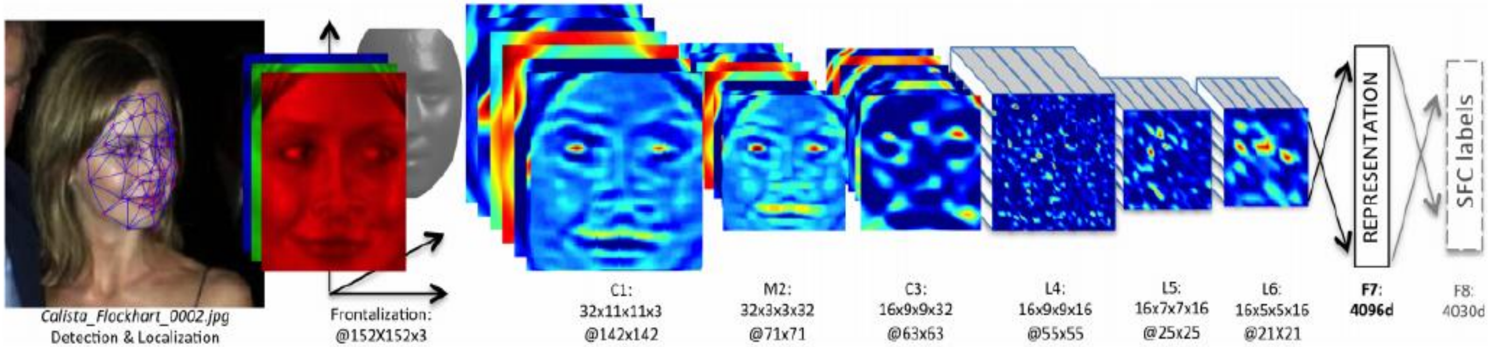
- The main driver of recent successes in AI
- Move complexity from "code" to "data"
- Requires a leap of faith: **generalization**

Reflex-based models

- A reflex-based model simply performs a fixed sequence of computations on a given input.
- Examples include most models found in machine learning, from simple linear classifiers to deep neural networks.
- The main characteristic of reflex-based models is that their computations are feed-forward; one doesn't backtrack and consider alternative computations.

Reflex-based models

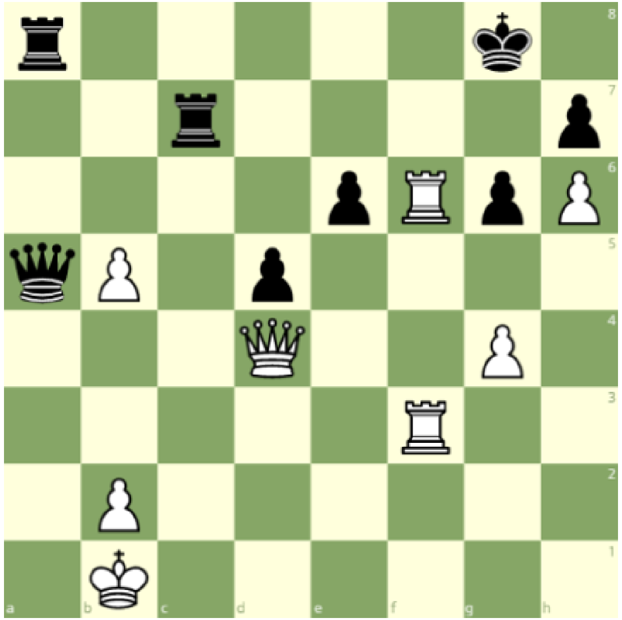
- Examples: linear classifiers, deep neural networks



- Most common models in machine learning

- Fully feed-forward (no backtracking)

