Artificial Intelligence I

Dr Sean Holden

Computer Laboratory, Room FC06

Telephone extension 63725

Email: sbh11@cl.cam.ac.uk

www.cl.cam.ac.uk/~sbh11/

Introduction: what's AI for?

Homo Sapiens = *"Man the wise"*

What is the purpose of Artificial Intelligence (AI)?

If you're a *philosopher* or a *psychologist* then:

- To *understand intelligence*.
- To understand *ourselves*.

However, we're neither—we're scientists/engineers, so...

Introduction: what's AI for?

From our perspective:

- To understand why our brain is small and (arguably) slow, but incredibly good at some tasks.
- To *construct* intelligent systems.
- To make and sell cool stuff.

This view *seems to be the more successful*.

AI is entering our lives almost without us being aware of it.

Introduction: now is a fantastic time to investigate AI

In many ways this is a young field, having only really got under way in 1956 with the *Dartford Conference*.

`www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html`

- This means we can actually *do* things.
- Also, we know what we're trying to do is *possible*.

Philosophy has addressed similar problems for at least 2000 years.

- *Can* we do AI? *Should* we do AI?
- Is AI *impossible*? (Note: I didn't write *possible* here, for a good reason...)

Arguably, philosophy has had relatively little success.

Aside I: philosophy (428 B.C. to present)

The philosophy of mind has a long history:

- *Socrates* wanted an algorithm (!) for *"piety"* prompting *Plato (428 B.C.)* to consider the rules governing rational thought. This led to the *syllogisms*.
- The possibility of *mechanical reasoning*: Ramon Lull's *concept wheels* (approx. 1315). Followed by various other attempts at mechanical calculators.
- Mind as a *physical system*: Rene Descartes (1596-1650). Is *mind* distinct from *matter*? What is *free will*? *Dualism*: part of our mind—-the *soul* or *spirit*— is set apart from the rest of nature.
- The opposing position of *materialism*: Wilhelm Leibnitz (1646-1716). Attempted to build a machine to perform mental operations but failed as his logic was too weak.

Aside I: philosophy (428 B.C. to present)

There is an intermediate position: mind is *physical* but *unknowable*. If mind is physical where does *knowledge* come from?

- Francis Bacon (1561-1626): *empiricism*. Leading to John Locke (1632-1704): *"Nothing is in the understanding, which was not first in the senses"*.
- In *A Treatise of Human Nature*, David Hume (1711-1776) introduced the concept of *induction*: we obtain rules by repeated exposure.
- This was developed by Bertrand Russel (1872-1970): *observation sentences* are connected to *sensory inputs*, and all knowledge is characterised by logical theories connected to these. *Logical positivism*.
- The *nature* of the connection between theories and sentences: Rudolf Carnap and Carl Hempel's *confirmation theory*.

Aside I: philosophy (428 B.C. to present)

Finally: what is the connection between *knowledge* and *action*? How are actions *justified*?

Aristotle: don't concentrate on the *end* but the *means*.

If to achieve the end you need to achieve something intermediate, consider how to achieve that, and so on.

This approach was implemented in Newell and Simon's 1972 *General Problem Solver (GPS)*.

Further reading, part I

Why do people like to argue that AI is *impossible*?

Why do people dislike the idea that humanity might not be *special*.

An excellent article on why this view is much more problematic than it might seem is:

"Why people think computers can't," Marvin Minsky. AI Magazine, volume 3 number 4, 1982.

Introduction: what's happened since 1956?

What's made the difference? We have a huge advantage in having reached a point where technology has matured sufficiently to allow us to *build things*.

- Perception (vision, speech processing...)
- Logical reasoning (prolog, expert systems, CYC...)
- Playing games (chess, backgammon, go...)
- Diagnosis of illness (in various contexts...)
- Theorem proving (Robbin's conjecture...)
- Literature and music (automated writing and composition...)
- And many more...

The simple ability to *try things out* has led to huge advances in a relatively short time. *So:* don't believe the critics...

Aside II: computer engineering (1940 to present)

To have AI, you need a means of *implementing* the intelligence. Computers are (at present) the only devices in the race. (Although *quantum computation* is looking interesting...)

AI has had a major effect on computer science:

- Time sharing
- Interactive interpreters
- Linked lists
- Storage management
- Some fundamental ideas in object-oriented programming
- and so on...

When AI has a success, the ideas in question tend to *stop being called AI*.

The nature of the pursuit

What is AI? This is not necessarily a straightforward question.

It depends on who you ask...

