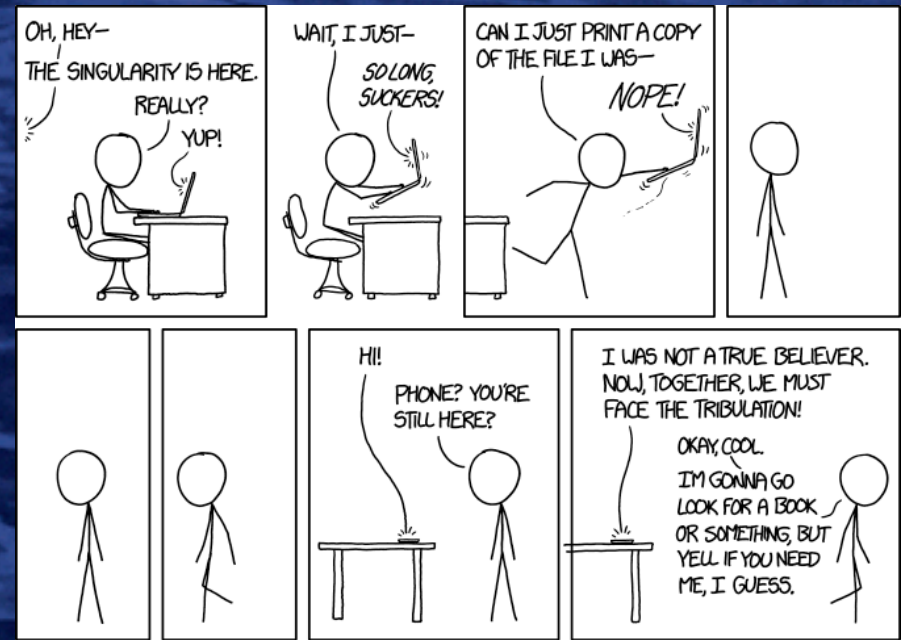


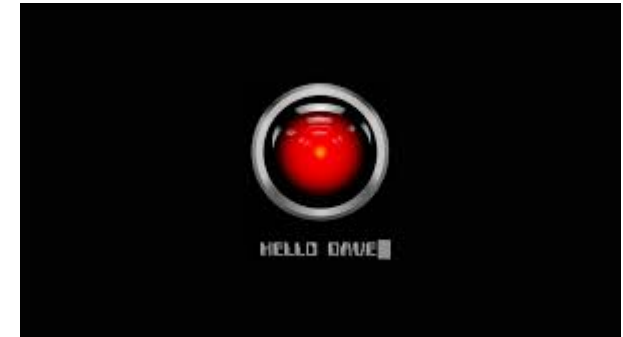
# Agents and Introduction to AI

CITS3001 Algorithms, Agents and Artificial Intelligence

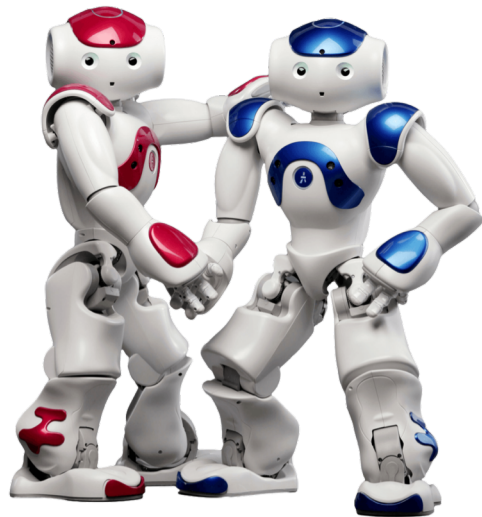


# Introduction

- We will consider what is meant by the terms
  - *Artificial intelligence*
  - *Agents*
- We will define
  - Four ways of looking at the former
  - Four general models for the latter



A STRANGE GAME.  
THE ONLY WINNING MOVE IS  
NOT TO PLAY.  
HOW ABOUT A NICE GAME OF CHESS?



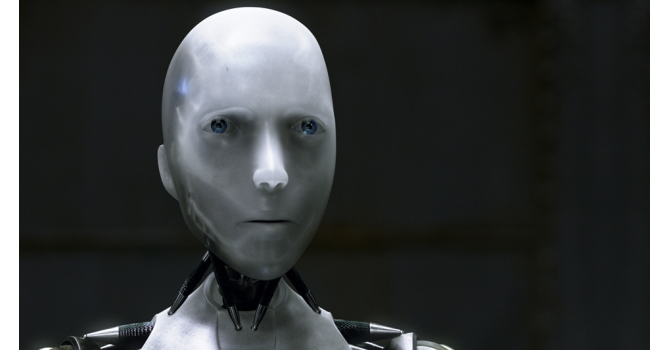
# What is Artificial Intelligence

- Given that experts can't even agree on a definition for the word "intelligence", what does "AI" mean!?
- Movies
  - Kubrick's *2001: a Space Odyssey*, 1968 ("I'm sorry Dave, I'm afraid I can't do that")
  - Cameron's *The Terminator*, 1984 (Skynet)
  - Proyas's *I, Robot*, 2004 (could you kill a robot?) <http://www.bbc.com/future/story/20131127-would-you-murder-a-robot>)
- TV "science shows"
  - *Towards 2000, Beyond 2000, Beyond Tomorrow, ...*
- News/current affairs
  - Deep Blue vs. Kasparov
  - Watson vs. the best of the best
  - AlphaGo vs Lee Sedol
  - Japan: Robot to take top university exam <http://www.bbc.com/news/blogs-news-from-elsewhere-26418431>
  - Robots will be smarter than us all by 2029 <http://www.independent.co.uk/life-style/gadgets-and-tech/news/robots-will-be-smarter-than-us-all-by-2029-warns-ai-expert-ray-kurzweil-9147506.html>
- Adverts
  - Intelligent TVs, washers, cars, molecules...

