[This is chapter 3 of
Forsyth, R.S. (1988). *From Here to Humanity*. Nottingham: Pathway Publications. pp. 31-45.

A slightly edited version of this chapter appeared as
Forsyth, R.S. (1988). The Trouble with A.I. *Artificial Intelligence Review*, 2(1), 67-77.
]

Chapter 3
SOME OF MY BEST FRIENDS ARE ROBOTS

Richard S. Forsyth

> "Men and women of the electronic age, with their desire to sweep along in the direction of technical change, are more sanguine than ever about becoming one with their electronic homunculus. They are indeed remaking themselves in the image of their technology, and it is their very zeal, their headlong rush, and their refusal to admit any reservation that calls forth such a violent reaction from their detractors." David Bolter -- *Turing's Man*.

The image of artificial intelligence (AI) has changed in recent years. These days the AI expert is as likely to wear a neat grey suit and talk about billion-dollar business opportunities as to sport a beard and tee-shirt and spend his time discussing knowledge representation. But do not be deluded: the hollow-eyed hacker, programming till dawn, represents the spirit of AI better than the smooth-talking salesman. The AI business bandwaggon is a financier's fad. The true objective of AI is what it was before the venture capitalists moved in and what it will remain after they have gone on to ransack other treasure-houses of learning: that objective is nothing less than the creation of superhuman beings.

Judging AI in terms of commercial products such as expert systems for dyspepsia diagnosis is like justifying the space programme by non-stick frying pans. Both enterprises are voyages of exploration. Saleable products devised along the way are merely spin-offs. AI is science fiction carried to its logical conclusion. The goal is the production of superhuman intellects -- i.e. thinking beings made through technology, by-passing the traditional method of raising children.

In this valuation of brainchildren above bodily children, in this attempt to reproduce without sex, AI is the culmination of an age-old yearning in western civilization. Pygmalion is said to have fashioned Galatea from marble and then fallen in love with her. Dr Frankenstein made his mythical monster from chopped-up corpses and then fell into despair. Between these two landmarks, legends about golems, talking heads and homunculi surfaced at regular intervals like a recurrent motif in a piece of music. ln our own time the film studios pump out a steady stream of robots that can walk, talk, love, hate and plot the overthrow of humanity.

It is not hard to understand this fascination with the creation of conscious creatures. In a biological sense that is the only purpose that our lives can be said to have. Because of its central importance, people have always tried to exercise control over reproduction from primitive fertility rites to modern methods of contraception. Industrialization accentuates this tendency, which ranges from straightforward obstetric practices to the visions of 'designer genes' conjured up by recent advances in biotechnology. Though we have not reached the 'brave new world' of human genetic engineering, techniques like artificial insemination (also known as AI), in vitro fertilization, abortion of genetically abnormal embryos and the like all in routine clinical use. Calls to restore 'natural childbirth' implicitly recognize that reproduction is fast becoming an industrial process. The overall trend is relentlessly towards ever increasing technological control, ln the ultimate it will replace blood, sweat, tears and other messy secretions by the purified atmosphere of the electronic factory. And that is where Al fits in.

But why should we want to make artificial intelligences? Why should healthy young graduates spend the hours of darkness hunched alone over computer terminals in a vain attempt to make a thinking being when they could spend the night in bed with a friend and very possibly succeed?

The answer is a single word: control, Ever since the Reformation our society has valued control over nature. We technologize things in order to control them better. Mass-production is preferred to craftsmanship because it produces more goods. Agribusiness takes over from farming (as farming took over from hunting and gathering) because it feeds more mouths. Automation, mass-production, industrialization -- these form the very bedrock of our culture. At present population levels we literally cannot live without them.

There is another reason for seeking AI. A baby can only become, at best, a person. We cannot expect a child to grow into anything much better than a Leonardo or a Mozart; and we cannot realistically expect even that. If we want to improve on the human race it is necessary to design something different, something that need not be biological at all. Edward Fredkin, formerly professor of electrical engineering at MIT, is a spokesman for this view of human destiny.

> "There are three events of equal importance, if you like. Event one is the creation of the universe. It's a fairly important event. Event two is the appearance of life. Life is a kind of organizing principle which one might argue against if one didn't understand enough -- shouldn't or couldn't happen on thermodynamic grounds or some such. And, third, there's the appearance of artificial intelligence. It's the question which deals with all questions. In the abstract nothing can be compared to it. One wonders why God didn't do it." (Fredkin, quoted by McCorduck, 1979.)