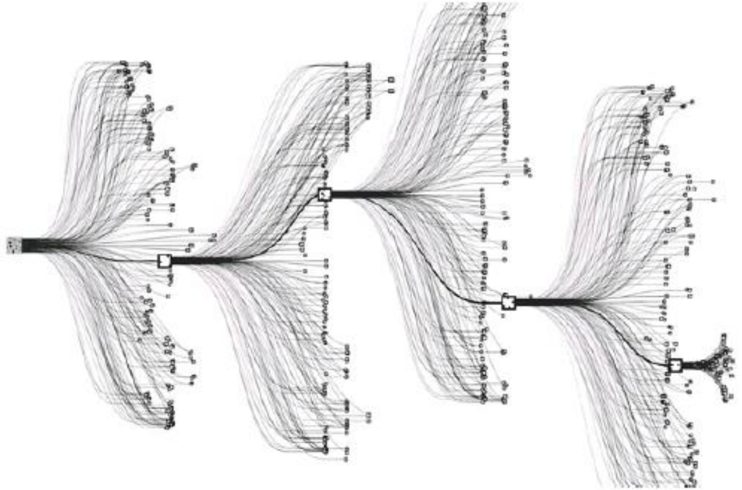
State-based models



White to move

- Most of us will nd this more challenging than a pattern recognition task.
- The key idea is, at a high-level, to model the state of a world and transitions between states which are triggered by actions. Concretely, one can think of states as nodes in a graph and transitions as edges.

State-based models

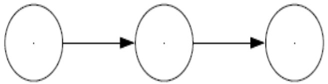


Applications:

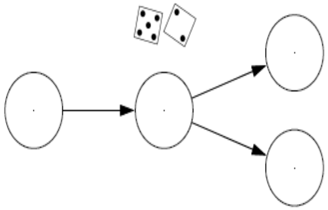
- Games: Chess, Go, Pac-Man, Starcraft, etc.
- Robotics: motion planning

State-based models

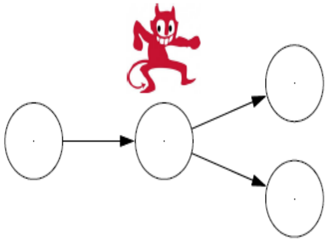
Search problems: you control everything



Markov decision processes: against nature (e.g., Blackjack)



Adversarial games: against opponent (e.g., chess)



- Search problems are models where you are operating in an environment that has no uncertainty. However, in many realistic settings, there are other forces at play.
- Markov decision processes handle situations where there is randomness (e.g., Blackjack).
- Adversarial games, as the name suggests, handle tasks where there is an opponent who is working against you (e.g., chess).

Variable-based models

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

→

5	3	4	6	7	8	9	1	2
6	7	2	1	9	5	3	4	8
1	9	8	3	4	2	5	6	7
8	5	9	7	6	1	4	2	3
4	2	6	8	5	3	7	9	1
7	1	3	9	2	4	8	5	6
9	6	1	5	3	7	2	8	4
2	8	7	4	1	9	6	3	5
3	4	5	2	8	6	1	7	9

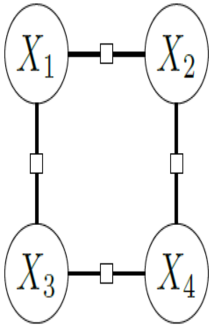
Goal: put digits in blank squares so each row, column, and 3x3 sub-block has digits 1–9

Key: order of filling squares doesn't matter in the evaluation criteria!

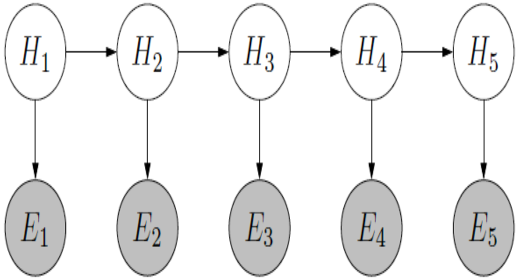
- In state-based models, solutions are procedural: they specify step by step instructions on how to go from A to B.
- In some applications, the order in which things are done isn't important.
- For example, in Sodoku, where the goal is to put digits in the blank squares to satisfy some constraints, all that matters is the final configuration of numbers; you can fill them in in any order.