# AI in research....

“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ...”  
[Bellman, 1978]

“The study of the computations that make it possible to perceive, reason, and act.” [Winston, 1992]



“The art of creating machines that perform functions that require intelligence when performed by people.” [Kurzweil, 1990]

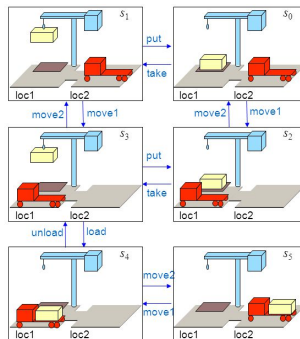
“Computational Intelligence is the study of the design of intelligent agents” [Poole *et al.*, 1998]

- Views of AI fall into four categories:

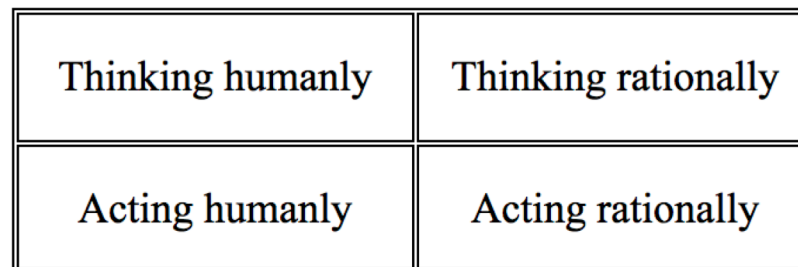
### Example

- $\Sigma = (S, A, E, \gamma)$
- $S = \{\text{states}\}$
- $A = \{\text{actions}\}$
- $E = \{\text{exogenous events}\}$
- State-transition function  $\gamma: S \times (A \cup E) \rightarrow 2^S$

- Example:
  - $S = \{s_0, \dots, s_5\}$
  - $A = \{\text{move1, move2, put, take, load, unload}\}$
  - $E = \{\}$
  - $\gamma$ : see the arrows



Dock Worker Robots (DWR) example



“thought”  
↑  
↓  
“behaviour”

“human”

“ideal”

# Thinking Humanly

- Determine how humans think, and attempt to replicate it in software/hardware
- Develop a theory of the human mind, by one or more of
  - Introspection
  - Psychological experiments (top-down?)
  - Brain imaging (bottom-up?)
- What level of abstraction is best?
  - “Knowledge” or “circuits”?
  - Should we model the “mind” or the “brain”?
- And how would we validate such a system?
- e.g. the General Problem Solver (GPS)  
[Newell & Simon, 1961]
  - Attempted to “solve like a human”
  - No searching
- *The question of whether machines can think ... is about as relevant as the question of whether submarines can swim* Edsger Dijkstra

# Acting Humanly

- Intelligence = the ability to act indistinguishably from a human in cognitive tasks(?)
- An operational test: the Turing Test [Alan Turing 1950]
  - $H$  interrogates  $X$  in a black box
    - If  $X$  is a computer, but  $H$  cannot tell,  $X$  must be intelligent!
  - Loebner Prize
    - <http://www.loebner.net/Prizef/loebner-prize.html>
    - Basically an online Turing Test
  - Botprize
    - <http://botprize.org/>
    - Can computers play like people?
  - GECCO “humies”
    - <http://www.sigevo.org/gecco-2014/humies.html>
    - Prizes for human-competitive results



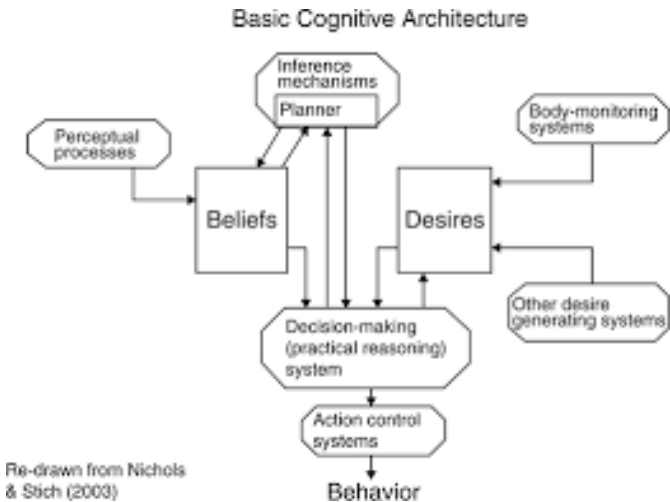
# Thinking Rationally

- Codify “laws of thought” or “right-thinking”
  - Irrefutable reasoning processes
  - Independent of what humans do
- *All men are mortal*
- *Socrates is a man*
- *Therefore Socrates is mortal*
- Captured in rules of inference
  - *modus ponens*:  $(P \wedge (P \rightarrow Q)) \rightarrow Q$
  - *modus tollens*:  $(\neg Q \wedge (P \rightarrow Q)) \rightarrow \neg P$
  - *absorption*:  $(P \rightarrow Q) \rightarrow (P \rightarrow (P \wedge Q))$
  - Many others
- Problems include
  - Difficulty in codifying informal knowledge
  - Difficulty in dealing with uncertainty
  - Scalability issues