We can find many definitions and a rough categorisation can be made depending on whether we are interested in:

- The way in which a system *acts* or the way in which it *thinks*.
- Whether we want it to do this in a *human* way or a *rational* way.

Here, the word *rational* has a special meaning: it means *doing the correct thing in given circumstances*.

Acting like a human

What is AI, version one: acting like a human

Alan Turing proposed what is now known as the *Turing Test*.

- A human judge is allowed to interact with an AI program via a terminal.
- This is the *only* method of interaction.
- If the judge can't decide whether the interaction is produced by a machine or another human then the program passes the test.

In the *unrestricted* Turing test the AI program may also have a camera attached, so that objects can be shown to it, and so on.

Further reading, part II

If you've never heard of *Alan Turing* then you really should find out about him, because he provided the *foundations for most of computer science*, did fundamental work in *AI* and was a major figure at *Bletchley Park* during the second World War.

Try:

`www-groups.dcs.st-and.ac.uk/~history/Biographies/Turing.html`

(It's not a tale with a happy ending...)

Acting like a human

The Turing test is informative, and (very!) hard to pass.

- It requires many abilities that seem necessary for AI, such as learning. *BUT*: a human child would probably not pass the test.
- Sometimes an AI system needs human-like acting abilities—for example *expert systems* often have to produce explanations—but *not always*.

See the *Loebner Prize in Artificial Intelligence*:

`www.loebner.net/Prizef/loebner-prize.html`

Thinking like a human

What is AI, version two: thinking like a human

There is always the possibility that a machine *acting* like a human does not actually *think*. The *cognitive modelling* approach to AI has tried to:

- Deduce *how humans think*—for example by *introspection* or *psychological experiments*.
- Copy the process by mimicking it within a program.

An early example of this approach is the *General Problem Solver* produced by Newell and Simon in 1961. They were concerned with whether or not the program reasoned in the same manner that a human did.

Computer Science + Psychology = *Cognitive Science*

Aside III: psychology (1879 to present)

Modern psychology began with the study of the human visual system performed by Hermann von Helmholtz (1821-1894).

The first *experimental psychology* lab was founded by his student Wilhelm Wundt (1832-1920) at the University of Leipzig.

- The lab conducted careful, controlled experiments on human subjects.
- The idea was for the subject to perform some task and *introspect* about their thought processes.

Other labs followed this lead. *BUT:* a strange—and fatal—effect appeared.

For each lab, the introspections of the subjects turned out to conform to the preferred theories of the lab.

Aside III: psychology (1879 to present)

The main response to this effect was *behaviourism*, founded by John Watson (1878-1958) and Edward Lee Thorndike (1874-1949).

- They regarded evidence based on introspection as fundamentally unreliable, so...
- ...they simply rejected all theories based on any form of mental process.
- They considered only *objective* measures of *stimulus* and *response*.

Learnt a LOT of interesting things about rats and pigeons!

Aside III: psychology (1879 to present)

The somewhat more sophisticated view of the brain as an *information processing device*—the view of cognitive psychology—was steamrollered by behaviourism until Kenneth Craik's *The Nature of Explanation* (1943).

The idea that concepts such as reasoning, beliefs, goals *etc* are important is re-stated.

Critically: the system contains a model of the world and of the way its actions affect the world.

Aside III: psychology (1879 to present)

stimuli converted to internal representation

↓

cognitive processes manipulate internal representations

↓

internal representations converted into actions

Thinking rationally: the "laws of thought"

What is AI, version three: thinking rationally

The idea that intelligence reduces to *rational thinking* is a very old one, going at least as far back as Aristotle as we've already seen.

The general field of *logic* made major progress in the 19th and 20th centuries, allowing it to be applied to AI.

- We can *represent* and *reason* about many different things.
- The *logicist* approach to AI.

This is a very appealing idea. *However...*

Thinking rationally: the "laws of thought"

Unfortunately there are obstacles to any naive application of logic. It is hard to:

- Represent *commonsense knowledge*.
- Deal with *uncertainty*.
- Reason without being tripped up by *computational complexity*.

These will be recurring themes in this course, and in AI II.

Logic alone also falls short because:

- Sometimes it's necessary to act when there's *no* logical course of action.
- Sometimes inference is *unnecessary* (reflex actions).

Further reading, part III

The *Fifth Generation Computer System* project has most certainly earned the badge of *"heroic failure"*.

It is an example of how much harder the logicist approach is than you might think:

"Overview of the Fifth Generation Computer Project," Tohru Moto-oka. ACM SIGARCH Computer Architecture News, volume 11, number 3, 1983.