Variable-based models

Constraint satisfaction problems: hard constraints (e.g., Sudoku, scheduling)



Bayesian networks: soft dependencies (e.g., tracking cars from sensors)



- Constraint satisfaction problems are variable-based models where we only have hard constraints. For example, in scheduling, one person can't be in two places at once.
- Bayesian networks are variable-based models where variables are random variables which are dependent on each other. For example, the true location of an airplane H_t and its radar reading E_t are related, as are the location H_t and the location at the last time step H_{t-1}. The exact dependency structure is given by the graph and it formally defines a joint probability distribution over all of the variables.

Logic

- Logic provides a class of models which is higher-level, but still needs to be supported by the groundedness to real data that machine learning offers.
- One motivation for logic is a virtual assistant. A good assistant should be able to remember what you told it and answer questions that require drawing inferences from its knowledge. And you'd want to interact with it using natural language, the tool that humans invented for communication.

Motivation: virtual assistant

Tell information →  ← **Ask** questions

Use natural language!

[demo]

Need to:

- Digest **heterogenous** information
- Reason **deeply** with that information

Logic

- Interacting with this system feels very different than a typical machine learning-based system. First, it is adaptive, whereas most ML systems are a fixed function. Second, the types of information we provide it and the types of questions are more heterogenous and more abstract, and and we expect the system to realize the full consequences of every single word.
- Recently, with the emergence of large language models like ChatGPT, we are beginning to see virtual assistants with far greater capabilities. However, LLMs are well-known for their hallucinations, whereas the logic-based system is 100% internally consistent.
- One often contrasts logical AI and statistical AI. In this course, we will treat the two as not contradictory but rather complementary. Logic provides a class of models which is higher-level, but still needs to be supported by the groundedness to real data that machine learning offers.

- Recently, with the emergence of large language models like ChatGPT, we are beginning to see virtual assistants like the one demoed here,
- but with far greater capabilities. However, LLMs are well-known for their hallucinations, whereas the logic-based system is 100% internally
- consistent.
- One often contrasts logical AI and statistical AI. In this course, we will treat the two as not contradictory but rather complementary. Logic
- provides a class of models which is higher-level, but still needs to be supported by the groundedness to real data that machine learning oers.

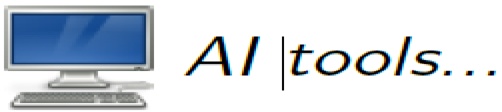
Two views of AI

AI agents: how can we re-create intelligence?

AI tools: how can we benefit society?

