Introduction to AI

For the man or woman on the Clapham Omnibus

Andrew Lea, MA(Cantab) FBCS FRSA

What is Artificial Intelligence?

When the term was first coined, it meant the endeavour of making computers think as people. As we thought that intelligence was the key characteristic of people, making computers solve "intellectual" tasks would be the key.

Research therefore focused on intellectual tasks such as translating language or playing chess. When computers were not very powerful, it was naturally assumed that to accomplish these tasks would require high levels of abstraction.

As researched progressed we discovered that many of these tasks could be accomplished with a very different type of programming, which (probably) does not represent the way in which humans address these problems. For example, Chess can be very effectively programmed using "adversarial search" (on which more later) which is absolutely *not* how chess masters think. Consequently, as these tasks were accomplished, and shown not to be "intelligent" (in the human sense), AI because almost "that which a computer cannot do".

Interestingly, it turned out that the things we find hard to do, like playing chess, are relatively easy tasks for computers. On the other hand, things we find easy, such as understanding what we see or how to run, are correspondingly hard for computers. (We have large areas of the brain dedicated to these tasks, which as they mostly operate *below* the level of consciousness therefore seem easy. In fact, lots of computation is going on, but beneath the surface on automatic.)

Now that computers are far more powerful, as computing becomes ubiquitous, and as the forces of competition cause companies to want to label their offerings as advanced, AI has come to simultaneously mean several contradictory things. These are:-

1. General AI, Strong AI, Real AI
2. Applied Artificial Intelligence
3. Algorithm Intelligence
4. Fake AI

Let us examine each of these in more detail… [Elements from my own experience are in square brackets.]

1. General, Strong or Real AI

This the most powerful form of AI, and was also what Artificial Intelligence also originally meant. This form of AI actively aims to incorporate the qualities of human thought. It therefore has attributes beyond intelligence, including self-awareness, sentience, ethics, emotions, humour.

In many ways the term "Artificial Intelligence" is inappropriate for this form of AI, partly because if any entity were genuinely intelligent then its intelligence would necessarily be real rather than artificial, and partly because intelligence is only one of the necessary attributes.

I therefore advocate and prefer the term **Machine Thought** or **Machine Sentience**, which would include Artificial Sentience.

2. (Applied) Artificial Intelligence

This is genuine AI, focused just on intelligence, and not the other attributes of thought. This is a collection of techniques, each aimed at specific tasks. As this is what we shall mean later on as "AI", it is worth defining it.

A working definition of AI is the acquisition, learning, manipulation, and exploitation of knowledge by systems…
- whose behaviours may change on the basis of experience;
- which are not constrained to be predictable.

These are important characteristics of AI, and is a distinguishing feature of AI from non-AI software. Misunderstanding these characteristics can lead people to suppose that AI can be regulated in the same way as non-AI software.

3. Algorithm Intelligence

Much software that is described in marketing as "AI" isn't: its simply software - possibly very complex and powerful - that is both marketed as and mistaken for AI. It can have the effect of AI, whilst lacking its fundamental characteristics. In many cases it may be more appropriate than actual AI.

4. "Fake" or "Marketing" AI

Unfortunately AI has become the latest marketing buzz-phrase for software. Consequently, "AI" is used to spice up uninspiring software and dull business ventures.

Concepts in Artificial Intelligence

A key issue in AI is the ***representation*** of the data. The correct representation makes the problem easier to solve. A particularly useful representation often used are ***graphs***: *nodes* joined by *edges*, a bit like dot-to-dot. They can represent many types of meaning (ie semantics), for example nodes might be people and places, and edges might represent links between them.

Another differentiation is where the ***knowledge*** came from. Broadly speaking there are three possibilities: from an expert, in the form of codified knowledge or "knowledge base"; from learning; or in the computer program itself. For some reason only the middle, known as machine learning, is popular.

Of great importance today is ***explainability***: the ability of a system to explain the reason behind its conclusions. It is often a legal concept too: regulations require that some decisions, such as granting or refusing a loan, can be explained. This can rule out some of the popular forms of machine learning.