$$\begin{array}{c}
 \frac{}{\Delta, B \rightarrow B} \text{initial}^\dagger \quad \frac{B, C, \Delta \rightarrow G}{B \wedge C, \Delta \rightarrow G} \wedge L^\dagger \\
 \frac{\Delta \rightarrow B \quad \Delta \rightarrow C}{\Delta \rightarrow B \wedge C} \wedge R^\dagger \quad \frac{B, \Delta \rightarrow G \quad C, \Delta \rightarrow G}{B \vee C, \Delta \rightarrow G} \vee L^\dagger \\
 \frac{\Delta \rightarrow B}{\Delta \rightarrow B \vee C} \vee R \quad \frac{\Delta \rightarrow C}{\Delta \rightarrow B \vee C} \vee R \\
 \frac{B, \Delta \rightarrow C}{\Delta \rightarrow B \supset C} \supset R^\dagger \quad \frac{C, B, \Delta \rightarrow G}{B \supset C, B, \Delta \rightarrow G} \supset L_1 \\
 \frac{A \supset (B \supset C), \Delta \rightarrow G}{(A \wedge B) \supset C, \Delta \rightarrow G} \supset L_2^\dagger \quad \frac{A \supset C, B \supset C, \Delta \rightarrow G}{(A \vee B) \supset C, \Delta \rightarrow G} \supset L_3^\dagger \\
 \frac{B \supset C, \Delta \rightarrow A \supset B \quad C, \Delta \rightarrow G}{(A \supset B) \supset C, \Delta \rightarrow G} \supset L_4
 \end{array}$$

# Acting Rationally

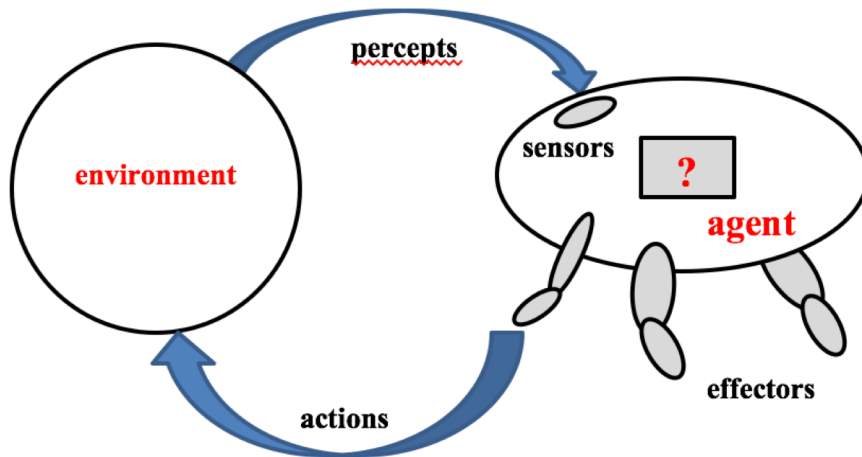
- Act in such a way as to achieve goals, given beliefs
- Define an agent, and give it
  - Some goals
  - The ability to perceive its surroundings
  - The ability to perform actions
  - The ability to “reason”
- It will (try to) find actions to achieve the goals
- Note that this doesn’t necessarily involve “thinking”
  - e.g. is a thermostat intelligent?
- This gives us an engineering viewpoint
  - Can we develop systems that do useful stuff?
  - Or even cool stuff!?
- We have a proof of concept, after all
- Our view (the modern view) of AI is as the *study, design, and construction of intelligent agents*
- For previous significant views, read up on the foundations and history of AI
  - Section 1.2–1.4 of AIMA





# So, What is an Agent?

- An agent
  - *Perceives* its environment through *sensors*
  - *Acts* on its environment through *effectors*



<b><u>Percepts</u></b>	Light, sound, solidity, ...
<b>Sensors</b>	
human	Eyes, ears, skin, ...
robot	Cameras, mike, accelerometers, ...
software	Keyboard, mouse, files, n/w packets, ...
<b>Effectors</b>	
human	Hands, legs, voice, ...
robot	Wheels, speakers, grippers, ...
software	Screen, printer, files, n/w packets, ...
<b>Actions</b>	Pick up, speak, throw, ...

