

# The Myth of AI Failure

CSRP 568

Blay Whitby  
Department of Informatics  
University of Sussex  
Brighton BN1 9QH, UK

December 2003

## Abstract

A myth has developed that AI has failed as a research program. Most myths contain some germ of truth but this one is exceptional in that it is more or less completely false. In fact AI is a remarkably successful research program which has delivered not only scientific insight but a great deal of useful technology.

One of the main reasons why people assert the failure of AI as a research program is the mistaken view that its main goal is to replicate human intelligence. Such a view is understandable. It is common to misread Turing's 1950 paper *Computing Machinery and Intelligence* as suggesting that the ultimate goal of AI should be the complete replication of human intelligence. AI researchers have also not done sufficient to make it clear that complete replication of human intelligence is not the ultimate goal of AI.

A further source of the failure myth is the need felt by many researchers to distance their approach to AI from other, usually previous, approaches. In many cases a fashionable approach to AI may give itself a new name - ALife would be a good example – and portray previous AI approaches as having failed.

In truth there is no failure to be explained. Almost every citizen in the developed world makes use of AI-derived technology every day. The fact that this AI technology is usually hidden in other technologies and works unobtrusively is a measure of just how successful AI has been. AI has also inspired, and continues to inspire, many other disciplines from linguistics to biology through the generation of scientifically useful data and concepts. The scientific work may still be at an early stage but its potential is great and failure myths should not be allowed to impede it.

## Introduction

This paper seeks to challenge a widespread myth that Artificial Intelligence (AI) has failed as research programme. This myth, it is claimed, has no basis in truth. AI is a remarkably successful research programme. As a science, AI has yielded many insights not only in the area of intelligent machines but also in many other branches of science. Considered as a technological enterprise, AI is even more successful and has contributed a surprising amount of useful technology.

It is important to challenge myths because they can greatly influence the popular image of a science. Even a false myth such as this can have an effect out of all proportion to its apparent significance on the way AI is seen outside the field.

The myth that AI has failed has two different groups of progenitors. The first group is composed primarily of those philosophers who see the entire enterprise as misguided, impossible, absurd, or as some combination of these. The second group is made up of those AI scientists who wish to distinguish their particular approach to AI from other approaches and feel that they will help reinforce this distinction by claiming that approaches other than theirs have failed.

Over the approximately fifty years that the expression Artificial Intelligence has been in use there have been numerous examples of members of both groups making the sorts of attack on AI research that might well lead to the creation of the myth that AI has failed and it would be tedious to enumerate them all.

Also it cannot be denied that over-enthusiastic claims by leading AI figures over the decades have very often created expectations that could not be met. In particular, some wildly optimistic estimates of how long certain developments would take to achieve have contributed to the failure myth. This is reprehensible and unfortunate. However, since I have observed this at length (for example in Whitby 1996b), I shall only nod towards it here and concentrate instead on the two groups of progenitors mentioned above.

A suitable example representative of the first group – the philosophers who see the whole project as misguided – is Hubert Dreyfus, particularly the claims made in Dreyfus 1998. An exemplar of the second group is Rodney Brooks – in particular the Brooks of *Intelligence without representation* (Brooks 1991). It is important to remember that what is under discussion here is *myth* – not explicit claims by Brooks and Dreyfus. It is also the case that Dreyfus is not the only philosopher who has given the impression that AI scientists are failing in their objectives. Similarly, Brooks is far from the only AI scientist who has felt the need to disparage other approaches to AI.

Hubert Dreyfus has criticised AI in most of his publications. The most explicit criticism closest to the claim that AI is a failed research programme is to be found in notes from a lecture he gave at the University of Houston on 27<sup>th</sup> January 1998 *Why Symbolic AI Failed: The Commonsense Knowledge Problem* (Dreyfus 1998). Dreyfus starts this lecture with the mistaken claims that in 1950 Turing predicted machinery indistinguishable from humans and that symbolic AI had adopted this goal. This is a misreading of both Turing's 1950 paper and of the goals of symbolic AI. Subsequent sections of this paper will attempt to set out just why these are misreadings. However, given these misreadings of what AI is all about, it is relatively easy for Dreyfus to spend the rest of the lecture arguing that it has clearly failed.

There is no need at this point to pick up on the 'symbolic' in the title of Dreyfus lecture. Subsequent sections will argue that singling out any particular approach to AI for the 'failure' tag is yet another mistake in understanding the overall nature of AI.

Brooks' contribution to the myth is even more indirect. In *Intelligence without Representation* (Brooks 1991) he clearly states AI had the initial goal of replicating human-level intelligence in a machine but that he and others believe that this level of intelligence is 'too complex and too little understood' to be attacked by the method of decomposing it into manageable sub-problems. This too, is the most obvious reading of the parable of the 747 that Brooks weaves into this paper.

Brooks' agenda in this paper is to replace the goal of the achievement of human-level intelligence through the route of dividing it up into manageable sub-problems with the goal of the incremental development of simple autonomous systems. This seems both a viable approach to AI and one which is fully compatible with the previous sub-division approaches. Nevertheless Brooks' tone in this paper makes an overly strong case against previous approaches to AI. He does not explicitly state that they have failed but presents a variety of reasons for believing that they cannot succeed.

