FIRST AUSTRALIAN ARTIFICIAL INTELLIGENCE CONGRESS

MELBOURNE 21 NOVEMBER 1986

LEGAL & ETHICAL ISSUES IN AI

MELBOURNE 21 NOVEMBER, 1986

LEGAL & ETHICAL ISSUES IN AI
The Hon. Justice Michael Kirby, CMG*

## AI - A RAPID ADVANCE

Artificial intelligence ("AI") has at last hit us in Australia. A venturesome article in The Age of 4 November 1986 declared that Victoria was emerging "as a world centre of artificial intelligence research".[1] If I were to question this bold assertion, my scepticism would doubtless be attributed to my northern origins. For all that, there is no doubt that artificial intelligence is on the march. We in Australia will not be immune from it. As with most technological developments, there will be great implications for society, for the law and for human rights.[2] These considerations are the focus of my contribution which is necessarily brief and selective.

The first thing to get clear is how quickly expert systems are coming. The Micro Electronics Monitor for 1984 suggested that in the next two to five years the world wide computer industry would produce a wave of AI products that would turn into a tidal wave by 1990. A forecast by the International Resource Development Inc of Norfolk, Connecticut in the United States, estimated that the United States market

for AI products and services would grow from a mere $66 million in 1983 to $8.5 billion in 1993 - and these are real, American dollars! The estimate suggested that the future AI market would be located primarily in the home, factories and offices. At the time of the report, about 50 expert systems had been built. Some were experimental. Others were in use in the companies which had built them. A few were for sale.[3] The eventual goal of all of the research on AI is to develop computer systems which surpass human capabilities in reasoning, problem solving, sensory analysis and environmental manipulation. Some AI commentators do not expect this goal to be achieved within 50 years. Some even doubt that the development of a fully sensing artificial intelligence will ever be achieved. But the lesson of the past twenty years in the development of informatics has been that things happen more quickly than you expect; that things become smaller than you expect; and that today's miraculous products are, almost by tomorrow, obsolete. The difficulty for society is that, however quickly artificial intelligence progresses, human intelligence is locked into cultural, linguistic and other prisons. The fundamental issue is the extent to which our human intelligence, and the human society it serves, will remain the master (and not become the servant) of the remarkable developments of AI.[4]

The extent to which it will ever be possible to establish a reasoning and sensing artificial intelligence is controversial. In his 1984 Reith Lectures, Professor John Searle of the University of California, talking of "Minds, Brains and Science", declared that the present conception of a digital computer - a machine whose operations could be

specified purely formally - eliminated the possibility that the machine could have thought with meaning. As Searle put it "the reason that no computer program can ever be a mind, is simply that a computer program is only syntactical and minds are more than syntactical. Minds are semantical in the sense that they have more than a formal structure. They have content".[5]

This suggestion of a basic limitation in the capacity of computers to think prompted the magazine Nature to comment that Searle's distinction between syntax and semantics was helpful only so far as it went. Unfortunately, declared Nature that was not very far.

> "It creates but does not bridge a gap ... The truth is that there already is a gap, but one that is already narrower than when the Lighthill report appeared 11 years ago. And the purpose of seeking to fill the gap is not to replace people by machines ... but to understand better how people function and, perhaps, to improve the automation of processes as a biproduct."[6]

Other commentators have urged that Searle was wrong to dismiss the increasing complexity of the machine as irrelevant. Professor Ernest W. Kent, in 1980, demonstrated clear comparisons between micro electronic processes and neurophysiological processes of the brain.[7] Dr. John Dawson, of the British Medical Association commented:

> "While Kent, I believe, correctly shares Professor Searle's view that there is a one to one correspondence between the function of the brain and mental experience and that 'a machine built along the lines of our present computing machinery would be unlikely to have conscious

experience on the basis of any similarity to the brain';
he goes on to say that "it is hard to conceive that the
nature of mind is such that carbon, hydrogen and oxygen
as opposed to silicon, copper and gold determine its
occurrence or non-occurrence. It seems more likely that
some aspect of the brain's larger scale construction is
essential, but which one? Complexity is a possibility. It
may be that mind is a property of self sustaining, self
organising data processing systems sufficiently complex
to support it. This possibility is probably the one that
most people who have considered the issue regard as the
principal candidate."[8]