In this godlike quest, AI researchers have so far been spectacularly unsuccessful; but that is not the point. The point is that AI workers are engineers rather than pure scientists. Unlike cognitive psychologists, who look to the computer for models of the brain, they regard the brain as merely a first approximation -- a rudimentary prototype that can be thrown away once it has been superseded by something better.

Naturally there are many individuals working in AI who do not subscribe to this promethean viewpoint. Their motivations may be as simple as wanting to obtain a Ph.D. or to make a voice-driven typewriter. But there is a hard core of AI researchers who really are committed to the idea of the ultra-intelligent machine (or UIM), and that is the idea which energizes the whole field. (See Good, 1982.) Moreover, seen within its social and historical context, the final destination of AI is clearly the UIM whatever the specific motives of the people involved in pushing it there. David Bolter puts it as follows.

> "The literature of the movement often gives the impression that practical applications are of secondary importance, that a project ceases to be AI as soon as it becomes practical. What really matters to the AI specialist is to realize in transistors the image of a thinking human being: to make an electronic man." (Bolter, 1986.)

Turing's Test

Will they ever succeed in this grandiose dream? One of the first people to confront this question was the mathematician Alan Turing, who in 1950 published a seminal paper entitled 'Computing Machinery and Intelligence'. In it he considered how to answer the question: can a machine think?

Turing envisaged a situation in which an interrogator sat in one room linked via a teletype line to two other people in other rooms, a man and a woman. The man and woman were pitted against each other in a game: the man's task was to bluff the interrogator that he was, in fact, a woman; the woman was required to assist the interrogator by being truthful. Turing adapted this diverting pastime for the purpose of operationalizing intelligence. If the imitation game were played between a man and a machine, with the machine trying to pass itself off as human, and the interrogator failed to distinguish between them, then the machine would deserve to be called intelligent. Turing forecast that this would in fact occur around 2000 AD.

For a long time I half-suspected Turing's test might be some sort of spoof, a refined intellectual con-trick played by Turing with the aim of proving to his own private satisfaction that the academic community would swallow a preposterous parlour game as the solution to a profound philosophical problem. After all, in the purely human version of the imitation game, if the man fooled the interrogator into thinking he was female, we would still not call him a woman. (Turing's homosexuality may have coloured his idiosyncratic opinions on this subject, as others have pointed out.)

Halpern (1987) has characterized Turing's test as a 'myth', but it is an enduring myth. One of the reasons for its endurance is the lack of plausible alternatives. When you try to think of something better, it is easier to appreciate Turing's difficulties, as I discovered one evening shortly before Christmas 1984.

On that occasion I fell into a dispute with Brian Morris (a robotics expert and computer journalist) in which he maintained that machines would be able to think by the year 2000 and I denied it. Both parties to the debate being well lubricated, a wager of £2500 was struck on the spot. But even with the aid of several more glasses of beer, we never managed to lay down the rules for deciding the momentous question when the time arrives. Once you are forced into the position of trying to devise something better, the Turing test does not seem quite so silly after all; so that is roughly what we fell back on to settle our bet. But even though I have a tidy sum riding on the outcome of Turing's prediction, I suspect we will not be sure which of us has won the bet. The test is not so definitive as it appears.

Indeed, some people have asserted that the test has already been passed. The following hilarious episode was recounted by Daniel Bobrow (in McCorduck, 1979).

When Bobrow worked for Bolt, Beranek and Newman in Boston in 1968 the following encounter took place. One Saturday morning around 9 a.m. a Vice President in charge of selling the Telcomp computing service arrived to use the computer for a demonstration. He found a note attached to the teletype that read 'TYPE ALL MESSAGES INTO THIS TELETYPE AND END THEM WITH A PERIOD'. This note was left by the programmer who had just got a version of the ELIZA program (devised by Joseph Weizenbaum (1967) as a parody of a nondirective psychotherapist) to run. The VP thought this meant that the teletype was linked to Bobrow's home and wanted to check with Bobrow whether he could use it. The following conversation ensued.

VP: Can I use the computer this morning?
COMPUTER: Why do you ask?
VP: I might be able to make some additional Telcomp sales.
COMPUTER: Why aren't you sure?
VP: My prospective customers have never seen a demonstration of the system.
COMPUTER: Does this mean very much to you?

VP: Of course it does.
COMPUTER: You seem very positive about that.
VP: This is ridiculous.
COMPUTER: What do you mean ridiculous?
VP: Is there any reason why I can't use the computer now?
COMPUTER: Tell me more about your prospective customers.
VP: Please dial me up on 491-1850

After the last remark the VP left out the period, which meant that the computer stopped replying. This so infuriated him that he called up Bobrow, woke him from sleep and said:

VP: Why are you being so snotty to me?
BOBROW: What do you mean why am I being snotty to you?

