Perspectives on Artificial Intelligence: Three Ways to be Smart

Inman Harvey
Evolutionary and Adaptive Systems Group, University of Sussex
inmanh@gmail.com

Abstract
Three different styles of achieving Artificial Intelligence are discussed and compared. The earliest, and best-known to the general public, is the computational approach to AI that takes the brain to be some form of computer. This focuses on a narrow form of intelligence, abstract reasoning. By contrast, Artificial Neural Networks sees the brain as a brain rather than computer. This focuses on categorization and pattern recognition. The newest perspective is Evolutionary Robotics. This takes the broadest view of intelligence and cognition, seeing it as adaptive behaviour in a physical world; cognition and the mind are not centred in the brain at all. Rather than three different methods of achieving similar goals, these perspectives are aimed in very different directions. The implications of this for developing artificially intelligent SmartData agents are discussed.

1. Introduction
The general public largely knows about AI, Artificial Intelligence, through Hollywood and the media. There are cultural biases – in the West, robots are often seen to be Terminator-like and threatening, whereas in Japan the norm is to see robots as friendly, as with the manga/tv character Astro Boy. Whichever way that cultural bias goes, unsurprisingly this has tended to exaggerate expectations on the technical side of what is achievable. AI researchers often get a raw deal; on the one hand they get denigrated for failing to live up to such expectations, but on the other hand many of their manifest successes have become so mainstream and familiar that people forget that they are really AI successes. As the late John McCarthy, a founding father of AI, said: "As soon as it works, no one calls it AI any more". Computer programming, cellular phone technology, online banking, Google, speech recognition – all these demonstrate AI that works.

For professionals working in AI there are several possible motivations. One set of motivations is technical and technological, to build equipment that helps us to do intelligent things. These may be best viewed as tools to extend and amplify human intelligence, as cars and planes extend our means of travel. Another set of motivations is scientific, to build machines based on theories of how our intellects or our brains work, to test just how far those scientific theories make sense. Since understanding just how we ourselves work – when it is we ourselves who have to do the understanding – is the most challenging of scientific puzzles, it is not surprising that there are strong disagreements on what approach to take. Here I shall discuss three competing approaches to AI, three perspectives on what the scientific challenge consists in. In comparing GOFAI (Good Old Fashioned AI, the computational variety), ANNs (Artificial Neural Networks) and ER (Evolutionary Robotics) it will be seen that the differences often lie not so much in different approaches to the same problem, but different perspectives on what the problem itself actually is. To a large extent they are tackling different projects; they count as different paradigms in the Kuhnian sense (Kuhn 1962).

When it comes to applying advances in these three different strands of AI to tackling some technical problem, the three different styles will have their advantages and disadvantages according to context. Exploiting SmartData so as to allow people to delegate some of their privacy concerns requires some AI in the form of sensitive contextual judgment. Here I shall discuss what possibilities lie within each of these three strands of AI. For the most part this is a general overview of AI, so the references given are mostly old and foundational works.

2. Classical GOFAI
Probably the most common understanding of AI is the widespread view that computers are the best examples of intelligent machines and that the brain is some form of computer (Feigenbaum 1986). In actual fact, many in the business – including the present author – consider these to be deeply misleading statements: my own view is that computers have one crucially important but extremely limited role in intelligence and that our brains have virtually nothing whatsoever to do with computers.

Intelligence evolved in the biological world to give different organisms more appropriate behaviour to survive, eat, escape predators, and provide for the next generation. Indeed on one view, in the broadest sense intelligence is
just the convenient label to apply to the property of having such adaptive behaviours. Different organisms, living in different environmental niches will have different behavioural requirements, and different types of intelligence. From our anthropomorphic perspective we naturally focus upon our own human world. What differentiates us most from other creatures is our sophisticated use of language; indeed, we live in a sea of language as a fish lives in water. Humans are also tool users, and the precursor of so much of our intelligent abilities was the invention of tools used with language. Whoever it was, perhaps a Sumerian first writing on a clay tablet some five thousand years ago, can also be considered the great-grandfather of computing; for computing is the most recent update on linguistic tools for manipulating data. The printing press made copying data cheap, and the electronic revolution has made copying virtually free.