Dawson concludes that mankind will probably in due course
build a machine which will achieve a recognisable
consciousness. He quotes Kent's definition that consciousness
in ourselves has at least two distinct dimensions:

"It comes and goes, we can have it or not, and we can
have it in varying degrees from intense alert attention
and concentration to drowsy relaxation verging on sleep.
Additionally and independently of the level or degree of
consciousness, our conscious activity has content. Our
minds are filled with mental events of all sorts ...
given what is known about the relationship between
reported states of conscious awareness and arousal and
the physical measures of forebrain activities ... it
seems reasonable to identify the existence of the state
of consciousness as a primitive mental experience with
the operations of the recticulo-cortical circuitry and
the state of maintained conscious activity with the
physical activity around the feed back loop."[9]

But whether or not in our lifetime a machine will be developed with a recognisable consciousness, it is plain that the fifth generation of informatics will go a long way down that track. As Dawson points out, his BBC micro computer beats him at chess everytime. The capacity of "thinking machines" grows apace. It will continue to do so. The issue, is therefore, the social and ethical implications of this remarkable development.

IMPLICATIONS FOR PEACE AND SURVIVAL

In the nuclear age, it is obvious that the grandest and most important moral issue is that of the world's survival. In a week in which the United States Secretaries of State and Defense have acknowledged their ignorance of foreign policy initiatives being taken by the President, thoughts must begin to run to the human control which exists over the means of nuclear war. If people so high are (or say they are) ignorant of major and sensitive foreign policy developments, can we have certainty that "on our side", those who have control over the means of life and death, have an appropriate, well informed line of command.

In the same week, the German Chancellor (Dr. Kohl) has criticised as intolerable the leakage of chemicals into the Rhine from a chemical factory in Basl, Switzerland. This instance, and the earlier nuclear leakage at Chernobyl, demonstrate the increasing inter-dependence of nations. In a scientific age, we are all inter-related. Transborder problems grow in number, complexity and danger.

One of the chief implications of artificial intelligence is obviously for defence systems. Indeed, there is a joke told by J.A. Campbell of the University of Exeter. He tells it thus:

For the U.S.A., the joke says that the organisation which
will have the biggest share of the funds at its disposal,
move fastest and syphon off the best of what is obviously
a limited stock of talents and experience in AI to work
on its own choice of applications, is the Pentagon. Hence
the American effort will be self handicapping as far as
commercial competitiveness is concerned. The British
section of the same joke says that the Government will
tie up the relatively few equivalent AI specialists for
several years on committees to decide how to handle the
notional funds when they may or may not be available
...''10

It is clear that if expert systems take control of
defence preparedness, they may put human life and health at
risk on a very large scale if ever they go wrong. Yet they will
not be subject to the same rules of control as for civil
applications. The primary argument for introducing AI in
defence systems is that there are some dangers, sequences of
warning signals and so on which are just too fast for humans to
be able to react quickly enough. However, as Campbell points
out, there are some situations, even in military history, where
doing nothing would have been a preferred solution to acting in
a way which, on paper, was orthodox and appropriate but which
led onto military disaster. Judgment is the key here. The
interposition of human sensitivity and evaluation is important.
Yet how will this be possible if systems are automatically set
off, on the excuse that the human mind would be just too slow
to perceive the danger? At stake here is nothing less than the
survival of mankind. There cannot be a more serious and
important ethical question before us.