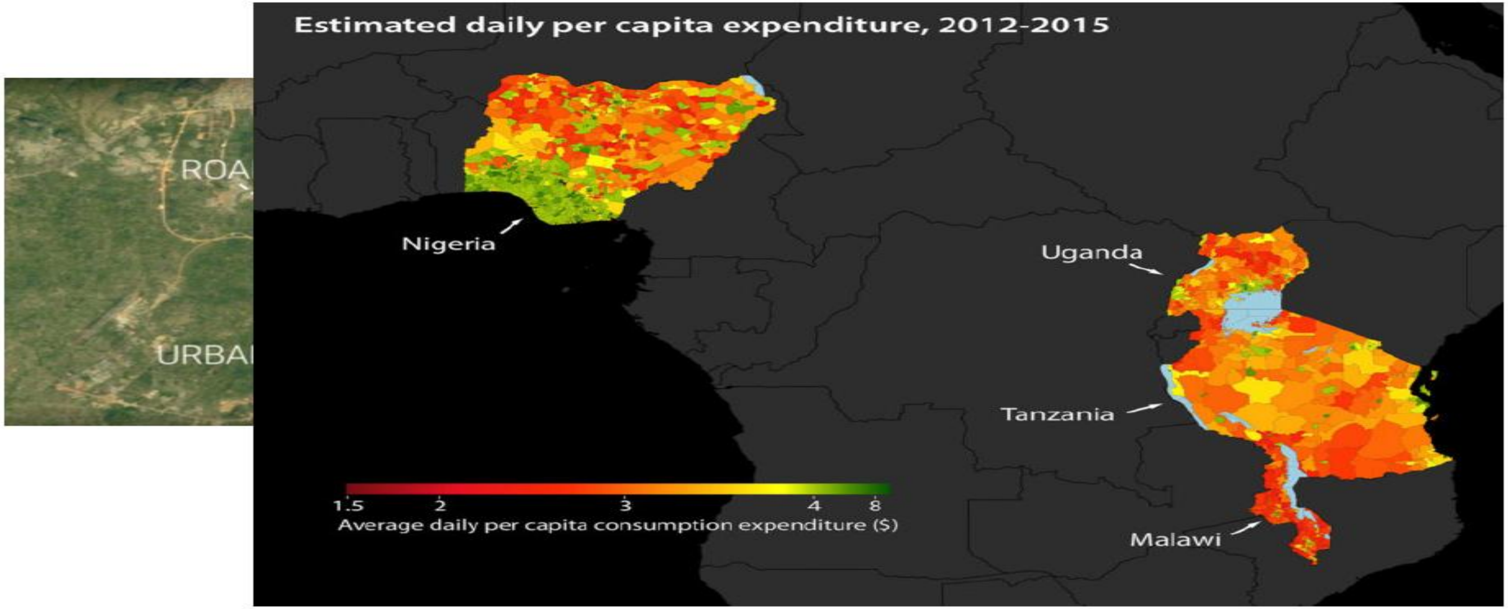
- There are two ways to look at AI philosophicaly.
    - The first is what one would normally associate with the AI: the science and engineering of building "intelligent" agents. The inspiration of what constitutes intelligence comes from the types of capabilities that humans possess: the ability to perceive a very complex world and make enough sense of it to be able to manipulate it.
    - The second views AI as a set of tools. We are simply trying to solve problems in the world, and AI techniques happen to be quite useful for that.
- However, both views boil down to many of the same day-to-day activities (e.g., collecting data and optimizing a training objective), the philosophical differences do change the way AI researchers approach and talk about their work. And moreover, the conation of these two can generate a lot of confusion. (Ref: CS221)

AI Tools



- The AI agents view is an inspiring quest to undercover the mysteries of intelligence and tackle the tasks that humans are good at. While there has been a lot of progress, we still have a long way to go along some dimensions: for example, the ability to learn quickly from few examples or the ability to perform commonsense reasoning.
- At the same time, the current level of technology is already being deployed widely in practice. These settings are often not particularly human-like (targeted advertising, news or product recommendation, web search, supply chain management, etc.)

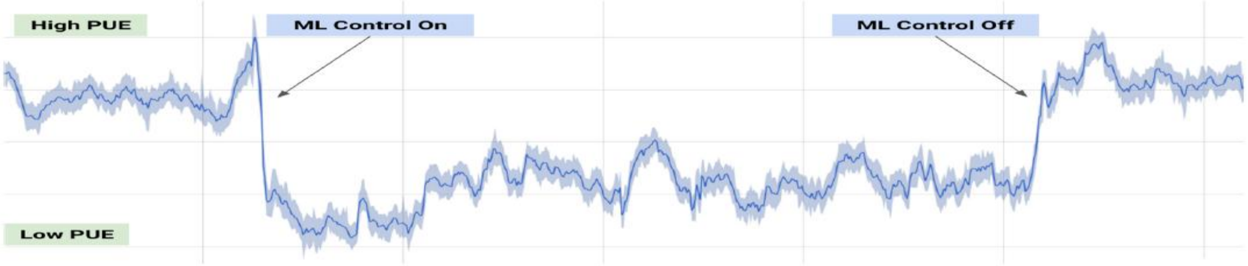
Predicting poverty



- The computer vision techniques, used to recognize objects, can also be used to tackle social problems. Poverty is a huge problem, and even identifying the areas of need is difficult due to the difficulty in getting reliable survey data. Recent work has shown that one can take satellite images (which are readily available) and predict various poverty indicators.