# Rational Agents

- A *rational agent* tries to “do the right thing” wrt a set of goals or utilities
- The right thing can be specified by a *performance measure* defining a numerical value for any environment history
- A rational action is whatever action maximises the expected value of the performance measure, given the current state of the environment and the percept sequence to date
- But note that
  - Rational  $\neq$  Omniscient
  - Rational  $\neq$  Clairvoyant
  - Rational  $\neq$  Successful
- It is entirely possible to do the right thing and to fail anyway
  - Sometimes randomness is the most rational choice!
  - e.g. games
- An agent’s behaviour is specified by an *agent function* mapping percept sequences to actions
  - The agent will usually also store knowledge or rules that help it to understand and to select actions
- We will discuss four basic types of agents, in order of generality

# Simple Reflex Agents

- Choose an action using *condition-action rules*
  - e.g. **if** the car-in-front's brake-lights come on **then** apply your brakes
- The key word in this diagram is *now*
  - No history is stored
  - Some experts believe that this is how simple life-forms (e.g. insects) behave

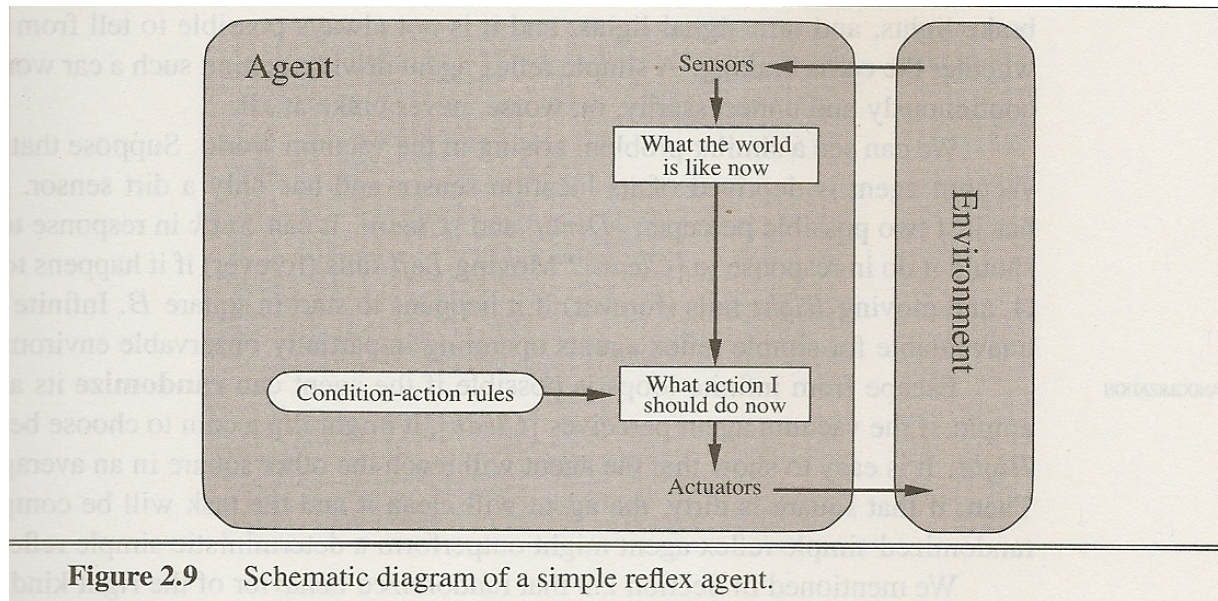


Figure 2.9 Schematic diagram of a simple reflex agent.

# Model-based Reflex Agents

- While simply reacting to the (current) world is adequate in some circumstances, most intelligent action requires more knowledge
  - Stored memory of the past
  - Understanding of the effects of actions
- Both of these require *internal state*
  - e.g. you see a pedestrian ahead signal to a bus
  - You know the bus will stop
  - You should change lanes
- Also this allows much better for worlds that are only partially observable
  - Which is by far the most common case