B. Shackel of Loughborough University explained his
concern in a way that is relevent:

"[W]e are asking machines to take responsibility for
complex decisions and delicate judgments ... But are we?
I know of no instance where this is happening to a
significant degree and I know of many cases where it
could happen but does not. For example, the Victoria Line
would be entirely automatic with no staff on trains
themselves, but of course there is at least one driver to
handle the unpredictable or unpredicted, such as the
variation in public human behaviour. Again flight deck
automation is now such that the pilot of a 747 at the
start of the Heathrow runway could switch to automatic
and touch nothing until the plane came to a halt at the
end of the landing at Los Angeles, and in principle all
the taxiing could be automatic also. However, the pilots
will be there for a quite a time yet, to handle the
unpredictable or unpredicted, such as the American
Airlines DC10 captain who successfully landed after the
cargo door blew out (ten months before the disasterous
crash outside Paris)."[11]

So the first ethical question is how we interpose similar
skilfull human judgment in the critical decisions of life and
death which could affect the whole planet. Surrendering the
entire future of civilisation to artificial intelligence in the
field of "defence" is not, at least at this stage, morally
acceptable.

## IMPLICATIONS FOR THE PROFESSIONS

In an essay on reforming the professions, I pointed out
some years ago that the vulnerability of the professions was

linked to what has hitherto been their special strengths: the fact that practitioners act as special repositories and disseminators of specialist knowledge. But if that knowledge can be integrated into automated systems an important issue arises for the future of the professions. John Dawson of the British Medical Association considers it likely that the impact of expert systems on medical practice will not be to create more jobs for the orthodox professional. It will be to create more jobs for the computer specialist and systems engineers. Furthermore, he pointed out, it will raise questions about who should accept responsibility for the success or failure of a patient's treatment.[12] The probability seems to me to be that information systems will move from an adjunct to professional practice to the actual control of some professional activity. I will discuss the impact on my own profession later. But computer aided design in engineering and architecture and computer monitoring of intensive care patients is already with us. According to Dawson, hospitals will be used increasingly only by patients requiring surgical procedures and intensive care. The areas of clinical practice in medicine that require manual skills (such as surgery endoscopy, anaesthetics and obstetrics) appear less at risk than others. Counselling skills, for example, in the care of the dying and mentally ill are likely to increase in importance as "scientific" medicine becomes more machine dominated.[13]

But as machines take over the control of monitoring and directing patient care, an important moral question will be the philosophy that is written into the software programs. What, for example, will be the philosophy written into the monitoring

of a grossly retarded or defective neonate? Dawson asks - only
partly in jest - whether it will be possible to buy Catholic or
Scottish Presbyterian software or software which reflects a
Jewish philosophy or a humanist one? These are not really
humorous questions. As we involve machines in the interface
with human life, we are dealing with very delicate ethical
questions and ones upon which human judgment, evaluation and
ethical decision making have hitherto been considered vitally
important. One of the other ethical questions which Dawson
raises is the loss of pluralism. While machine monitoring of
prescription practices and standards of treatment may, in a
macro sense, improve the care of patients, there is a danger
that it will introduce a single standard. Many leaps in
treatment of sensitive questions have depended upon the
courageous individual who sees things differently. Will this be
possible, or so easy, in a situation largely controlled by a
software program with its element of predictable automaticity?

## IMPLICATIONS FOR THE LAW AND THE QUALITY OF MERCY

In my own profession, there are doubtless many changes
which will come with AI, and most for the better. They will
include standardisation and equality of treatment and true
access to accurate decision making because of the interposition
of informatics. Some decisions lend themselves to automated
treatment. Thus, for example, the qualifications for
citizenship may be so treated - at least in the first instance.
Under the British Nationality Act, the provision for British
citizenship may be computed automatically thus:

"For every individual x date y individual z and section
of the Act w.

X acquires British citizenship by Section 1.1a on date y -
If x is born in U.K. on date y and y is after the Act
takes effect.

And x has a parent z

And z is a British citizen.

By section w on date y."[14]

A similar approach could be taken to the Australian law
on this and many other subjects. Many laws will doubtless be
rewritten in order to reduce the judgmental or discretionary
element and to increase the element of automaticity. For
example, I foresee the reduction of personal  injury damages so
that entitlements to compensation will be in the form of social
security payments. These are more readily translated into
automated form. No element of evaluation is required in
determining simple issues of entitlement to weekly payments
according to pre-injury salary and the number of dependents.