The VP angrily read out the dialogue and could get no response but laughter. It took a while to convince him he had been talking to a computer. Of course the VP was not aware that any trial was in progress, nor had he a second (human) interlocutor as a reference point. But an even closer approximation to success in the imitation game was reported by Colby and colleagues (1971) His team wrote a program to simulate a paranoid schizophrenic which conducted the following chilling dialogue with a human doctor.

DOCTOR: Why are you in hospital?
PARRY: I shouldn't be here.
DOCTOR: Who brought you here?
PARRY: The police.
DOCTOR: What trouble did you have with the police?
PARRY: Cops don't do their job.
DOCTOR: What do they do instead?
PARRY: Cops arrest the wrong people.
DOCTOR: They arrest the right ones sometimes.
PARRY: The law should get those Italian crooks.
DOCTOR: Are you Italian?
PARRY: My background is British but I was born in this country.
DOCTOR: Do you know any Italian crooks?
PARRY: I try to avoid the underworld.
DOCTOR: Are you afraid of the underworld?
PARRY: The Mafia are the worst gangsters.
PARRY: I have incriminating evidence against the Mafia.
DOCTOR: Are you sure of all this?
PARRY: You don't believe me?
DOCTOR: I am doing my best,
PARRY: (after a pause) Can't you think of a decent question to ask?
DOCTOR: You seem hostile to me.
PARRY: Some of your remarks annoy me.
DOCTOR: I am sorry.
PARRY: Forget the excuses, you guys are all alike.

You may be thinking that since Weizenbaum's program (1967) gives a passable imitation of a psychotherapist and Colby's does the same for a paranoid patient, why not put them together? Doing exactly that was a favourite amusement among computer specialists using the ARPA network during the 1970's (though it is believed that no cure was effected). Berry (1983) quotes an amusing 'dialogue of the deaf' between PARRY and ELIZA.

PARRY: In my opinion you have a one-track mind.
ELIZA: What makes you think I have a one-track mind?
PARRY: I used to go to the tracks every week.
ELIZA: What does that suggest to you?
PARRY: I avoid racetracks now because the races are fixed.
ELIZA: Is that the real reason?
PARRY: I'd rather not discuss it.
ELIZA: What else would you like to discuss.
PARRY: That's something I don't want to go into.

Interchanges such as those quoted above have been seriously put forward as evidence that Turing's test has been passed. This suggests that the imitation game is perhaps too flexible a yardstick to constitute an acceptable definition of machine Intelligence. Even if it is passed one day in its strict form, the argument over machine cognition will persist. Really we have no satisfactory way of defining thinking without reference to people. To put it in a nutshell: we don't know what intelligence is, but we're going to build it anyway.

Electronic Brainstorms
We cannot frame a watertight definition of thinking which encompasses all the qualities of human thought; but we can define rationality, and it is in the nature of AI that it explores the outer boundaries of what is rational.

AI researchers work on many and varied problems. Some try to get computers to understand natural language or interpret scenes relayed via TV camera. Others work on game-playing or expert systems. There is not space to do justice to the full diversity of the field here, so we concentrate on one subfield that exemplifies both the glories and the failings of AI -- machine learning. Readers desiring a more thorough survey of AI in general can consult excellent textbooks by Charniak & McDermott (1985), by Elaine Rich (1983) or by Patrick Winston (1984).

Until quite recently AI systems have conspicuously lacked learning ability. Most chess-playing programs, for instance, while they may give human masters a run for their money, cannot learn from their mistakes. If they make a blunder in one game, they will go on making the same blunder in future games until the programmers fix it, normally by reprogramming. This is the antithesis of intelligence as humans understand it. In short, current AI programs are *idiots savants*, and can only improve by something more akin to brain surgery than learning. However, a small but growing band of researchers have started to tackle this key problem. In the 1980's machine learning became respectable again, and some significant findings have been reported.

One teachable machine is known as WISARD. It is a vision system that was developed at Brunel University. WISARD can be taught to make the rather subtle distinction between smiling and frowning human faces (Aleksander & Burnett, 1984). Technically WISARD is quite simple, but it gains from being implemented as special-purpose hardware. In fact it out-performs any other visual recognition system on the market.

Essentially it works by sampling a large number (over 32000) of feature detectors, each of which monitors the status of a group of eight 'pixels' or picture-points in the image. During the training phase it records, for each feature-detector, what configurations of 8 bits (e.g. 10111101) are associated with each type of pattern. Having been trained, it identifies a fresh pattern by totting up the number of features 'voting' for each class of pattern and picks the one with the highest score.

