

# AITA : The Roots, Goals and Subfields of AI

© John A. Bullinaria, 2003

1. The Goals
2. The Roots  
Philosophy, Logic/Mathematics,  
Computation, Cognitive Science/Psychology,  
Biology/Neuroscience, Evolution
3. The Sub-fields  
Learning, Neural Networks, Evolutionary Computation,  
Vision, Robotics, Expert Systems, Speech Processing,  
Natural Language Processing, Planning, ...
4. Common Techniques  
Representation, Learning, Rule Systems, Search, ...

## The Goals

“Artificial Intelligence (AI) is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit characteristics we associate with intelligence in human behaviour – understanding language, learning, reasoning, solving problems, and so on.” (Barr & Feigenbaum, 1981)

***Engineering Goal*** To solve real world problems using AI techniques such as knowledge representation, learning, rule systems, search, and so on.

***Scientific Goal*** To determine which ideas about knowledge representation, learning, rule systems, search, and so on, explain various sorts of real intelligence.

Traditionally, computer scientists and engineers have been more interested in the engineering goal, while psychologists, philosophers and cognitive scientists have been more interested in the scientific goal. It makes good sense to be interested in both, as there are common techniques and the two approaches can feed off each other. In this module we shall attempt to keep both goals in mind.

# The Roots

Artificial Intelligence has identifiable roots in a number of older disciplines, particularly:

Philosophy

Logic/Mathematics

Computation

Psychology/Cognitive Science

Biology/Neuroscience

Evolution

There is inevitably much overlap, e.g. between philosophy and logic, or between mathematics and computation.

By looking at each of these in turn, we can gain a better understanding of their role in AI, and how these underlying disciplines have developed to play that role.

# Philosophy

**~400 BC** Socrates asks for an *algorithm* to distinguish piety from non-piety.

**~350 BC** Aristotle formulated different styles of *deductive reasoning*, which could mechanically generate conclusions from initial premises, e.g.

Modus Ponens          If  $A \Rightarrow B$  and  $A$  then  $B$   
                                 If  $A$  implies  $B$  and  $A$  is true then  $B$  is true  
                                 *when it's raining you get wet and it's raining then you get wet*

**1596 – 1650** Rene Descartes idea of mind-body *dualism* – part of the mind is exempt from physical laws. Otherwise how do we have *free will*?

**1646 – 1716** Wilhelm Leibnitz was one of the first to take the *materialist* position which holds that the mind operates by ordinary physical processes – this has the implication that mental processes can potentially be carried out by machines.

## Logic/Mathematics

**1777** Earl Stanhope's Logic Demonstrator was a machine that was able to solve syllogisms, numerical problems in a logical form, and elementary questions of probability.

**1815 – 1864** George Boole introduced his formal language for making logical inference in 1847 – Boolean algebra.

**1848 – 1925** Gottlob Frege produced a logic that is essentially the first-order logic that today forms the most basic knowledge representation system.

**1906 – 1978** Kurt Gödel showed in 1931 that there are limits to what logic can do. His *Incompleteness Theorem* showed that in any formal logic powerful enough to describe the properties of natural numbers, there are true statements whose truth cannot be established by any algorithm.

**1995** Roger Penrose tries to prove the human mind has non-computable capabilities.

# Computation

**1869** William Jevon's Logic Machine could handle Boolean Algebra and Venn Diagrams, and was able to solve logical problems faster than human beings.

**1912 – 1954** Alan Turing tried to characterise exactly which functions are capable of being computed. Unfortunately it is difficult to give the notion of *computation* a formal definition. However, the Church-Turing thesis, which states that a *Turing machine* is capable of computing any computable function, is generally accepted as providing a sufficient definition. Turing also showed that there were some functions which no Turing machine can compute (e.g. Halting Problem).

**1903 – 1957** John von Neumann proposed the *von Neuman architecture* which allows a description of computation that is independent of the particular realisation of the computer.

**~1960s** Two important concepts emerged: *Intractability* (when solution time grows at least exponentially) and *Reduction* (to 'easier' problems).

## Psychology / Cognitive Science

Modern Psychology / Cognitive Psychology / Cognitive Science is the science which studies how the mind operates, how we behave, and how our brains process information.

Language is an important part of human intelligence. Much of the early work on *knowledge representation* was tied to language and informed by research into linguistics.

It is natural for us to try to use our understanding of how human (and other animal) brains lead to intelligent behaviour in our quest to build artificial intelligent systems. Conversely, it makes sense to explore the properties of artificial systems (computer models/simulations) to test our hypotheses concerning human systems.

Many sub-fields of AI are simultaneously building models of how the human system operates, and artificial systems for solving real world problems, and are allowing useful ideas to transfer between them.

## Biology / Neuroscience

Our brains (which give rise to our intelligence) are made up of tens of billions of neurons, each connected to hundreds or thousands of other neurons. Each neuron is a simple processing device (e.g. just firing or not firing depending on the total amount of activity feeding into it). However, large networks of neurons are extremely powerful computational devices that can *learn* how best to operate.