## Science versus Technology

It is useful at this stage to introduce the distinction between AI as a scientific enterprise and AI as a technology. This is not an absolute distinction. It is not always possible to say where science ends and technology begins – particularly in the case of AI. However, there is an important philosophical principle captured in the quip that 'the existence of twilight does not mean that there is no difference between night and day'. AI as a science clearly has very different goals from AI as a technological enterprise.

The myth that AI has failed concerns both AI as science and AI as technology. In caricature we might say that the myth is that : "AI has failed to replicate human intelligence and therefore it has failed to give a scientific account of intelligence in general. Because it has failed to replicate human intelligence it has also failed in its technological goal which was to produce human-like robots"

To show that this is false requires showing both that AI has scientific goals other than the replication of human intelligence and that its technological goals are not primarily the production of human-like robots.

## AI as Science

For the purposes of this paper, AI as science can be defined as the scientific study of all forms of intelligence and the attempt to replicate it or parts of it in a machine. 'All forms of' includes, at least, animal, human, and machine intelligence. It is different from and a more worthy scientific enterprise than the study of *only* human intelligence or animal intelligence. The reason for this is that the attempt to recreate natural abilities in artefacts of various sorts leads to a far deeper scientific understanding of the general scientific principles underlying those natural abilities than would any mere observation of their natural occurrence.

The classic twentieth century illustration of this principle would be the study of flight. At the time that the early aviation technologists succeeded in building flying machines, biology textbooks generally explained the abilities of birds by stating that birds could fly because they had the power of flight. This account of bird flight was gradually modified over the century as reliable aerodynamic data was produced from the study of aircraft. We now know that birds, insects, and flying mammals fly by complying with the same set of scientific laws as do aircraft [1]. It is to be hoped that the twentieth century will see a similar development in the scientific study of intelligent behaviour.

AI scientists sometimes advance an even stronger version of my claim that study natural abilities through building them into artefacts is more worthy science than merely observing them in nature. This is the claim that *only if* one can build a copy of a natural ability – usually some facet of natural intelligent behaviour - can one claim scientific understanding of that ability.

The claim here is merely that the requirement for understanding intelligence at a level general enough to enable replication or partial replication in an artefact is a scientifically productive requirement. All other things being equal, we should prefer this approach to natural science over one which seeks only to observe.

Unfortunately AI scientists, rather like early astronomers, tended to look at the subject from their own perspective. That is to say that they saw human intelligence as the starting point and wanted to develop the scientific study of intelligence from that starting point. Just as it was natural for pioneer astronomers to assume that the earth was at the centre of the universe, so it has been natural to assume that human intelligence is at the centre of the universe of intelligence.

Science can be hard on such arrogance. Just as it turns out that we live on an undistinguished rock revolving around a middle class star far from the centre of a very ordinary galaxy, so it seems there is nothing central or optimal about our intelligence – or at least our present superficial analyses of it.

An important contributor to the mistaken view that AI should be primarily concerned with the replication of human intelligence was the so-called Turing test. I have argued at length (for example in Whitby 1996a) that it is a misreading of *Computing Machinery and Intelligence* (Turing 1950) to view AI as being concerned with the replication of human intelligence. It is probably unnecessary to repeat that here. It will suffice to say that Turing's 1950 paper is primarily about the change in human attitudes that computing machinery might bring about in the years between 1950 and 2000. It most certainly does not say that the goal of any new (in 1950) science should be indistinguishably to replicate human intelligence. This goal would be as stupidly anthropocentric as astronomy based on the earth as the centre of the universe. Science has to be interested in the whole space of intelligence. After fifty years, AI as science is beginning to suggest that human intelligence may be very far from the centre of the space of possible intelligences.

### **AI as Technology**

Just as it is arrogant and anthropocentric to see human intelligence as central to the scientific study of intelligence so it is anthropocentric to see the goal of AI as technology as 'making machines that behave just like us'. Technology should exist only to improve the lives of humans. It is far from clear that human-like machines have any part to play in this. Since there are about six billion examples of human intelligence available it would be much better for AI technology to set about building useful machines. In this it has been quietly but highly successful.

As a matter of history, it seems reasonable to subsume all of modern computing technology as AI technology. Until the 1950s, the very word 'computer' meant a human who performed routine calculations. A major motivation for the pioneers of computing – in particular Alonzo Church and Alan Turing - was the understanding of human intelligence. In 1948, at the time Turing was actually writing *Computing Machinery and Intelligence* (Turing 1950) he was also engaged in writing code for the first, and at that time the only, general purpose digital computer in the world. Historically, AI as science seems to predate computing. It was AI as science that gave birth to modern computing.

It is true that the infant computing industry was soon deflected into pursuing goals rather different from that of the scientific understanding of intelligence. Most notable among these would be the optimization of military and commercial communication and decision-making. It is for this reason that AI has come to be seen as something distinct from computing.