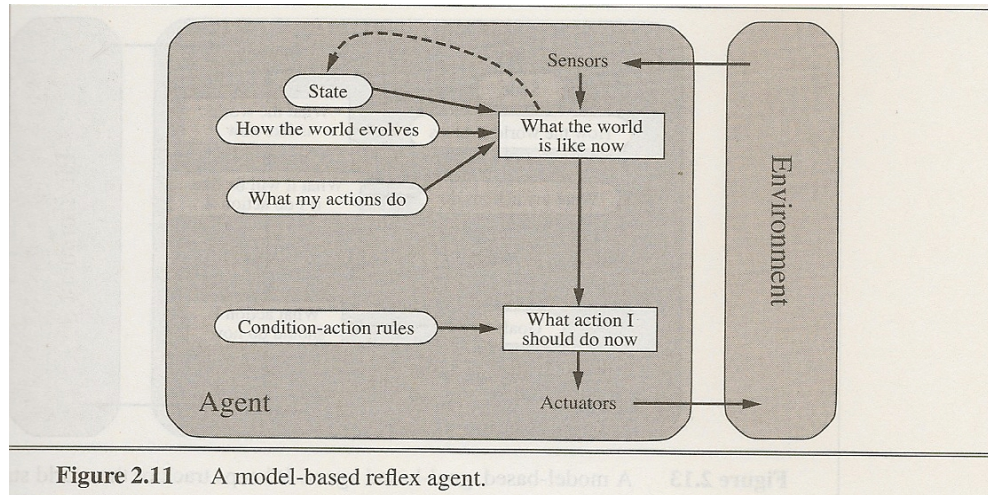
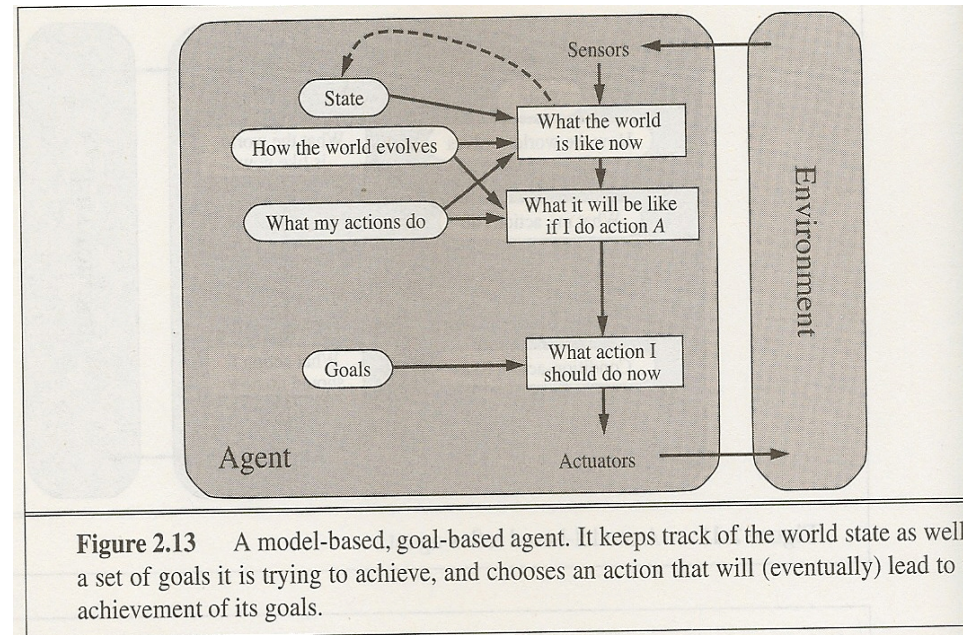


Figure 2.11 A model-based reflex agent.

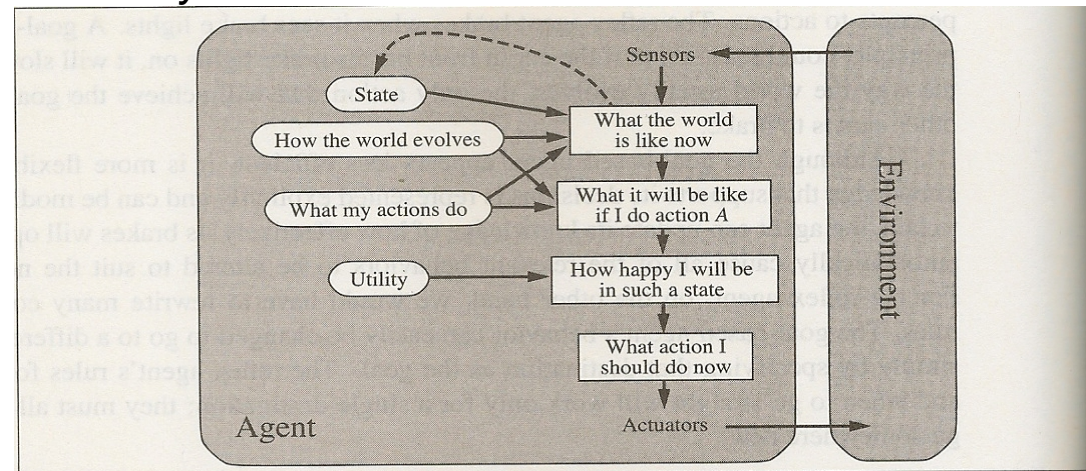
# Goal-based Agents

- Reacting better to the changing world is an improvement
  - But what are we trying to achieve?
- Intelligent (and some other!) beings have *goals*
  - e.g. at a junction, which way do we turn?
- Achieving goals involves predicting the future
  - If I do this action, how will that change the world?
- Some goals are simple
  - Star Trek: boldly go where no one has gone before
- Other goals are complex and require *planning*
  - Star Wars: defeat the Empire!
- Planning is fundamental and usually requires *search*