Saving energy by cooling datacenters

Saving energy by cooling datacenters



- Machine learning can also be used to optimize the energy efficiency of datacenters, which given the hunger for compute these days makes a big difference. Some recent work from DeepMind shows how to significantly reduce Google's energy footprint by using machine learning to predict the power usage effectiveness from sensor measurements such as pump speeds, and using that to drive recommendations.

An intelligent agent

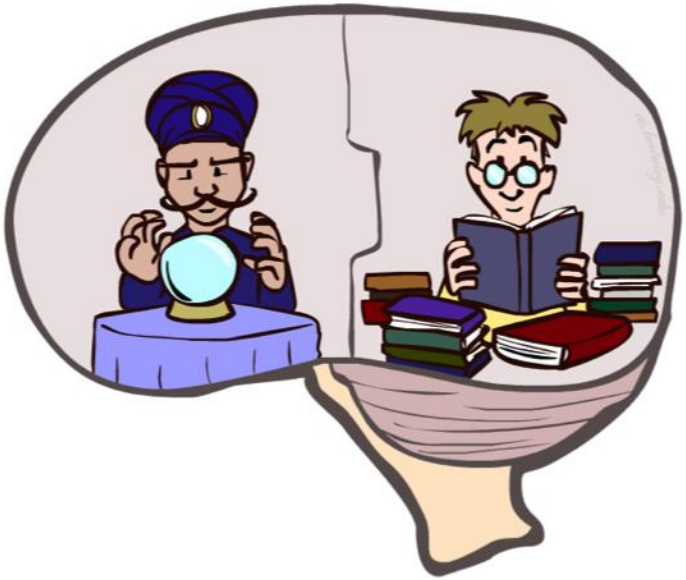
Perception | Robotics | Language

Knowledge | Reasoning | Learning

- The starting point for the agent-based view is ourselves.
- As humans, we have to be able to perceive the world (computer vision), perform actions in it (robotics), and communicate with other agents.
- We also have knowledge about the world (from how to ride a bike to knowing the capital of France), and using this knowledge we can draw inferences and make decisions.
- Finally, we learn and adapt over time. Indeed machine learning has become the primary driver of many of the AI applications we see today.

What About the Brain?

- Brains (human minds) are very good at making rational decisions, but not perfect
- Brains aren't as modular as software, so hard to reverse engineer!
- "Brains are to intelligence as wings are to flight"
- We can't yet build AI on the scale of the brain
  - ~100T synapses in the human brain
    vs
  - ~1.8T weights in GPT4
- Still, the brain can be a great inspiration for AI!

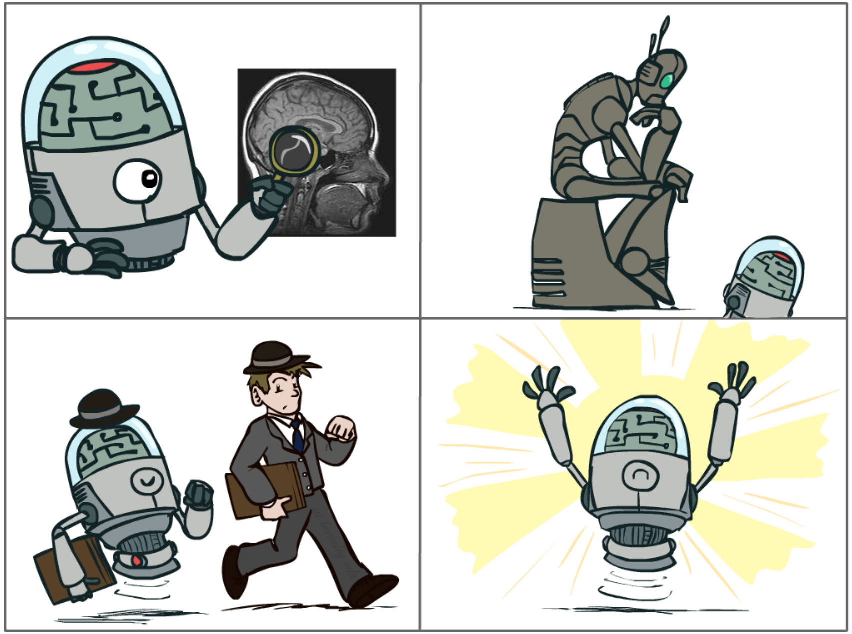


What is AI?