Spoken language is transient; to maintain verbal histories requires persistent teaching of stories from one generation to the next. But to make a semi-permanent record, on clay or stone or papyrus, allows people to store and transmit information through distance and through time. This made government, accounting, urbanisation, education and science possible. The tools of writing did not merely record spoken language, but extended what we could do with it. We would not be able to multiply 57 by 239 without pen and paper, or an abacus or some such tool. Here we are following a sequence of instructions, and storing intermediate pieces of information that are going to be re-used later down the line. Such step-by-step procedures are named algorithms, in honour of the 9th century Persian mathematician al-Khwarizmi.

In the 19th and early 20th century computers were employed in their thousands to carry out such algorithms – for at that time the term 'computer' meant somebody like a clerk employed in an accountancy or insurance office to perform by hand immense and tedious calculations by following detailed set procedures. Following Turing's insights in the 20th century, we now use the term computer for those machines that can mechanically carry out such algorithms; at least computer scientists do. There is a clear definition, illustrated in about its simplest form by a Turing machine, that provably has immense power (a specific Universal Turing machine may be made to emulate any other arbitrary Turing machine) and provably has specific limitations (the Turing Halting problem, the undecideability of many questions that can be set). Throughout this paper I am using the term 'computer' (after the brief nod above to its origins as meaning a clerk) in this technical sense of a particular kind of machine. Hence I want to maintain the clear distinction between computing machines and other sorts of machines. A grandfather clock, keeping time via a swinging pendulum and powered by a falling weight, is most certainly not a computer in this sense, since it is not following any algorithm or prescribed set of instructions.

Hence when I assert that a brain is not a computer, I take the position that the brain can indeed be thought of as a machine – but just not a computing machine, the algorithm-following sort of machine in the category of a Turing machine. A brain is a crucially important part of a human being, and it is largely in virtue of having a brain that we humans can (usually rather slowly and inadequately) carry out computations; we can indeed perform long multiplication with pencil and paper, provided we can remember the instructions we learnt at school. But it is then us humans carrying out the algorithm, following the list of instructions, and not our brains doing so.

Computers, the modern electronic versions, are superb tools that extend our human intelligence into regions that were previously unthinkable; they can carry out pre-specified tasks of great complexity at high speed. From a technical perspective, a major advantage of computing approaches to AI is that they encourage modular approaches to complex problems, and this allows extensions and advances to be easily bolted onto or built on top of existing technologies. Indeed, GOFAI is itself built on top of previous developments in tool development stemming from technological advances, from the invention of writing through printing presses, clockwork, steam engines and electronics.

Computer programmers are taught to break problems down into smaller self-contained modules that can then be treated as building blocks that communicate between each other. Snippets of code are used multiple times within one program, are copied into different programs and are reused across different platforms. This means that computer science is largely an incremental discipline in which the discoveries and advances of previous decades are used as the basis for building further advances. This gives strength in depth to GOFAI computational approaches, more so than the other perspectives we shall discuss later. GOFAI and computing has been the technological advance par excellence in the 20th century; indeed the contrasting AI approaches below have only been made feasible through the use and availability of computing advances. Moore’s Law describes how technical advances have reliably scaled computing performance further ahead each year; hardware and software developments have helped each other symbiotically.
So what are the negative aspects? It looks like GOFAI is tailored towards only a small subset of what humans can do, in focusing purely on well-defined problems where all the information is in principle available and the need is to crunch systematically through all the possibilities. The game of chess typifies this kind of problem, which is why the success of Deep Blue against world chess champion Kasparov in 1997 was such a milestone. Its ability to search through all the possible future moves up to 6 or 8 moves ahead at lightning speed ultimately defeated the human.