WISARD works at high speed because it carries out many operations in parallel. One of its intriguing features is that it exhibits 'emergent properties'. That is to say, quite sophisticated behaviour emerges from simple (but numerous) elementary components. Emergent properties of this sort are characteristic of the biological kingdom. Indeed life and mind have both been described as emergent properties of matter.

A rather different self-improving system worthy of note is the AQ11 program (Michalsky & Chilausky, 1980) which found better rules for soybean diagnosis than a human expert in plant pathology. To test the system its authors collected 630 questionnaires describing the symptoms of plants with a variety of diseases. Each plant was described on 35 attributes (or 'descriptors') such as seed colour, seed size, condition of roots etc. The correct diagnosis for each specimen was also recorded.

AQ11 generates rules in a language whose elementary items are called 'selectors'. A selector asserts one feature of a plant, e.g. [seed discoloration = absent] which states that there is no discolouration of the seeds. A rule is a combination of one or more selectors such as that below (D3) which the system produced as the rule for classifying Rhizoctonia Root Rot.

D3:
[leaves = normal] [stem = abnormal]
[stem cankers = below soil line]
[canker lesion color = brown]
OR
[leaf malformation = absent] [stem = abnormal]
[stem cankers = below soil line]
[canker lesion color = brown]

This rule consists of two descriptions linked by an OR. Each description, in this case, contains four selectors. Logical AND is implied between selectors: i.e. all four terms in one or other of the descriptions have to be true for this rule to apply. Actually this particular rule would be more compact if the system knew that the selector [leaf malformation = absent] is a special case of [leaves = normal].

Rule D3 and 14 others were automatically generated from 290 training instances presented to the AQ11 system. This training set was extracted from the 630 cases available in total. The remaining 340 cases were used later to assess the performance of the computer-generated rules. Whereas the human expert's diagnostic rules gave the correct first-choice disease on 71.8% of the test cases, AQ11's were correct in 97.6% of their first choices.

AQ11 works by trying to build up a description (a conjunction of selectors) that covers as many positive instances of the class under consideration while excluding all the negative instances. This is purely a process of symbolic manipulation: each new example given to the system causes it to amend its current description in various ways (e.g. by adding or deleting selectors) to fit the facts better.

Another impressive learning system, called Meta-Dendral, devised rules about mass-spectrometry sufficiently original to merit publication as a contribution to the science of organic chemistry (Buchanan & Mitchell, 1978). It discovered new cleavage rules for describing how molecules of the ketoandrostane group split up in a mass spectrometer. These rules were subsequently incorporated into the Dendral expert system (Buchanan & Feigenbaum, 1978) which interprets mass spectra.

The mass spectrometer is an instrument that bombards chemical samples with accelerated electrons. This causes the chemical bonds holding the compound together to break. The fragments then pass through an electromagnetic field that separates fragments with low charge and high mass (which are not deflected much) from those with low mass and high charge (which are). The result is a plot of intensity against mass-to-charge ratio.

Trained chemists can identify the molecular structure of the substance thus treated by examining the plot and noting significant peaks and troughs. Dendral does the same, essentially by testing hypothesized structures against a computer simulation of a mass-spectrometer. Meta-Dendral was designed to enhance Dendral's performance by finding new fragmentation rules, which it did successfully.

Meta-Dendral has two main subprograms -- Rulegen which generates new fragmentation rules and Rulemod which tidies up Rulegen's output. Rulegen starts with the most general rule possible (that some atomic bond will break) and gradually makes it more specific. It repeatedly creates 'offspring' of existing rules by making them more specific -- i.e. stating which bonds will break under what circumstances. A descendant rule is retained if:

> it predicts fewer fragmentations per molecule than its parent;
> it still predicts fragmentations for at least half the molecules under consideration;
> it predicts fragmentations for as many molecules as its parent (unless the parent was 'too general').

This kind of 'learning' is based on the process of scientific induction. An even more obviously scientific discovery program is Bacon.4 (Langley et al., 1986). Bacon.4 has 'rediscovered', among other things, Ohm's Law, Archimedes's Principle of displacement, Newton's Law of gravitational attraction and Kepler's Laws of planetary motion.

The program is presented with information such as that tabulated below, which contains Borelli's data about the orbits of Jupiter's moons.

Moon	**distance**	**period**	**D / P**	**D^2 / P**	**D^3 / P^2**
Io	5.67	1.769	3.203	18.153	58.15
Europe	8.67	3.571	2.427	21.036	51.06
Ganymede	14.00	7.155	1.957	27.395	53.61
Callisto	24.67	16.689	1.478	36.459	53.89

(Note that the original figures contained inaccuracies.)