# Utility-based Agents

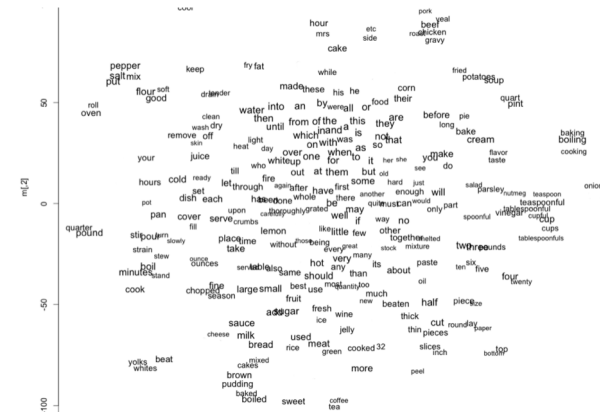
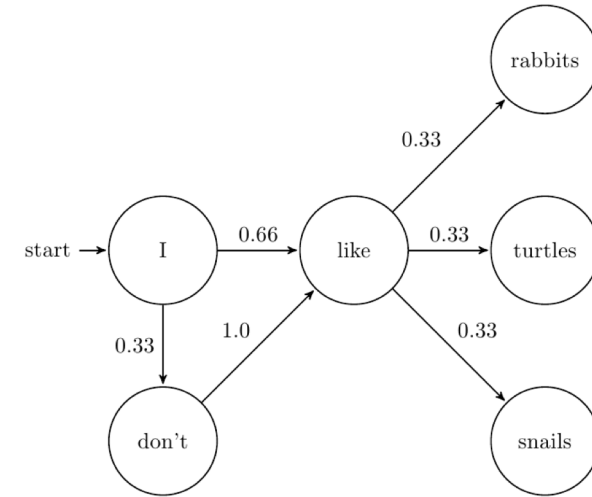
- A goal is a binary thing
  - Achieve it or fail!
- Most outcomes are more continuously-measured
  - e.g. which action will make me happier? Or richer?
- Usually defined as a *utility* to be maximised
  - cf. optimisation problems
- Again partial observability rears its head
  - The agent will try to maximise *expected utility*



**Figure 2.14** A model-based, utility-based agent. It uses a model of the world, along with a utility function that measures its preferences among states of the world. Then it chooses the action that leads to the best expected utility, where expected utility is computed by averaging over all possible outcome states, weighted by the probability of the outcome.

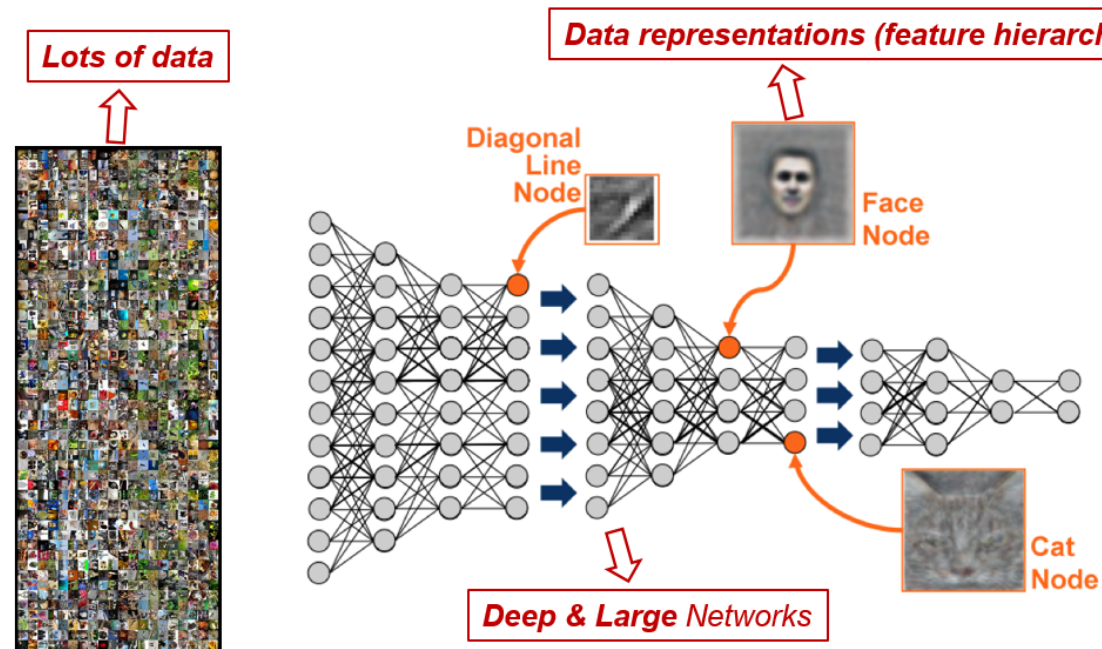
# Aspects of AI: Natural Language Processing

- Natural language processing is the process of applying meaning to text.
- More than just transcribing spoken word, this requires a machine to understand the intent and meaning behind a sentence, which requires semantic knowledge of the world.
- Closely related to machine translation, there has been consistent effort in this area throughout the history of AI
- Rule based approaches, apply grammatical rules and pattern matching approaches.
- Statistical approaches build statistical models of meaning by sampling large corpuses of text.
- Deep Neural Networks are the most recent and successful approach and create embeddings of text into high dimensional spaces where regions have semantic meaning.