Any position in a chess game can be unambiguously described, but the real world of everyday human existence is open-ended. All relevant context to a problem has to be incorporated into a computational approach, and unlike chess there is no clear way to decide in advance what might be relevant. How do we frame the problem? – this is known as the Frame Problem and has proven the nemesis for many attempts to build robots that can negotiate the real world. The GOFAI approach works with static problems of the type “given this static snapshot of a chess game, what is the next best move” and does not have a natural fit to realtime dynamic issues, such as crossing a road or playing basketball. The standard GOFAI tactic is to reduce dynamics to a rapid sequence of static problems. Biological creatures, including humans, live in a dynamic world and have been crafted by natural evolution to behave adaptively in it. The GOFAI approach ignores such evolutionary issues that have shaped our human intelligence.

Though chess computers can indeed beat a chess grandmaster through a GOFAI approach, interestingly it looks like human grandmasters mostly use a very different strategy, the ability to recognize at a glance significant patterns in the arrangement of pieces on the board and to relate these to games they have played or watched before. This ability turns out to be rather difficult to replicate with a GOFAI approach and AI has largely turned to the next paradigm in our list for this.

3. Artificial Neural Networks

We are good at recognizing patterns, and our brains are instrumental in enabling this. So neuroscience has been influential in providing inspirations for techniques for doing machine vision. Statisticians are also experienced at recognizing patterns in data, and sometimes ANNs can be seen as a method of ‘doing statistics’ in a brainlike fashion (Rumelhart et al. 1986; McClelland et al. 1986; Arbib 1995). In vision, the receptive rods and cones on our retina can be replaced by an array of pixels sampled in a CCD camera. This can be considered as a two-dimensional array of neuron-like units, each of which passes along multiple weighted copies of its signal strength to each of the units in a second layer. After further minimal processing of summed activations at each of the units in this second layer, a similar process passes on activations to a further layer.

This setup, combined with some simple rules for adjusting the weights between units during a training process, can result in a relatively simple ANN with the potential for the third, or output, layer to perform some categorization of the images that were presented to the retina or first layer. More sophisticated versions may have different numbers of layers, may have the neuron connections going both backwards and forwards between layers, and may have more sophisticated training regimes. What they will have in common is the use of several arrays of relatively simple processing units, passing signals between each other in a highly parallel fashion, with relatively simple local processing done at each unit.

This is clearly a very different style of machine from a basic computer. When we compare the artificial intelligence of a typical such ANN with the artificial intelligence of a GOFAI computer, and relate this to human intelligence, we can see that they are doing rather different jobs. Broadly speaking, the neural network is doing the sort of job that a chess player does when she recognises which piece on the board is a king or a pawn, and what sort of pattern of pieces is spread across the board. Any logical or abstract chess reasoning by the chess player, as to the consequences of further moves, broadly corresponds to a rational, computational next stage.

These two jobs do, however, have one thing in common. They are both sequential rather than real-time dynamical jobs. If there is a flow of images forming the input stream to an artificial neural network doing machine vision, then typically this is broken down into a sequence of snapshots. Each static snapshot is then categorised. Similarly with the subsequent computational task: an abstract description of the pattern of pieces on the chessboard provides a static context for the reasoning as to what is the choice of next move.

The advantages of ANNs for the jobs on which they are typically used are manifold. This is now a mainstream optimisation method for many types of pattern recognition, including speech recognition as well as computer
vision. As with many statistical methods, they can be very appropriate for handling noisy inputs; the
categorisation tasks they are doing require them to discern the signal through the noise. Over the last 25 years or
so, the body of knowledge about neural networks has built up to the extent that it is becoming more feasible to
assemble together different styles of simple neural networks to make a more complex whole. This potential
modularity, though, is still very minor in extent compared the major advantages modularity brings to conventional
computer methods. Disadvantages of ANNs include the fact that training them to find the appropriate weights on
the connections between neurons typically takes a very long time.

ANNs typically do not handle temporal issues naturally – in effect they often cheat by looking at successive
snapshots; however the real neural networks in our brains do allow us to behave adaptively in a dynamic world.
Our third AI perspective takes account of this.