Artificial Intelligence is implemented on computers and is therefore a class of computer program. Some computer programs are ***algorithms***: they exactly solve a problem; adding two integers, for example. Other programs are ***heuristics***: they yield (hopefully) good guesses to a problem, which cannot be guaranteed to be the best possible answer. The reason for using heuristics is that,

whilst not as exact, the ***running time*** of an algorithm (for example, trying all possibilities) may simply take far too long; and a good guess may be good enough. Sometimes we are in luck, and can ***optimise*** a slow algorithm, making it fast enough.

When looking at data AI often considers the ***features*** of an item. The features of a document might be its words; the features of a face might be the inter-eye distance, eye and hair colour, nose length; the features of a letter might be the strokes of the pen. ***Assertions*** are a similar concept, meaning "things I know to be true", ie facts I can assert.

AI systems which learn often have to be ***trained***.

We often want AI systems to ***generalise***; to learn off specific examples (eg pictures of *specific* cats) to identify the general (to recognise pictures of *any* cat). Sometimes this process overshoots the mark (eg only recognising the example cats as cats) which is called ***over-fitting***.

AI programs are, ultimately, computer programs; and often very conceptually complex too. Recursion is frequently used in AI programming. (This is true even if the AI program was itself written by an AI, rather than a person - more on this later too.) One popular techniques is ***recursion***, in which code is defined in terms of itself. By way of example, factorials (eg 5! = 5 x 4 x 3 x 2 x 1 = 120) can be defined iteratively, as I just did, or recursively: 5! = 5 x (5-1)! Naturally recursion has to be stopped otherwise it goes on for ever, and the computer runs out of memory. So we would write something like n! = (n-1)! if n > 1 otherwise 1

The Artificial Intelligence Landscape

In this section we systematically explore the range of (applied) AI techniques. There are several ways of doing this, so you will probably find other descriptions. The naming of AI techniques can be confusing, because sometimes they are named after the techniques, sometimes after the representation, and sometimes by their application.

Reasoning and Expert Systems

Expert Systems use codified knowledge to reason about a problem. They have the great benefit of being able to explain the reason for their conclusions.

In this area we have:-

Inference Engines. Once very popular, these systems use "forward" or "backward" chaining of reasoning based on knowledge bases. Forward changing means, "given the facts (or assertions) I currently have, what else can I deduce?". Backward chaining means the opposite: hypothesise about a conclusions, and then reason backwards to see if the evidence backs it up.

Case Based Reasoning is all about finding similar, already known, cases to the one in hand. It can be very powerful: for example in finding medical papers which also describe this unusual and exact set of symptoms.

Machine Learning

This is generally what most people unknowingly mean by "AI". Machine Learning itself can be broken into three areas:

Supervised Learning

Supervised Learning systems *learn from examples.*

One frequent task is classification. A **Baysean** email spam filter is told which emails are spam and which are not. It extracts the features of each email (for example, the set of words) and learns which features associate with spam and which do not. There are many types of classifier systems, and another complicated though powerful technique are the so-called **support vector machines**.

Artificial **Neural Nets** (or ANNs) are very popular. They are inspired by the neurones in the human brain and consist of multiple layers of neurones, connected by weights. In a typical three layer neural net there is an input layer, which takes in the features; a hidden layer; and an output layer, which announces the result. They are trained by being presented with examples and the outputs which correspond to those examples; internally they adjust the weights until they produce the correct output for those examples. The hope is that they have then generalised the rules from features to outputs, so that when shown new examples, they give the correct output. In other words, ANNs are approximation functions.

Supervised learning can be used to create or induct **decision trees**, which are closely related to expert systems. They have the great advantage of being able to explain their decisions by tracing the route from root to the conclusion.

Unsupervised Learning

Unsupervised learning does not learn from examples. Instead, it answers the question "what can we learn from this data?"

A frequent form is **clustering**, which answers the question "what clusters do these data points belong to?". Unfortunately many clustering algorithms produce different results according to initial random numbers so they are not deterministic. Worse yet they require, as an input, the number of clusters. [This seems to me the most important characteristic, so I have developed a *deterministic* clustering algorithm which also finds the number of clusters.]