Nonetheless AI as technology, even when considered as distinct from computing, has been spectacularly productive. The repeated adoption of AI innovations by the computing industry has rather hidden just how successful AI has been in producing useful and effective technology. Often quoted examples are the introduction of time-sharing by central processors (CPUs) and the technique of fast-prototyping of software. In fact the majority of computing innovations owe some degree of inspiration from AI.

In addition there are many examples of successful AI technology being incorporated into other forms of technology - from programs which analyse financial transactions to programs which animate characters in computer games. There is also a slow but steady development of autonomous robots. Many examples of successful AI technology could be listed. As an illustration for the purposes of this paper one particularly informative AI success will be described in detail in the following section.

## **From Machine Learning to Data Mining**

Christopher Columbus put to sea in 1492 with the intention of finding a short cut from Europe to India. He didn't find it; indeed it was not there to be found. What he discovered was a whole new (to Europeans at least) continent which was of far more profound historic consequence than any short sea-route could ever be.

The development of the AI technology known as data mining is closely analogous to Columbus' discovery. During the 1980s there was much research into an area of AI known as machine learning. The reason for this was that the power of AI systems containing and using real-world knowledge had been amply demonstrated. There was but one drawback to such systems. Real-world knowledge invariably proved very difficult and costly to acquire. This so-called 'knowledge acquisition bottleneck' meant that knowledge-based systems were almost impossibly expensive and restricted.

An obvious short-cut would be to enable the machines to learn for themselves and much research was conducted in and around this area. Just as in the short sea route to India, machine learning as originally conceived proved an impossible goal. In passing, it should be remarked that the goal was perhaps too anthropocentrically conceived in the first place. Also the main reason for this failure was the fact that the process of learning turned out to be far more complex than anybody had ever imagined before attempting to get machines to do it. Machine learning is a paradigm justification of some of the claims made in previous sections.

However, although the goal of a machine that could learn as effectively as humans seem to do remained unfulfilled, much was achieved. The main technological product of research into machine learning was a wide and diverse collection of programs that could sift and generalize from messy real-world data.

To call these programs 'diverse' is no mere diplomatic nod to their creators. They were developed under very different assumptions about what constitutes learning and with a fair amount of antipathy between the different camps. In the study of intelligence, just as in global exploration, we cannot always avoid the intrusion of the human tendency towards factionalism.

During the late 1990s the various successful techniques for exploring data were combined and given user-friendly interfaces. This technology became known as data mining and proved extremely useful. A data mining package enables a human to search a vast amount of confused or unstructured data and to extract the useful nuggets – the 'diamonds' in the mine of data. By the 1990s electronic trading records and accounts had generated large databases that were too large to be searched by hand. However, this should not be seen as the only application for data mining. All manner of scientific and social data can be mined in the same way as commercial data.

Commercial success did not follow instantly for the pioneers of data mining just as it did not for Columbus. They had developed a totally new type of technology and it had no obvious market. In the end, however, the immense usefulness of the technology itself generated a market which made small fortunes for the pioneers and remains highly profitable today.

This particular technology is illustrative of many of the arguments made in this paper. Firstly it is a highly useful and highly successful technology which stems entirely from research in AI. Secondly it starts from the fairly anthropocentric goal of getting machines to learn more like humans but actually yields a technology which is an extremely powerful information handling tool. This tool increases the power of those humans who use it and will probably be essential for us all if there really is an 'Information Age'

Lastly it provides an existence disproof of the claim that any particular AI technique has failed. One of the most interesting and obvious features about a data mining tool is that most of the major techniques of AI are combined in one single product. A data mining product combines AI techniques from symbolic AI, from neural inspired computation, and from evolutionary computing all in one user-driven package. Proponents of each of these techniques were (and in some cases still are) given to calling any one or more of the other techniques a hopeless failure. The fact that they can all be built into a single package on a

computer desktop makes these failure claims look rather parochial and silly. The fact that it is such a useful package should silence the failure myth completely. Unfortunately the susceptibility of humans to rational persuasion is something that AI as science has yet to fully explore.

## Notes

1. This best account of the overlap between bird and aircraft aerodynamics is undoubtedly 'The simple Science of Flight from Insects to Jumbo Jets' by Henk Tennekes. (Tennekes 1997)

## References

Brooks , R.A., 1991, *Intelligence without Representation*, Artificial Intelligence **47**. pp. 139-159

Dreyfus, H., 1998, *Why Symbolic AI Failed: The Commonsense Knowledge Problem* , available at: [www.hfac.uh.edu/phil/garson/DreyfusLecture1.htm](http://www.hfac.uh.edu/phil/garson/DreyfusLecture1.htm)

Tennekes 1997, *The simple Science of Flight from Insects to Jumbo Jets*, MIT Press, Cambridge MA.

Turing, A.M. 1950, 'Computing Machinery and Intelligence', *Mind*, vol.LIX, No. 236. pp.433-460 .

Whitby, B.R. 1996a, *The Turing test: AI's biggest blind alley?* Millican and Clark (Eds) *Machines and Thought, The Legacy of Alan Turing Vol.1*, Clarendon, Oxford.

Whitby B.R. ,1996b, *Reflections on Artificial Intelligence: The Social, Legal, and Moral Dimensions*. Oxford: Intellect Books.