4. Evolutionary robotics

This approach to AI counts as the new kid on the block compared to conventional computational approaches and
to neural networks (Harvey et al. 1997, 2005; Nolfi and Floreano 2000). However it is not competing directly
against them since it is typically aimed at doing an entirely different job. Instead of confining itself to the
reasoning part of human intelligence, or to the categorisation of sensory inputs, ER is normally applied to
designing the complete perception-action loop of a complete (artificial) organism or robot in a dynamic
environment.

Darwinian evolutionary theory leads us to believe that the brains and bodies of humans, animals, indeed all
organisms, were designed through evolution to produce adaptive behaviour so that they could survive and produce
offspring within their various environmental niches. In ER we try to do something comparable using artificial
evolution. Just as natural evolution started with the simplest possible organisms and used these as a basis for
evolving further complexity, so we start with the evolution of the simplest possible artificial organisms. Darwinian
evolution is clearly one inspiration for this; another inspiration is the field of Cybernetics, which is the study of
control and regulation in animal and machine.

A typical experiment might start with deciding on some physical environment containing goals and obstacles that
we want an evolved robot to navigate through successfully. The criteria for success must be clearly measurable.
This might involve accumulating “energy” or “food” that is available intermittently at different places and counts
as reward; the excess energy spent through bad navigation or collision through obstacles counts as a penalty.
Rather simplistically, these calculations of fitness will be translated through into the probability that this particular
robot has offspring in the next generation, compared to other robots with greater or lesser fitness.

Each robot within the population has a genotype, or string of artificial DNA, that systematically determines the
configurations of its brain or control system; sometimes it may also determine the robot sensors and motors and
other bodily characteristics, sometimes these may remain fixed across the whole population and so not need
further genetic specification. The philosophy of ER is, subject to these constraints, for the human to play no
further role in designing the robot. So typically in the first generation the genotypes are specified randomly and
hence produce a random bunch of robot designs. Despite this unpromising start, each such random robot can be
evaluated for just how fit is its behaviour (or lack of behaviour) in the environment. In this and subsequent
generations, the fitter members of the population are given a greater representation than the less fit, through
becoming parents and thereby passing on more of their genes (their artificial DNA) to the subsequent generation.

The assessment of fitness does not depend on a breakdown of the cognitive tasks of the robot into visual
recognition, followed by abstract reasoning. The fitness depends upon the global dynamic behaviour of the robot
over its allotted lifespan. If the experiment has been set up appropriately, the mindless operation of artificial
evolution – The Blind Watchmaker – tends to increase the fitness over successive generations, from the initial
random behaviour to something more appropriate.

The types of artificial neural networks being evolved as “brains” or control systems for these robots are typically a
different style from the (timeless) information-filters of the earlier ANN paradigm. They are more in the
cybernetic tradition of being in effect real-time dynamical systems. These generate real-time dynamical behaviour
continuously, in contrast to the sequence of static snapshots that the GOFAI approach uses. It can be argued that
therefore this ER approach, though clearly simplistic, is closer in spirit to the real-time dynamics of real
organisms than the other two AI paradigms discussed above.
This goes some way towards explaining why ER is often an AI approach of choice when it comes to doing very basic, in principle, scientific studies of the basics of cognition. Fundamental questions can be tackled, some of them closer to the philosophy of mind than to conventional robotics: e.g. “how can we understand the origins of communication between agents, real or artificial?” It fits into the tradition of embodied cognition: we are not directly concerned with the manipulation of abstract concepts, but rather with the behaviour of a real physical robot in a real physical world. We may sometimes be using computer simulations, but these are explicitly simulations that model some real physics. Since the real world contains noise and unpredictability, noise is typically added to the simulated physics.

The advantages of ER compared to the other paradigms largely relate to this natural fit with the notion of embodied cognition and an enactive approach to cognition (Varela et al. 1991), together with its clear appeal to the use of an artificial evolutionary technique inspired by the natural evolution that produced biological organisms. One immediate consequence of this is that it minimizes the degree to which human design prejudices might bias control systems or artificial brains to reflect current scientific fads or fashions. For the fundamental questions about cognition to which ER is often addressed, this lack of bias is a clear advantage. A disadvantage shared with a more conventional use of ANNs is that just as the latter often requires excessive amounts of training time, the former often requires excessive amounts of evolutionary time. A further disadvantage is that often an evolved brain is opaque to scientific understanding. We may know that it works reliably, even with noisy conditions, but still not understand just what are the core features that make it work.