Reinforcement Learning does what it says on the tin. The robot arm, for example, has a poor table of torques to apply to move the arm like this, but through trial and error it is able to improve the table.

Natural Language Processing

Natural Language Processing is all about "understanding" natural languages, like English or French.

Speech recognition - changing sound to text - is actually very hard. It is, however, a well defined task and with many millions of pounds of research and more powerful computers has now been comodotised. Interestingly modern mobile phones are powerful enough to do this task, yet major phone vendors still send the sound back to central servers to be converted to text. One has to wonder why.

Natural language understanding operates on text, which may have originated as text, or as speech that has been converted to text. Again, this is surprisingly hard, and attempts built on school-grammar failed because, surprisingly, traditional grammar is it would seem not how language works! There are several approaches from artificial neural nets through to "parsing" the text using recursive grammar-like techniques.

Machine translation (eg English to French) is a mature use of natural language processing. Originally grammatical attempts were tried, which failed due to the complexity. Now days we

"teach" systems to learn to translate, for example by using a parallel text corpus, such as EuroParl, which have each sentence in multiple languages. [I wrote a translation system which depressingly took only a few minutes to learn each language pair.]

[I often use deep semantic representations (in this case, semantic graphs) in natural language processing. The language is translated into the graph, reasoning occurs on the graph, and answers are generated into natural language. This approach is inherently multi-lingual, and I am using it as a foundation in artificial sentience, making a snap-shot of you with which your descendants, even if they don't speak your language, could interact.]

Image Processing

Image Processing is an application of other AI techniques, together with techniques from the signal processing domain, such as convolutions and deconvolution. Important problems of image processing and computer vision are segmentation (where) are the different parts of each image?) and recognition (and what are those parts?). Frequently neural nets are used for these problems.

Search Space Exploration

Many problems in AI reduce to finding a needle in a haystack in some form. In other words, there are many candidate solutions, most of which are useless, and only a few good solutions.

(In really horrible problems there may be only one solution or, worse yet, none, so no amount of searching can find it yet it can be really hard to prove there are none.)

Route finding is a frequent problem, applicable to many abstract problems such as helping virtual characters in a game decide where to go next. It is also one we can appreciate ourselves: how to find our way from say Victoria Station to 221b Baker Street through the maze of streets (hey, they could be represented as a graph!) and endless possible routes. As ever, there are many route-finding algorithms. One of them starts at the origin, and grows a network of where we could get to towards the destination, and simultaneously starts at the destination and grows a network of we could get back to towards the original. The first bit of these two networks to bump into each other is the shortest route. This is roughly what your satnav does. [Our satnav also appears to have a wicked sense of humour, probably placed there by a programmer having a bad day.]

Adversarial Search is used in programming computers to play chess or Go. In essence, it very quickly thinks though all the moves it could do (trying to "maximise" its own score), considers for each of those all the moves you could reply with (trying to "minimise" its own score), considers for each of those all the moves it could reply with (again trying to "maximise" its own score), … and so on. It can't do this to the end of the game for Chess or Go (but could for 0s-and-Xs) because the universe would end too soon because of the number of possibilities. It therefore stops at some depth and applies a "static evaluation function" to score the position. This Turing-Shannon approach is also called the *Mini-Max algorithm*. Naturally there are ways of speeding it up or optimising it, and the principle one is called the *Alpha-Beta algorithm*. In essence, alpha-beta says once I know I'm not going to play this move, because its worse than another one I already know about, I need not explore it any further.

Optimisation

Mathematically, optimisation is about finding the minimum value of a function. The travelling salesman problem is the classic example: the salesman must visit a set of cities, travelling the least total distance (the quantity to minimise). In which order should he choose to visit those cities? There are many many problems which are, in effect, variants of this problem.

There are many approaches to solving this problem. To do it exhaustively needs all possible routes to be checked (the so-called "**British Museum procedure**") but for any large dataset that simply takes too long as there are n factorial possibilities.