To emulate Kepler, and discover that the square of the orbital period is proportional to the cube of the distance from the primary (Jupiter in this case) the program looks for concomitant variation. When it finds a regular association between one column and another it creates a third column in the table by simple mathematical operations such as multiplication or division. For example, D varies with P, so D/P is created. This varies inversely with P, so D^2/P is created; and so on. Eventually it creates a column like the last with a nearly constant value (within a pre-set tolerance of 7.5%) which is what it was seeking. In a sense it has recapitulated Kepler's finding that D cubed is proportional to P squared.

Bacon.4's method for finding constant terms can be encapsulated in just three heuristic rules:

1. If column X has a near-constant value, then formulate a law involving X;
2. Else, if X increases as Y increases, then consider the ratio X/Y and go to step 1;
3. Else, if X increases as Y decreases, then consider the product X*Y and go to step 1.

This method is sufficient to discover many other scientific laws, given suitable data.

None of the laws found by Bacon.4 have been formerly unknown; but the Eurisko program (Lenat, 1982) made a discovery for which a patent was subsequently granted. When Eurisko was applied to the design of VLSI (Very Large Scale Integration) circuits it started with several heuristic rules carried over from previous tasks. One of these said, in effect: if a concept is interesting then try to make it more symmetrical.

Applied to a two-dimensional device it led to the production of a three-dimensional version which has three axes of symmetry instead of just one. That three-dimensional device (a NAND/OR gate) has since been successfully fabricated.

Eurisko made a genuine innovation by the application of a single rule -- a preference for symmetry. (See also Michie & Johnston, 1985.) We must, however, guard against assuming that Eurisko's preference for symmetry is the same as, say, Ptolemy's preference for perfect circular motion in describing the paths of the planets. The computer manipulates formal symbols according to well-defined rules. It 'prefers' a symmetrical construct to an asymmetric one or a simple proof to a complex one in the same way that a chess program 'wants' to capture opposing pieces or 'likes' to keep its king safe from attack.

Yet it is hard not to be struck by these first stumbling steps towards the computer as scientist. Already machines have made non-trivial discoveries, as we have seen, merely by playing with symbols according to fairly simple rules. What lies over the horizon, as techniques become more sophisticated and computers get faster?

Science is one of the noblest monuments of the human imagination. We usually speak of scientific insight or inspiration in tones of awe. Genius is regarded as something mysterious, and those who possess it are treated with reverence (at least after their deaths) as if somehow touched by holy fire. But here we have machines starting to encroach on this sacred territory.

In this context it is worth contrasting Bacon.4's rediscovery of Kepler's third law with the way Kepler did it himself. Both man and program travelled down many blind alleys of fruitless conjecture. For the computer such dead ends are 'nodes pruned off the search tree' by its backtracking algorithm, in decent obscurity. Kepler, oddly enough, was perversely proud of these way-stations on his journey to enlightenment.

> "Yet it gives me pleasure to remember how many detours I had to make, along how many walls I had to grope in the darkness of my ignorance until I found the door which lets in the light of truth." (Kepler, quoted in Koestler, 1964.)

False trails were part of his apprenticeship.

While the computer pursues a literal-minded form of symmetry (or constancy), Kepler was inspired by a quasi-mystical Pythagorean faith in the 'Harmony of the World' (the title of the book in which his third law was published).

> "Why waste words? Geometry existed before the Creation, is co-eternal with the mind of God, is God himself." (Kepler, quoted in Koestler, 1964.)

He took the ancient idea of the 'music of the spheres' and showed, to his own satisfaction, that the solar system was a celestial music box, rejoicing that "the heavenly motions are nothing but a continuous song for several voices (perceived by the intellect, not by the ear)". So both human and inhuman scientists rely on aesthetic principles, but in the machine's case they are extremely superficial.

Kepler actually recorded for posterity the circumstances in which he discovered his third law.

> "On 8 March of this present year 1618, if precise dates are wanted [the solution] turned up in my head. But I had an unlucky hand and when I tested it by computations I rejected it as false. In the end it came back again to me on 15 May and in a new attack conquered the darkness of my mind; it agreed so perfectly with the data which my seventeen years of labour with Tycho's observations had yielded that at first I thought I was dreaming, or that I had committed a *petitio principi*." (Kepler, quoted by Koestler, 1964.)

As Koestler points out, Kepler had sought some kind of linkage between a planet's distance from the sun and the length of its year for the best part of two decades.