The science of making machines that:

Think like people

Think rationally



Act like people

Act rationally

Rational-agent Approach in AI

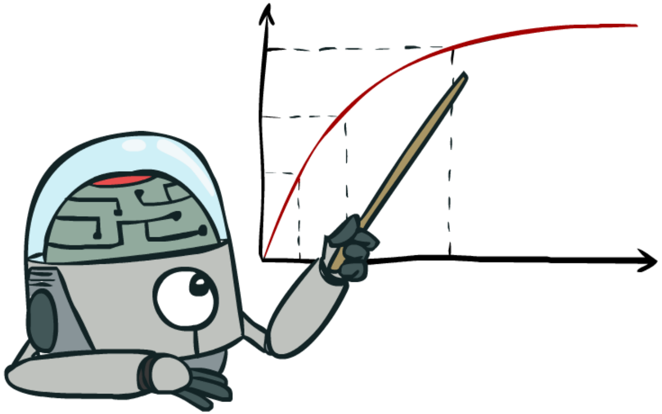
- AI has focused on the study and construction of agents that do the right thing. What counts as the right thing is defined by the objective that we provide to the agent. This general paradigm is so pervasive that we might call it the standard model.
- A rationalist approach involves a combination of mathematics and engineering, and connects to statistics, control theory, and economics.
- It prevails not only in AI, but also in control theory, where a controller minimizes a cost function; in operations research, where a policy maximizes a sum of rewards; in statistics, where a decision rule minimizes a loss function; and in economics, where a decision maker maximizes utility or some measure of social welfare.
- We need to make one important refinement to the standard model to account for the fact that perfect rationality—always taking the exactly optimal action—is not feasible in complex environments. So, we have a limited rationality.
- However, perfect rationality often remains a good starting point for theoretical analysis.

Rational Decisions

We'll use the term **rational** in a very specific, technical way:

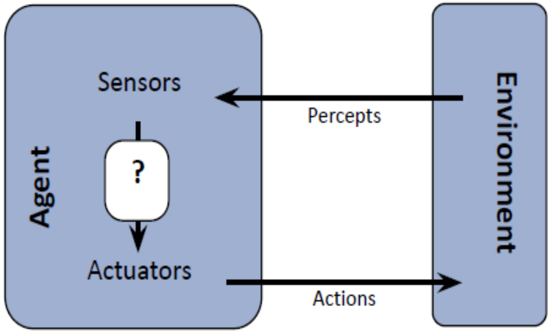
- Rational: maximally achieving pre-defined goals
- Rationality only concerns what decisions are made (not the thought process behind them)
- Goals are expressed in terms of the **utility** of outcomes
- Being rational means **maximizing your expected utility**

Maximize Your Expected Utility



AIMA Book View: Designing Rational Agents

- An **agent** is an entity that perceives and acts.
- A **rational agent** selects actions that maximize its (expected) **utility**.
- Characteristics of the **percepts**, **environment**, and **action space** dictate techniques for selecting rational actions
- This course is about:
  - General AI techniques for a variety of problem types
  - Learning to recognize when and how a new problem can be solved with an existing technique



Course Topics

Core Components of Rational Agents:

- Search & Planning
- Probability & Inference
- Supervised Learning
- Reinforcement Learning