Aside III: mathematics (800 to present)

To be *scientific* about AI three areas of mathematics are needed: computation, logic, and probability.

Logic:

- To the likes of Aristotle, a philosophical rather than mathematical pursuit.
- George Boole (1815-1864) made it into mathematics.
- Gottlob Frege (1848-1925) founded all the essential parts of *first-order logic*.
- Alfred Tarski (1902-1983) founded the theory of reference: what is the relationship between *real* objects and those in logic.

Aside III: mathematics (800 to present)

Computation:

- Concept of an algorithm: Arab mathematician *al-Khowarazmi*.
- What are the limits of algorithms? David Hilbert's (1862-1943) *entscheidungsproblem*.
- Solved by Turing, who (with others) formulated precisely what an algorithm *is*.
- Ultimately, this has lead to the idea of *intractability*.
- Kurt Godel (1906-1978): theorems on completeness and incompleteness.

Aside III: mathematics (800 to present)

Probability:

- Gerolamo Cardano (1501-1576): gambling outcomes.
- Further developed by Fermat, Pascal, Bernoulli, Laplace...
- Bernoulli (1654-1705) in particular proposed probability as a measure of *degree of belief*.
- Bayes (1702-1761) showed how to *update* a degree of belief when *new evidence* is available.
- Probability forms the basis for the modern treatment of *uncertainty*.
- The *decision theory* of Von Neumann and Morgenstern (1944) combines uncertainty with action.

Acting rationally

What is AI, version four: acting rationally

Basing AI on the idea of *acting rationally* means attempting to design systems that act to *achieve their goals* given their *beliefs*.

What might be needed?

- To make *good decisions* in many *different situations* we need to *represent* and *reason* with *knowledge*.
- We need to deal with *natural language*.
- We need to be able to *plan*.
- We need *vision*.
- We need *learning*.

And so on, so all the usual AI bases seem to be covered.

Acting rationally

The idea of *acting rationally* has several advantages:

- The concepts of *action*, *goal* and *belief* can be defined precisely making the field suitable for scientific study.

This is important: if we try to model AI systems on humans, we can't even propose *any* sensible definition of *what a belief or goal is*.

In addition, humans are a system that is still changing and adapted to a very specific environment.

Rational acting does not have these limitation.

Acting rationally

Rational acting also seems to *include* two of the alternative approaches:

- All of the things needed to pass a Turing test seem necessary for rational acting, so this seems preferable to the *acting like a human* approach.
- The logicist approach can clearly form *part* of what's required to act rationally, so this seems preferable to the *thinking rationally* approach alone.

As a result, we will focus on the idea of designing systems that *act rationally*.

Other contributions: linguistics (1957 to present)

B. F. Skinner's *Verbal Behaviour* (1951) set out the approach to *language* developed by the behaviourists.

It was reviewed by Noam Chomsky, author of *Syntactic Structures*:

- He showed that the behaviourists could not explain how we understand or produce sentences that we have *not previously heard*.
- Chomsky's own theory—based on syntactic models traceable as far back as (350 B.C.), did not suffer in this way.
- Chomsky's own theory was also formal, and could be programmed.

Other contributions: linguistics (1957 to present)

This overall problem is considerably harder than was realised in 1957.

It requires knowledge representation, and the fields have informed one another. A classic example:

"Time flies like an arrow"

"Fruit flies like a banana"

Other contributions: economics (1776 to present)

How should I act, perhaps in the presence of adversaries, to obtain something nice in the future?

- Adam Smith: *An Inquiry into the Nature and Causes of the Wealth of Nations* (1776).
- When we say *"something nice,"* how can the *"degree of nice-ness"* be measured?

  This leads to the idea of *utility* as a mathematical concept.

  Developed by Leon Walras (1834-1910), Frank Ramsey (1931) and John Von Neumann and Oskar Morgenstern (1944).

Other contributions: economics (1776 to present)

- For *large* economies:

  Probability theory + utility theory = decision theory

- *Game theory* is more applicable to *small* economies.

  In some games it turns out to be *rational* to act (apparently) randomly.

- Dealing with *future* gains resulting from a sequence of actions: operations research and *Markov decision processes*, the latter due to Richard Bellman (1957).

Unfortunately it is computationally hard to act rationally.

Herbert Simon (1916-2001) won the Nobel Prize for Economics in 1978 for his work demonstrating that *satisficing* is a better way of describing the actual behaviour of humans.