> "Without such a correlation the universe would make no sense to him, it would be an arbitrary structure. If the sun had the power to govern the planets' motion, then that motion must somehow depend on their distance from the sun; but how? Kepler was the first who saw the problem -- quite apart from the fact that he found the answer to it, after twenty-two years of labour. The reason why nobody before him had asked the question is that nobody had thought of cosmological problems in terms of actual physical forces." (Koestler, 1964.)

So we have some similarities and some salient differences between human and electronic discovery. The first difference is that the computer's search for regularity (e.g. a constant term for Bacon.4) is a pale shadow of the mystic obsession that drove Kepler to seek geometric harmony in the skies. Secondly, Kepler's innovation was primarily to ask, rather than answer, the right question. It took him a long time to get it right, but once he had decided to look for a mathematical relation between a satellite's distance and its period, someone was bound to find it eventually. Once the correct goal had been formulated the rest, one is tempted to say, was a mechanical process. Kepler set new goals, which present discovery programs (with the arguable exception of Eurisko) cannot do.

Another illuminating feature of Kepler's scientific career is that he was only dimly aware of his main achievements. The findings that gave him most pleasure -- his geometric 'explanation' of why there are only five planets and his elucidation of the celestial music -- are today regarded as quaint historical curiosities. It required another great mind to see in Kepler's work what would be usable by future generations. "Not the least achievement of Newton was to spot the Three Laws in Kepler's writings, hidden away as they were like forget-me-nots in a tropical flowerbed" (Koestler, 1964). In other words, Kepler hit on his major discoveries, like many other scientists, almost by accident.

The limits of Rationality

Today's discovery programs, as we have seen, cannot rival the thought processes of great scientific thinkers. But it is not unrealistic to suppose that one day computers will be more efficient at making scientific discoveries than people -- churning out more new laws per day than any human researcher. This could be one of the most significant developments in AI.

> "Scientists who are quite willing to accept the computer as a workhorse for calculations may have reservations about its role as fellow voyager on the great ocean of truth -- especially if it sails rapidly over the horizon of human comprehension. It is bad enough for a machine to make scientific discoveries that nobody ever thought of: that has already happened to a limited extent. But in due course computers may well discover facts that are literally beyond our understanding." (Forsyth & Rada, 1986.)

If computers can one day excel the best human brains in science it will be a triumph for AI. One of the crowning glories of the human mind will have been formalized, and will therefore cease to appear mysterious. This is one way, among many, in which AI may devalue venerated human abilities. The question is: should we worry?

One man who has worried deeply about it is Joseph Weizenbaum, professor of computer science at MIT and himself an early contributor to AI. His position is that the drive to formalize all aspects of the mind in terms of rational calculation is dangerous. "I would argue that, however intelligent machines may be made to be, there are some acts of thought that ought to be attempted only by humans" (Weizenbaum, 1984). Weizenbaum is a heretic in what he calls the 'temple of technology'.

> "But science may also be seen as an addictive drug. Not only has our unbounded feeding on science caused us to become dependent on it, but, as happens with many other drugs taken in increasing dosages, science has been gradually converted into a slow-acting poison. Beginning perhaps with Francis Bacon's misreading of the genuine promise of science, man has been seduced into wishing and working for the establishment of an age of rationality, but with his vision of rationality tragically twisted so as to equate it with logicality." (Weizenbaum, 1984.)

Weizenbaum's arguments arise from a deep knowledge of modern technology. They can only be sketched here. But his central point is that computer worship, and the mechanistic model of mankind that goes with it, damages us all. It reduces us to instances of the economists' fictitious 'rational man' or something like it -- automata whose sole aim is to maximize an objective payoff function in our dealings with the world. To put it another way: computer technology (and AI in particular) is instrumental in promoting beliefs that lead to an atrophy of the human spirit.

Weizenbaum's criticisms have provoked a number of replies. Kenneth Colby (1976), who felt himself personally the target of some of Weizenbaum's charges, wrote: "over the past four centuries the scientific community has come to mistrust suppressions of inquiry, not only because they protect the status quo but because so often the finger-wagging moralist has turned out himself to be morally confused, piously self-serving, and irresponsibly blind to the consequences of his own oppressive actions". There seems almost as little meeting of minds in the controversy between Weizenbaum and Colby as in the exchanges between the programs, ELIZA and PARRY, which the two protagonists created.

Colby apparently sees Weizenbaum as advocating a kind of scientific censorship, unenforceable and undesirable, and wades in on the side of unrestricted enquiry. But there must be some limits to research. No one can condone the experiments on pain and resistance to cold carried out on prisoners by German and Japanese medical researchers during World War Two. And the tests on live human volunteers of susceptibility to radioactive fallout, conducted in the USA till 1971, have been made illegal. AI workers in their 'ethically neutral' thought bubble are, in the main, perfectly happy to take funding for work on 'smart' weapons and missile control systems just as long as they can get on with what interests them. It is as if thermonuclear weapons had never been invented.