Although natural evolution clearly evolved complex human brains through incrementally adding bits and modifying simpler ones, it has had nearly 4bn years to evolve humans. With artificial evolution we are at the equivalent of being right back near the origin of life. We don’t have the easily assemble building blocks that computational programs have, and this slows down progress.

5. Comparing the 3 perspectives

We have already listed some of the individual advantages and disadvantages. But overwhelmingly the comparison is not between three different methods for achieving similar objectives, since they are each aimed at very different ends. The computational approach is appropriate for logical reasoning in very clearly defined domains; if you wanted to build a reliable pocket calculator, use GOFAI. ANNs are appropriate for categorization and pattern recognition; if you want to recognize faces from photos, then this is the method of choice. But if you want a real-time control system for a robot in a dynamic environment, that respects the principles of embodied cognition that many believe characterize real animals, then ER may well be the method of choice.

All three methods count as different kinds of AI, but clearly they lay different emphases on just what counts as intelligence. The GOFAI approach focuses on a narrow definition of intelligence as abstract reasoning, ANNs focus on the ability to recognize patterns, and how this skill can be trained. ER is based on the broadest possible interpretation as adaptive behaviour. They have different issues with the need to contextualise problems. In everyday life, some of our problems, typically the abstract ones, are so unambiguously described that they are context-free. Multiplying 2 by 6 has the same answer whether we dealing with cows or kilograms, and this is tailor-made for GOFAI methods. ANNs, when used for instance for machine vision, necessarily have to cope with some degree of noise and contextual variation; if doing face-recognition, the face should be recognized in a wide variety of backgrounds, of orientations and of lighting conditions. In some sense the job is to separate the categorization from the noisy context. The ER approach typically goes even further in this direction, since here context is everything; the job is to match appropriate behaviours to context.

6. Three approaches to SmartData

So far we have discussed varieties of intelligence, and the different approaches to the artificial versions thereof, to AI, in very general terms. If we focus on applications of SmartData to the personal control of information, to privacy concerns and the protection of data, we can see how these different approaches might assist.

The GOFAI approach already plays a central role in protecting information. Information here typically means data in electronic bits, and the most common everyday form of protection is encryption providing a (nearly) impenetrable safe box, with passwords providing the key. The intelligence, in deciding when to use the key
oneself or whether to give somebody else access, rests entirely with the human user. The data is dumb but the human, we hope, is smart. This lays burdens on the human.

Your average person has many such boxes of data, for personal documents on their own computer, for financial data on bank and supplier systems, perhaps health and official data on medical and governmental systems. One problem is the desirability to use different strong passwords on different systems, set against the practical infeasibility of committing all these to memory. There are password management tools that help users organize their passwords in such a way that they only have to remember one master password; perhaps this is also combined with a further authentication device such as biometric recognition, or some physical dongle.

Once data has been released to a third party, this normally means further control of that data – and future distribution of copies – has been relinquished to that third party. There are information rights management tools (e.g. Zafesoft) that can restrict those rights to further use the data, by embedding the encrypted data within custom software that the third party can use only so long as the original data owner permits. This is only partial control, however – as soon as the data is made visible and usable by a third party, they can find ways to extract it from that control.

Given all these circumscribing issues with their partial solutions, there remains an overriding human element of making the decision of when to open some box, and just how much of the contents to release under what terms to somebody else. When these are rare and important decisions, the data owner can give their full attention to them. Unfortunately such decisions are needed on a daily, sometimes hourly, basis, and we may not be able to give our full and undivided attention to assessing the context in which they occur. In an ideal world, we might have a 24-hour personal and confidential assistant to whom we trusted the delegation of such context-sensitive decisions; can AI somehow add elements of this to a SmartData system?