The field of *Connectionism* or *Neural Networks* attempts to build artificial systems based on simplified networks of simplified artificial neurons. The aim is to build powerful AI systems, as well as models of various human abilities.

Neural networks work at a *sub-symbolic* level, whereas much of conscious human reasoning appears to operate at a *symbolic* level.

Artificial neural networks perform well at many simple tasks, and provide good models of many human abilities. However, there are many tasks that they are not so good at, and other approaches seem more promising in those areas.



# Evolution

One advantage humans have over current machines/computers is that they have a long evolutionary history.

Charles Darwin (1809 – 1882) is famous for his work on evolution by *natural selection*. The idea is that fitter individuals will naturally tend to live longer and produce more children, and hence after many generations a population will automatically emerge with good innate properties.

This has resulted in brains that have much structure, or even knowledge, built in at birth. This gives them at the advantage over simple artificial neural network systems that have to *learn* everything. Computers are finally becoming powerful enough that we can simulate evolution and *evolve* good AI systems. We can now even evolve systems (e.g. neural networks) so that they are good at learning.

A related field called *genetic programming* has had some success in evolving programs, rather than programming them by hand.

## Sub-fields of Artificial Intelligence

AI now consists many sub-fields, using a variety of techniques, such as:

Machine Learning – e.g. decision tree learning, version space learning

Neural Networks – e.g. brain modelling, time series prediction, classification

Evolutionary Computation – e.g. genetic algorithms, genetic programming

Vision – e.g. object recognition, image understanding

Robotics – e.g. intelligent control, autonomous exploration

Expert Systems – e.g. decision support systems, teaching systems

Speech Processing– e.g. speech recognition and production

Natural Language Processing – e.g. machine translation

Planning – e.g. scheduling, game playing

Most of these have both engineering and scientific aspects. Many of them you will hear about in this module. Others, that you will not study till later years are:

# Speech Processing

As well as trying to understand human systems, there are also numerous real world applications: speech recognition for dictation systems and voice activated control; speech production for automated announcements and computer interfaces.

How do we get from sound waves to text streams and vice-versa?

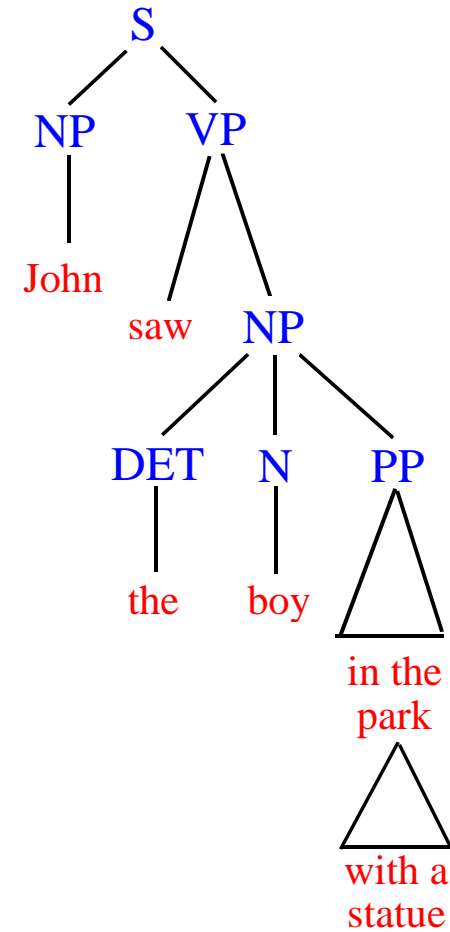
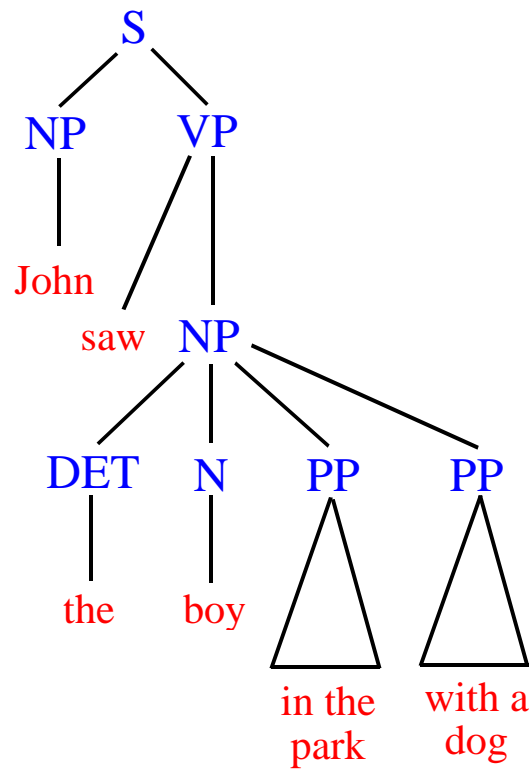
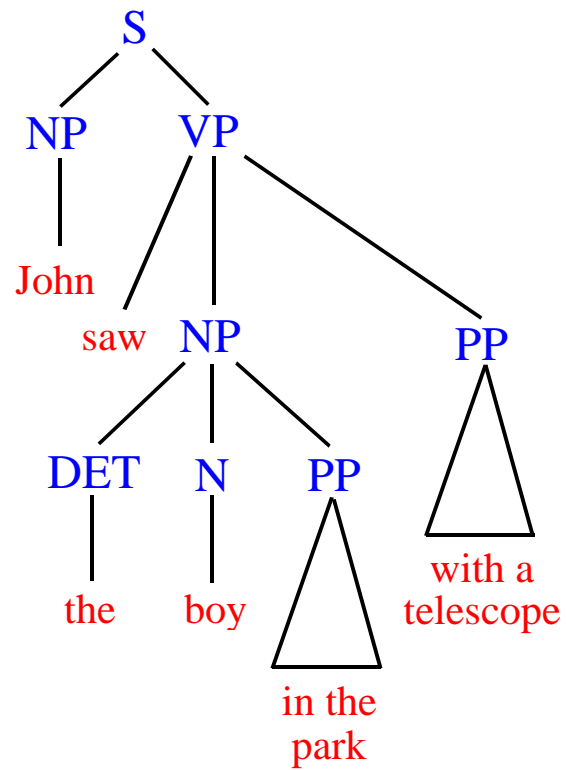