Other leading lights of AI are equally dismissive of Weizenbaum. Most of them simply cannot see what he is getting at. For example, McCorduck (1979) quotes an anonymous colleague of Weizenbaum as saying: "I have the impression that if Joe could do science, he wouldn't be doing this.... Joe hasn't produced science, so he's got to do something, I wish he hadn't chosen this."

This attitude reveals the kind of mentality required to 'do science' in today's AI laboratories.

Others respond to Weizenbaum by saying things like "it is certainly true that computers (or anything else) should not be applied for immoral ends" (Taylor, 1976). They can grasp the idea that AI products might be misused. They cannot understand that the conceptions on which AI rests debase human values. AI scientists still imagine themselves living behind a veil of pure objectivity. When they walk up to that flimsy curtain and hand over the fruits of their labour, they disclaim any responsibility for the misdeeds of their paymasters on the other side.

In short, the AI community simply shrugs off Weizenbaum's attacks and carries on as before. The few who go so far as to consider moral questions legitimate at all generally stop short of the major issues and concentrate on such things as who should take the blame if a medical expert system delivers an erroneous diagnosis.

The Cult of Artificial Intelligence

In general, then, AI scientists ignore critics like Weizenbaum and the Dreyfus brothers (1985) while hurrying on towards the realization of their age-old dream -- a mind in metal, the UIM. How will it all end?

It may turn out to be a wild goose chase, like the alchemists' quest for an elixir of life.

> "A cynic could well argue that current research in AI is doing nothing more than providing jobs for a small but expanding elite in esoteric, specialist topics or large defence-oriented areas, with there being little hope that the vast majority of us will ever benefit in some concrete way from the research, where by 'benefit' we mean an accepted and proven gain rather than one which is advertised as being a gain by the elite who spend our money researching it. The cynic may well have a valid point." (Yazdani & Narayanan, 1984.)

If the cynic is correct, we can heave a sigh of relief. Humanity will have had a lucky escape if the worst thing about AI turns out to be that it wastes money on remuneration for a self-perpetuating clique. For the greatest danger is not failure to produce the goods, but success. One day there really may be computer systems which surpass humans in all modes of rational thought.

> "The effect will be similar to what happens when a modern mining company descends on a remote valley in Papua New Guinea to rip out a mountain or two.... Suddenly [the natives] are confronted with space-age technology. Their arts and crafts become redundant overnight. Witch-doctoring, face-painting, ear-piercing, hunting -- all the skills they take pride in become meaningless. They convert to a cash economy; take menial jobs in the mine; and spend their free hours in the company bar, consuming prodigious quantities of beer. One day, after the ore runs out, the mining facility moves on to fresh deposits, leaving a scar on the landscape and another wrecked culture on the slag-heap of civilization," (Forsyth & Naylor, 1985.)

The UIM, if it ever exists, will be our intellectual superior by definition, but it will not have a personality in any but a metaphorical sense. One thing we can be sure of is that its intelligence will be alien to us. Paradoxically, a quest which had its emotional roots in the striving for greater and greater control over reproduction may force us to abdicate control over reproduction altogether.

> "I suspect there will be very little communication between machines and humans, because unless the machines condescend to talk to us about something that interests us, we'll have no communication.... They won't have much effect on us because we won't be able to talk to each other. If they like the planet and don't want to leave, and they don't want it blown up, they might find it necessary to take our toys away from us, such as our weapons." (Fredkin, quoted by McCorduck, 1979.)

Then again, since they may have been created under a military budget, they may use our 'toys' for games of their own. The robot is the ideal Kamikaze pilot. There is no reason to be confident that UIMs will prove any more beneficial to our survival as a species than homo sapiens was to Neanderthal Man. It is generally accepted among palaeontologists that the disappearance of the Neanderthalers is best explained by assuming they were wiped out by modern humans. Today we use 'Neanderthal' as a term of abuse. They were inferior: we are the clever ones. Why do we see it this way? Is it because our brains are bigger? In fact, our brains are, on average, slightly smaller. The proof of our intellectual superiority is that we survived at their expense. In effect, genocide is raised to the status of the ultimate intelligence test.

Here at last we have found the basis for an unarguable test to replace Turing's quaint imitation game, an operational definition that accords with the lessons of history: machines will be more intelligent than people when they exterminate us. AI triumphalists may turn out to be right. If we really deserved the label *sapiens* we would be preparing seriously for such a possibility now.