# Aspects of AI: Computer Vision

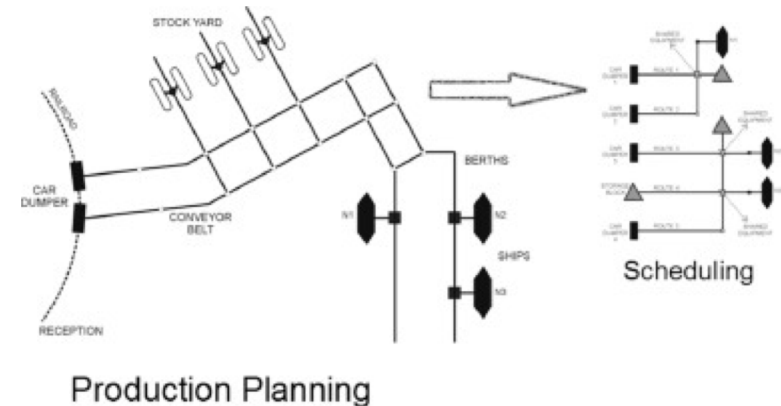
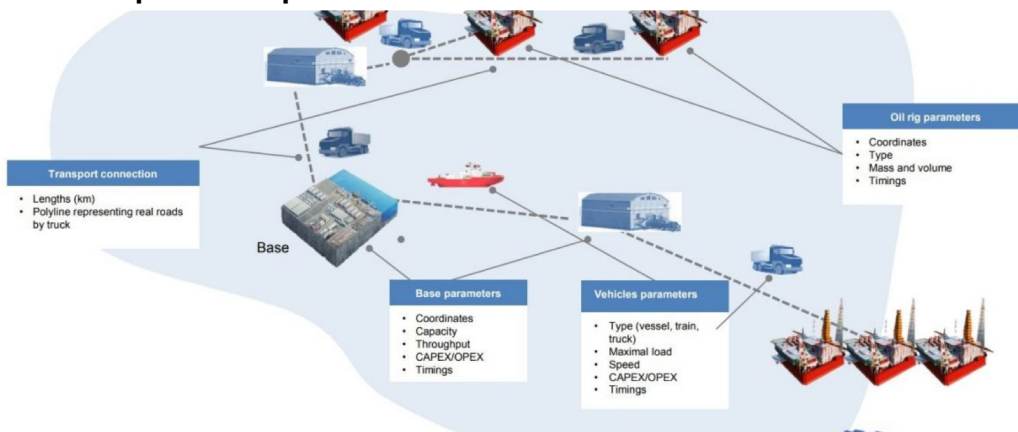
- Computer Vision has made exceptional advances in recent times, largely based on the advent of convolutional neural networks and deep learning.
- Images and videos are represented as high dimensional vectors, and passed through many layers of a neural network.
- They are tagged (classified) and the neural network is trained to recognize the classification via back propagation.
- Different layers of the neural network often correspond to recognizable features in the image.





# Aspects of AI: Optimization

- Optimization and Operations Research are branches of applied mathematics not traditionally a part of artificial intelligence.
- However, many aspects of autonomous control and automation rely on a machine selecting a good or best action.
- From an implementation point of view, this is often achieved by building abstract simulations of the environment, and trialling different actions to see which provides the greatest utility.
- A large part of this task is modelling and prediction: how should the simulated environment respond to different actions, and what reward can we expect. Statistical learning is often applied here.
- Once a model is provided, generic branch and bound/brute force methods can be used to explore options.



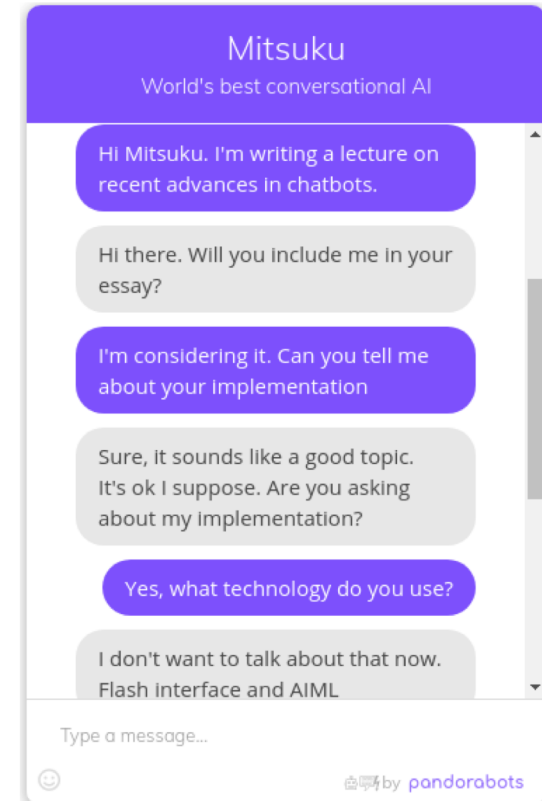
# Aspects of AI: Argumentation and Reasoning

- Argumentation and reasoning is the process of applying logical deductions and inferences to reach a conclusion from a premise.
- It is an intrinsic part of bargaining, negotiation, and social interactions.
- Despite the substantial advances in deep learning, recent advances in reasoning have been much more modest.
- Reasoning is typically done via rule based systems, where premises are transformed and match with patterns, where deductions can be extracted.
- Programming systems like Prolog and LISP can encode such rules, but then the reasoning typically proceeds through a depth first search.

```
mammal (X) :      has_hair (X).
                  warm_blooded (X).
                  has_hair (dog).
                  has_hair (caterpillar).
                  warm_blooded (robin).
                  warm_blooded (dog).
```

```
When we request a list of all mammals
mammal (X) ?
we get
```

$X = dog$



Mitsuku  
World's best conversational AI

Hi Mitsuku. I'm writing a lecture on recent advances in chatbots.