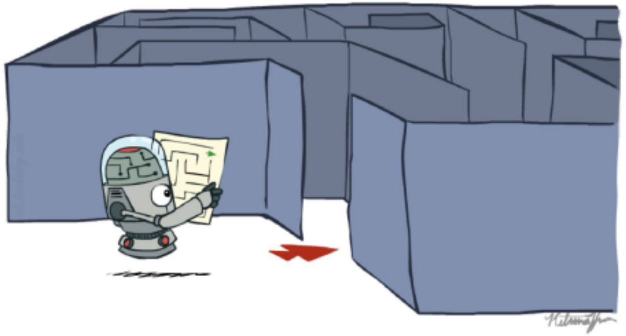
Course Topics

Search & Planning

Probability & Inference

Supervised Learning

Reinforcement Learning



How can I use my ***model*** of the world to find a ***sequence of actions*** to achieve my ***goal***?

sequence of actions to achieve my ***goal***?

Course Topics

- Search & Planning
- Probability & Inference
- Supervised Learning
- Reinforcement Learning

How can I make sense of ***uncertainty***?

Course Topics

Search & Planning

Probability & Inference

Supervised Learning

Reinforcement Learning

How can I learn a ***model*** of the world from ***data***?

Course Topics

Search & Planning

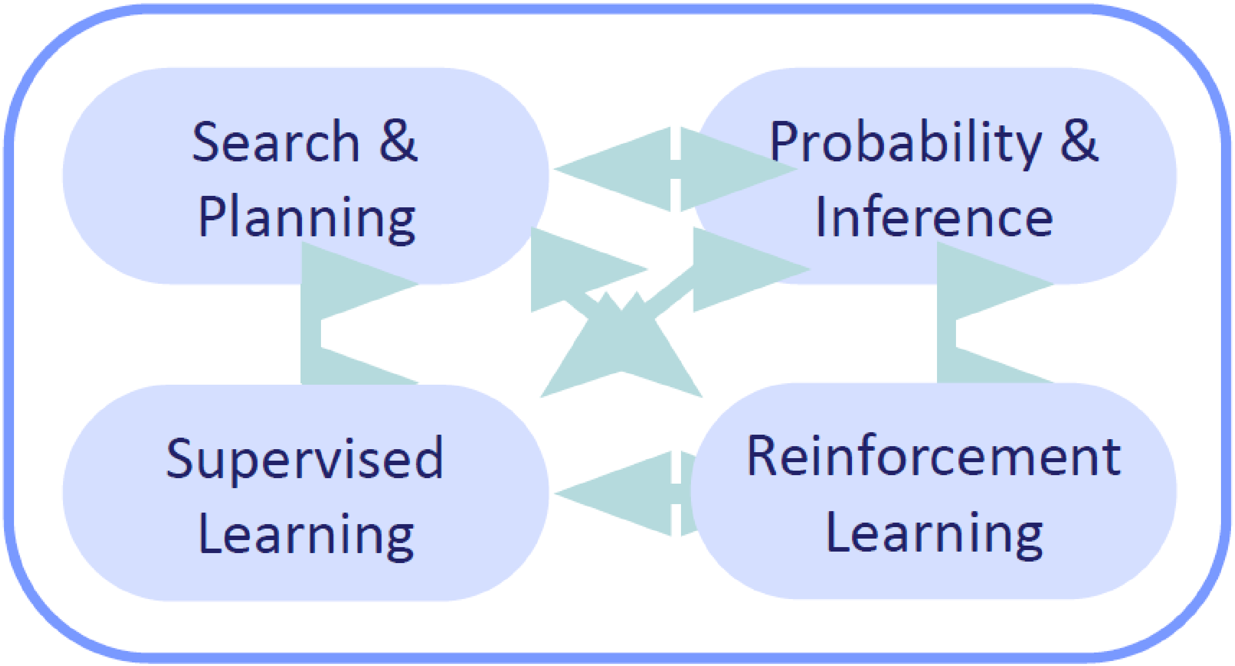
Probability & Inference

Supervised Learning

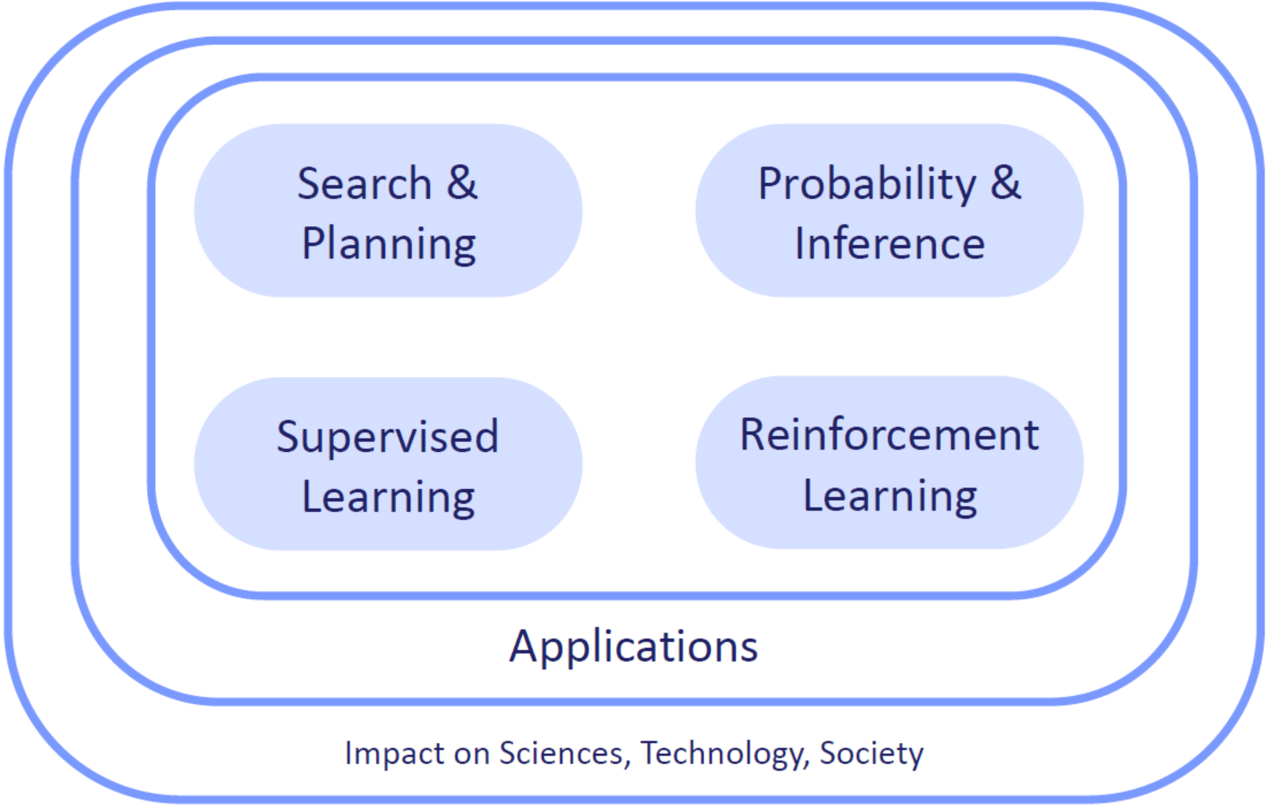
Reinforcement Learning

How can I learn a ***policy*** for any situation so that I can ***maximize utility***?

Course Topics



Course Topics



Tentative Course Plan

- As time permits, we plan to cover the following topics in this course:
  - Various search techniques to optimize the utility (including adversarial, and online settings)
  - Constraint satisfaction problems
  - Automated logical inference
  - Inference under uncertainty and Bayes nets
  - Temporal probability models (such as the Markov models)
  - Introduction to Machine Learning, Deep Learning, Reinforcement Learning
  - Introduction to Natural Language Processing, Deep Learning for Natural Language Processing
  - Introduction to Computer Vision, Robotics

Table of Contents

Table of Contents for the US Edition (or see the Global Edition)