We have seen that the computational approach is not well-matched to issues of context sensitivity, so let us now assess how ANNs and ER might be able to assist.

6.1 ANNs and Context Sensitivity

The decision to release some data to a third person can happen in the street or store, or increasingly frequently online. When dealing with people, we assess their manner, the surroundings they are associated with, what previous interactions we have had with them; “is this a reliable looking store or restaurant, should I give them my credit card details?” When dealing online we have less to go on, just the text interactions. In fact it is very much like the Turing Test, where we are engaging in online conversation with an entity whose credentials we have to judge – and where the potential bad guys are very likely trying hard to persuade us that they are good. This is a form of categorization, where we may expect the pattern recognition capacities of ANNs and similar methods to be of potential use.

A computational approach to categorization might be in effect box-ticking, a list of clear criteria that need to be passed. But ANN approaches that are resilient to noise, that accumulate weights of evidence for and against possible categories over a multitude of hints and clues, are much more promising. ANNs can indeed be analysed in terms of acting as Bayesian statistical machines accumulating evidence in just such a fashion, and coming out with some degree of confidence in a conclusion. Anyone who uses a spam filter for their email, for instance the one that Gmail uses automatically, will be benefitting from just these techniques. There are all sorts of clues, in the headers and in the text of an email, that tend to make it more or less plausible as genuine. As humans, collectively, we are fairly good, though not perfect, at spotting these ourselves. The human labeling of potential spam forms the training set for what is in effect an ANN, a Bayesian filter, to categorise as spam or legitimate (Sahami et al. 1998).

There are some grey areas where particular emails might be rejected as spam by some, yet others might still want to receive these. On top of the collectively trained consensus there is scope for fine-tuning by individuals according to their own tastes. Both the global training and the personal fine-tuning are driven by human overview that corrects for where the current filter makes a wrong choice. Spam filtering is generally a context-determined binary yes/no decision, but the same techniques can be extended to finer multiple discriminations, for instance to different levels of privacy. In so far as privacy concerns online, based on text interactions, mirror the essentials of
spam filtering in email, then such techniques of Bayesian filtering in the spirit of ANNs stand out as a principled method with a body of technique already developed.

6.2 ER and Context Sensitivity

As discussed above, ER has been aimed at generating adaptive real time behaviour in a dynamic physical context. As far as I am aware there has been no ER work to date where the environment that gives context to action has been textual. That is not the sort of goal that ER is currently aimed at, since it works in a real physical environment of action, or a simulation that incorporates the essentials of the physics.

Some other aspects of security and privacy issues might come closer. Recognising people from their gait when they walk would be one form of real time categorization; similarly recognizing signatures from the temporal pattern in which they are written, and perhaps identifying people from their speech-patterns through time. All these have real time dynamics at their core.

At a more philosophical level, ER experiments can provide a cognitive science forum in which to investigate just how problems, that may be posed to animals, humans or artificial agents, relate to the contexts in which they are posed. It looks to me that the invention of writing, perhaps on clay tablets in Sumer some 5000 years ago, marked a major shift towards storing information in a relatively context-free manner. Ultimately we may hope that ER can tackle such an important transition in artificial agents, but it is too early at the moment.

7. Summary

Different perspectives on what natural human and animal intelligence is, on where the core problems of cognition lie, have translated into different approaches to the production of artificial intelligence. The computational approach, here called GOFAI, has focused on rational thought on problems abstracted from their context. Brain-inspired approaches such as ANNs have focused on just how that abstraction from context can take place, in pattern recognition. An ER approach is not so much concerned with abstraction at all, being more concerned with behaviour in a dynamic physical world.

When it comes to finding AI approaches that might assist in the development of SmartData, in particular for automating the recognition of context so as to decide what levels of privacy to observe, this appears to have a closest fit to the concerns met by ANNs. For making decisions based on text there are existing techniques for spam detection and filtering in email that can be adapted. Bayesian filters for doing this may not be explicitly presented as ANNs, but they belong very much to the same paradigm.

References