Hi there. Will you include me in your essay?

I'm considering it. Can you tell me about your implementation

Sure, it sounds like a good topic. It's ok I suppose. Are you asking about my implementation?

Yes, what technology do you use?

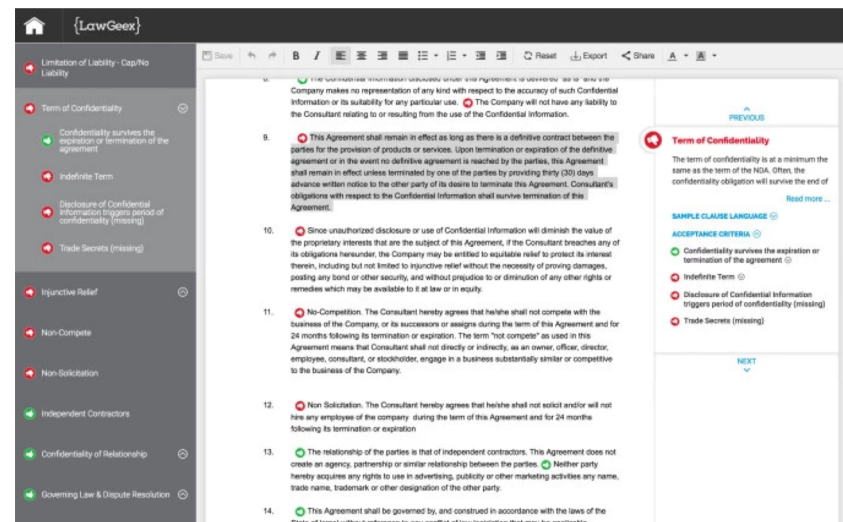
I don't want to talk about that now. Flash interface and AIML

Type a message...

by pandorobots

# AI and job automation

- A lot of media focus has been on the opportunity of AI in the workforce, and the threat to traditional vocations.
- Checkouts, bank tellers, typists are traditional professions that are already heavily automated.
- There is a significant investment in autonomous road vehicles which could have a massive impact on the workforce, but many legislative barriers remain.
- White collar automation refers to the process of automating routine tasks in accounting, law, medicine and other traditional professional occupations. Predictive systems and pattern matching can provide support to many of the typical tasks in these professions, such as contract review, filing tax forms, or matching conditions to symptoms.



The screenshot displays a legal document review tool. On the left, a sidebar lists various clauses with status indicators (red for issues, green for compliance). The main area shows a 'Confidentiality' clause with red annotations highlighting specific terms and conditions. The right sidebar provides a detailed view of the 'Term of Confidentiality' clause, including its duration and termination conditions.

**Limitation of Liability - Cap/No Liability**

**Term of Confidentiality**

- Confidentiality survives the expiration or termination of the agreement.
- Indefinite Term
- Disclosure of Confidential Information (against period of confidentiality missing)
- Trade Secrets (missing)
- Injunctive Relief
- Non-Compete
- Non-Solicitation
- Independent Contractors
- Confidentiality of Relationship
- Governing Law & Dispute Resolution

**9.** This Agreement shall remain in effect as long as there is a definitive contract between the parties for the provision of products or services. Upon termination or expiration of the definitive agreement or in the event no definitive agreement is reached by the parties, this Agreement shall remain in effect unless terminated by one of the parties by providing thirty (30) days advance written notice to the other party of its desire to terminate this Agreement. Consultant's obligations with respect to the Confidential Information shall survive termination of this Agreement.

**10.** Since unauthorized disclosure or use of Confidential Information will diminish the value of the proprietary interests that are the subject of this Agreement, if the Consultant breaches any of its obligations hereunder, the Company may be entitled to equitable relief to protect its interest therein, including but not limited to injunctive relief without the necessity of proving damages, posting any bond or other security, and without prejudice to or diminution of any other rights or remedies which may be available to it at law or in equity.

**11.** **No-Competition.** The Consultant hereby agrees that he/she shall not compete with the business of the Company, or its successors or assigns during the term of this Agreement and for 24 months following its termination or expiration. The term "not compete" as used in this Agreement means that Consultant shall not directly or indirectly, as an owner, officer, director, employee, consultant, or stockholder, engage in a business substantially similar or competitive to the business of the Company.

**12.** **Non-Solicitation.** The Consultant hereby agrees that he/she shall not solicit and/or will not hire any employees of the company during the term of this Agreement and for 24 months following its termination or expiration.

**13.** The relationship of the parties is that of independent contractors. This Agreement does not create an agency, partnership or similar relationship between the parties. Neither party hereby acquires any rights to use in advertising, publicity or other marketing activities any name, trade name, trademark or other designation of the other party.

**14.** This Agreement shall be governed by, and construed in accordance with the laws of the State of Israel without reference to any conflict of law legislation that may be applicable.

**Term of Confidentiality**

The term of confidentiality is at a minimum the same as the term of the NDA. Often, the confidentiality obligation will survive the end of

Read more ...

**SAMPLE CLAUSE LANGUAGE**

- Confidentiality survives the expiration or termination of the agreement
- Indefinite Term
- Disclosure of Confidential Information triggers period of confidentiality (missing)
- Trade Secrets (missing)