The 'Three Laws of Robotics' (Asimov, 1950) have been touted as safeguards against such contingencies; but the protection they offer has been exposed as paper-thin by their own author.

Of course, nothing so concrete as a UIM may come out of AI, but merely by holding it up as a worthy goal, AI degrades humanist values. It is, to use an old-fashioned word, idolatry. Instead of seeking God, we seek to manufacture gods. Legend has it that such ventures invariably have devilish results, and folklore in this respect is probably wiser than the brittle cleverness of our brightest technicians.

The trouble with AI is that it states that, in principle and eventually in practice too, human minds can be defined as formal systems. Everything of importance can be made explicit as rules which a machine can follow. This feeds our culture's excessive love of quantification, which in turn reinforces the socio-economic doctrine that money is the measure of all things. AI also encourages the instrumentalist approach to human beings. After all, a person is just a mechanism, and an obsolescent one at that. As Isaac Asimov has put it (quoted by Moody, 1986): "the human species was simply the most efficient way that nature could find to build the silicon chip".

Inhumanity is the whole point of AI. Its contribution to our cultural values can be summed up roughly as:

1. Nothing subjective or personal really counts;
2. People are information processors which do not happen to be particularly efficient.

References

Aleksander, I. & Burnett, P. (1984). *Reinventing Man*. Penguin Books, Middx.
Asimov, I. (1950). *I, Robot*. Gnome Press, New York.
Berry, A. (1983). *The Super-Intelligent Machine.* John Cape, London, UK.

Bolter, J. D. (1986). *Turing's Man: Western Culture in the Computer Age.* Penguin, Middx, UK.
Buchanan, B. & Feigenbaum, E. (1978). Dendral and Meta-Dendral: their Applications Dimension. *Artificial Intelligence*, 11, 5-24.
Buchanan, B. & Mitchell, T. (1978). Model-directed Learning of Production Rules. In: Waterman & Hayes-Roth, eds. P*attern-Directed Inference Systems*. Academic Press, New York.
Charniak, E. & McDermott, D. (1985). *Artificial Intelligence.* Addison-Wesley, Mass.
Colby, K. (1976). *On the Morality of Computers Providing Psychotherapy.* UCLA Memo AI HMF-6, Los Angeles, USA.
Colby, K., Weber, S. & Hilf, F. D. (1971). Artificial paranoia. *Artificial Intelligence* **2**, 1–25.
Dreyfus, H. & Dreyfus, S. (1985). *Mind Over Machine.* Macmillan, NY, USA.
Forsyth, R. & Naylor, C. (1985). *The Hitch-Hiker's Guide to Artificial Intelligence.* Chapman & Hall, London, UK.
Forsyth, R. & Rada, R. (1986) *Machine Learning: Applications in Expert Systems and Information Retrieval.* Ellis Horwood, Chichester, UK.
Halpern, M. (1987). Turing's test and the ideology of artificial intelligence. *Artificial Intelligence Review* **1**, 79–93.
Koestler, A. (1964). *The Sleepwalkers.* Penguin, Middx, UK.
Langley, P., Simon, H. A., Bradshaw, G. L. & Zytkow, J. (1986) eds. *Scientific Discovery: Computational Explorations of the Creative Process*. MIT Press, Mass.
Lenat, D. (1982). The nature of heuristics. *Artificial Intelligence*, **19**(2), 189-249.
McCorduck, P. (1979). *Machines Who Think.* W.H. Freeman, San Francisco, USA.
Michalski, R. & Chilausky, R.L. (1980). Learning by Being Told and Learning from Examples.... *International Journal of Policy Analysis & Information Systems*. 4(2), 125-161.
Michie, D. & Johnston, R. (1985). *The Creative Computer*. Penguin Books, Middx.
Moody, G. (1986). Book review. *Practical Computing* **9**(12).
Rich, E. (1983). *Artificial Intelligence*. McGraw-Hill, New York.
Taylor, R. H. (1976). Letter. *Datamation*, June issue, 35–36.
Turing, A. (1950) Computing machinery and intelligence. *Mind*, **59**, 236–250.
Weizenbaum, J. (1967). Contextual understanding by computers. *Communications of the ACM*, **10**(9), 474-480.
Weizenbaum, J. (1984). *Computer Power and Human Reason*, 2nd edition, Penguin, Middx, UK.
Winston, P. (1984). *Artificial Intelligence*, 2nd edition. Addison-Wesley, Mass.
Yazdani, M. & Narayanan, A. (1984). *Artificial Intelligence: Human Effects.* Ellis Horwood, Chichester, UK.