Other contributions: neuroscience (1861 to present)

Nasty bumps on the head

↓

We know that the brain has something to do with consciousness

Experiments by Paul Broca (1824-1880) led to the understanding that localised regions have different tasks.

Around that time the presence of *neurons* was understood *but* there were still major problems.

For example, even now there is no complete understanding of how our brains store a single memory.

More recently: EEG, MRI and the study of single cells.

Other contributions: cybernetics and control theory (1948 to present)

Ktesibios of Alexandria (250 BC)

The first machine to be able to modify its own behaviour was a water clock containing a mechanism for controlling the flow of water.

- James Watt (1736-1819): governor for steam engines
- Cornelius Drebbel (1572-1633): thermostat
- *Control theory* as a mathematical subject: Norbert Wiener (1894-1964) and others.

Other contributions: cybernetics and control theory (1948 to present)

Interesting behaviour caused by a *control system* minimising *error*

$$\text{error} = \text{difference between } \textit{goal} \text{ and } \textit{current situation}$$

More recently, we have seen *stochastic optimal control* dealing with the maximisation over time of an *objective function*.

This is connected directly to AI, but the latter moves away from *linear*, *continuous* scenarios.

What's in this course?

This course introduces some of the fundamental areas that make up AI:

- An outline of the background to the subject.
- An introduction to the idea of an *agent*.
- Solving problems in an intelligent way by *search*.
- Solving problems represented as *constraint satisfaction* problems.
- Playing *games*.
- *Knowledge representation, and reasoning*.
- *Learning* using *neural networks*.
- *Planning*.

What's in this course?

Strictly speaking, AI I covers what is often referred to as *"Good Old-Fashioned AI*.

Historically, the nature of the subject changed a great deal when the importance of *uncertainty* became fully appreciated.

Roughly speaking, AI I covers material up until that point.

AI II covers more recent material.

What's *not* in this course?

- The classical AI programming languages *prolog* and *lisp*.
- A great deal of all the areas on the last slide!
- Perception: *vision*, *hearing* and *speech processing*, *touch* (force sensing, knowing where your limbs are, knowing when something is bad), *taste*, *smell*.
- Natural language processing.
- Acting on and in the world: *robotics* (effectors, locomotion, manipulation), *control engineering*, *mechanical engineering*, *navigation*.
- Areas such as *genetic algorithms/programming*, *swarm intelligence*, *artificial immune systems* and *fuzzy logic*, for reasons that I will expand upon during the lectures.
- *Uncertainty* and much further probabilistic material. (You'll have to wait until next year.)

Text book

The course is based on the relevant parts of:

Artificial Intelligence: A Modern Approach, Second Edition (2003). Stuart Russell and Peter Norvig, Prentice Hall International Editions.

NOTE: the 3rd edition has recently become available. However it seems at present to be both expensive and difficult to obtain in the UK, so I'm still recommending the 2nd edition.

Interesting things on the web

A few interesting web starting points:

The Honda Asimo robot: `world.honda.com/ASIMO`

AI at Nasa Ames: `www.nasa.gov/centers/ames/research/exploringtheuniverse/spiffy.html`

DARPA Grand Challenge: `ai.stanford.edu/~dstavens/aaai06/montemerlo_etal_aaai06.pdf`

2007 DARPA Urban Challenge: `cs.stanford.edu/group/roadrunner`

The Cyc project: `www.cyc.com`

Human-like robots: `www.ai.mit.edu/projects/humanoid-robotics-group`

Sony robots: `support.sony-europe.com/aibo`

NEC “PaPeRo”: `www.nec.co.jp/products/robot/en`

Prerequisites

The prerequisites for the course are: first order logic, some algorithms and data structures, discrete and continuous mathematics, basic computational complexity.

DIRE WARNING:

In the lectures on *machine learning* I will be talking about *neural networks*.

This means you will need to be able to *differentiate* and also handle *vectors and matrices*.

If you've forgotten how to do this *you WILL get lost—I guarantee it!!!*

Prerequisites

Self test:

1. Let

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} a_i x_i^2$$

   where the a_i are constants. Can you compute $\partial f/\partial x_j$ where $1 \leq j \leq n$?

2. Let $f(x_1, \ldots, x_n)$ be a function. Now assume $x_i = g_i(y_1, \ldots, y_m)$ for each x_i and some collection of functions g_i. Assuming all requirements for differentiability and so on are met, can you write down an expression for $\partial f/\partial y_j$ where $1 \leq j \leq m$?

If the answer to either of these questions is "no" then it's time for some revision. (You have about three weeks notice, so I'll assume you know it!)