One process is called **hill climbing**. Start with a random order, and try each possible change of the order (eg each possible swap of two cities). Keep the best swap. This is the new order; keep going until no more improvements occur. This will often, normally even, result in a local optimum: an ordering from which all possible steps are worse (so it is in a valley), yet is not in the deepest valley (the optimum). Depending on the problem, and luck, this may of course be good enough. One optimisation is simply to try from several different starting orderings, and keep the very best.

Two highly effective heuristics are both modelled on nature.

Simulated annealing models the way in which ice forms. It seems to start in different places, yet still join up into one sheet of ice. It works a bit like hill climbing, but has the concept of temperature. It starts hot and gradually cools down; whilst hot, it will sometimes go in the "wrong" direction. This approach seems to help it avoid local minimums.

Genetic Optimisation models each possible solution as a gene. In each generation they compete. The very best survive and mate to produce the next generation, crossing over their genes; the least fit die and are lost. In this way, the best solution "evolves". (It is surprisingly hard to make genetic optimisation and indeed any evolutionary system stable: the system must be very carefully designed and tuned.)

Predictive Analytics

Predictive Analytics is arguably not AI, but is often associated with it. It is about "learning" from previous examples in order to "predict" the future. It can be as simple as a linear regression model of the form y = ax + b; it climbs through statistics; and eventually ascends into the heights of applied artificial intelligence.

Issues in Artificial Intelligence

AI and related issues present many issues worthy of discussion. The following are my views or questions, and may not be generally accepted.

Humans – working with AI

- Applied AI delivers benefit when we ask it to do some, but not all, of a task. It can help drive cars more safely, or answer 80% of emails. Applied AI with a human in the loop is a force multiplier.
- I have measured AI to have a better mean accuracy than people, who excel at exceptions and subtle judgement, so the best design uses both. AI can effectively identify *when* to pull in people.
- AI, security and data protection cannot be retrofitted into a system because they are foundational.

Machine Learning Techniques – a subset of AI

- Deep learning is deep in so far as it has many layers; not in the sense of uncovering deep insights.
- But decision trees are explainable, and the inducted knowledge can be validated by experts.

- "Big data" is small compared to the number of possible Go or Chess positions an AI must consider.

Benefits of AI

- My calculations show AI could increase the effectiveness and reduce the costs of social services.
- AI could drive efficiency and effectiveness in many industries, such as logistics and aviation.
- By helping reduce energy consumption, AI might help reduce CO2 emissions and climate change.
- "Emergency" AI could make aviation safer, but replacing pilots with AI might reduce safety.

Employment

- Intuition, as well as reasoning can be modelled, so "intuition" based jobs may be vulnerable to AI techniques. More positively, perhaps "artificial intuition" could enhance human creativity.
- AI should reduce everyone's workload. The challenge (for government) is how to share this benefit, rather than have it enrich a few or be squandered on an ever-growing administrivia.
- AI may well create employment by opening up new fields of endeavour.

Reliability and Trust in AI

- Explainability must be a strong requirement for AI in the NHS or financial services.
- Suppose in a critical application the explainable algorithm is less accurate than the black box algorithm: which should we prefer? With recent advances, this dilemma is a very real possibility.
- AI is of its nature not predictable: it may learn, and is non-deterministic. "AI" for safety-critical applications such as aviation should be AI-inspired, yet deterministic predictable and verifiable.

AI Ethics

- AI ethics often parallels ethical issues in other technology or even human ethics
- Ought autonomous cars prefer the safety of its voluntary passenger, or involuntary pedestrians?
- AI ethical judgement would be learnt, acquiring any dataset bias, or inherit programmer's values.
- We might not deploy AI due to non-technical issues: whose fault is it if a self-driving car crashes?

Philosophical Questions

- If the best AIs need emotions as motivations, evolved programs might evolve emotions. Very intelligent AIs may have their own "artificial" and alien, desires.
- "Real" AI may arise as an uncontrolled emergent property of the Internet of Things.