Cen tre fo r Spee ch and Lan gua ge

How should we go about segmenting the stream into words? How can we distinguish between “Recognise speech” and “Wreck a nice beach”?

# Natural Language Processing

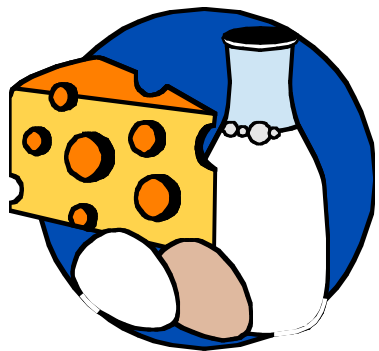
For example, machine understanding and translation of simple sentences:



is not as simple as you might think!

# Planning

*Planning* refers to the process of choosing/computing the correct sequence of steps to solve a given problem.



TO



BUT  
NOT



To do this we need some convenient *representation* of the problem domain. We can define states in some formal language, such as a subset of predicate logic, or a series of *rules*. A *plan* can then be seen as a sequence of operations that transform the initial state into the goal state, i.e. the problem solution. Typically we will use some kind of *search* algorithm to find a good plan.

## Common Techniques

Even apparently radically different AI systems (such as rule based expert systems and neural networks) have many common techniques. Four important ones are:

**Representation** Knowledge needs to be represented somehow – perhaps as a series of if-then rules, as a frame based system, as a semantic network, or in the connection weights of an artificial neural network.

**Learning** Automatically building up knowledge from the environment – such as acquiring the rules for a rule based expert system, or determining the appropriate connection weights in an artificial neural network.

**Rules** These could be explicitly built into an expert system by a knowledge engineer, or implicit in the connection weights learnt by a neural network.

**Search** This can take many forms – perhaps searching for a sequence of states that leads to a problem solution, or searching for a good set of connection weights for a neural network by minimizing a fitness function.

## Covering the Important Ideas

This module will build up a good background in AI as follows:

1. We shall start by looking at intelligence in humans, at how we go about studying human behaviour, and how we try to model/copy their neural processing.
2. Then we'll consider intelligent agents at higher levels of abstraction, and see in principle how we might build artificial intelligent agents.
3. The importance of efficient application dependent knowledge representations is soon clear, and we look in detail at semantic networks, frames, and production systems.
4. We then get an understanding of how the basic search techniques work.
5. Next we study expert systems – how they operate, how we can build knowledge into them, and how they might learn for themselves. Various associated techniques for dealing appropriately with uncertain information will be covered.
6. We end with a consideration of some of the philosophical questions concerning AI.
7. At appropriate points along the way will be guest lectures on Evolution, Robotics, Vision, Natural Language Processing, Computer Chess, and the Philosophy of AI.

## Overview and Reading

1. AI has inter-related scientific and engineering goals.
2. AI has its roots in several older disciplines: Philosophy, Logic, Computation, Cognitive Science/Psychology, Biology/Neuroscience, and Evolution.
3. Major sub-fields of AI now include: Machine Learning, Neural Networks, Evolutionary Computation, Vision, Robotics, Expert Systems, Speech Processing, Natural Language Processing, and Planning.
4. Major common techniques across many of these sub-fields include: Knowledge Representation, Learning, Rule Systems, and Search.

### Reading

1. Russell & Norvig: Sections 1.1, 1.2, 1.3, 1.4, 1.5
2. Nilsson: Sections 1.1, 1.2, 1.3, 1.5
3. Negnevitsky: Sections 1.1, 1.2, 1.3