But is it desirable to sweep away evaluation in such
matters? For example, how could any program ever be so designed
to compensate properly a person for cosmetic injury? Those
injuries impact different people in different ways. I doubt
that a program could ever be designed to take into account so
many idiosyncratic and personal variables involved in such a
loss, evaluated as open ended general damages.

Likewise, I doubt that a program could ever take into
account the myriad of unexpected events that occur in a
courtroom. In an application for leave to appeal from a
practice decision, the normal rule is that leave will not be
granted unless there is a clear error in the exercise of
discretion or some serious injustice shown which requires

remedy. In the course of a recent application to the Court of Appeal a litigant in person broke down in our presence. His collapse was an important consideration for at least two of the Judges in demonstrating the exhaustion of the litigant and in helping to establish the need for an adjournment to secure legal representation. I cannot conceive that such a factor could ever be written into an automated program. Human judgment requires a human face. It is responding to a complex of factual data. Some matters will be susceptible to automation. Others will not. It would be my hope that, at least in my lifetime, matters could generally come to a human decision maker to stamp onto the decisions which represent the exercise of power in society, the compassion and human understanding which only a human decision maker can offer, at least at this stage.

## OTHER IMPLICATIONS

There are many other moral questions which require consideration. One of them is the growing gap between the information rich and the information poor.[15] There are a number of considerations that are exacerbating this gap. The opinion has been expressed that the fifth generation is likely further to increase the importance of information as a commodity and yet, ironically, will make it harder for the underdeveloped countries to "catch up" because they will continually fall behind for want of access to information.

Other issues involve the impact of AI on deskilling or job enrichment;[16] its impact on organisational change;[17] and its implications for a multitude of personal, ecological, economic and political questions. These are not abstract issues. As usual, the technology is developed more quickly than our institutions are developing to provide the responses.

Shackel has concluded in words which I find attractive:
"However good AI may become, intelligence is not the
whole human. No human viewer has difficulty in
recognising the high intelligence of Mr. Spock in the
Star Trek TV series is not human; other human facets are
missing. Similarly AI takes no account of personality,
emotion, motivation and other characteristics which join
with intelligence to make the whole human being. But what
if we eventually come to AP, AE and AI 2 - automated
personality, automated emotionality and automated
illogicality? Might we then accept the resulting
"integrated artificial expert" as equivalent to the human
partner?[18]

## FOOTNOTES

\*    President of the Court of Appeal, Supreme Court, Sydney 1984-; Chairman, OECD Expert Group on Transborder Data Barriers and the Protection of Privacy, 1978-80; Governor, International Council for Computer Communication, 1984-.

1.    As claimed in The Age, 4 November 1986, 30.

2.    M.D. Kirby, Human Rights - The Challenge of the New Technology (1986) 60 Australian Law Journal 170.

3.    The Micro Electronics Monitor, April-September 1984. Compiled by the Technology Program of UNIDO, Austria.

4.    E. Mumford, Expert Systems and Organisational Change in British Computer Society, Specialist Group on Expert Systems, Proceedings of a Seminar on Social Implications of Artificial Intelligence and Expert Systems, Oxford, 10-12 May 1985, mimeo, 41, 44 (hereafter "Proceedings").

5.    J. Searle, "Minds, Brains and Science", in BBC Wreith Lectures, 1984, Lecture Two "Can Computers Think?"

6.    Nature, Vol 312, 29 November 1984.

7.    E.W. Kent, "The Brains of Men and Machines", McGraw-Hill Publications, New York, 1980.

8.    J. Dawson, "Expert Systems in Medicine" in Proceedings, 18.

9.    Kent, op cit n 7.

10.    J. Campbell, "Expert Systems and Professional Responsibility" in Proceedings, 23, 24.

11.    B. Shackel, "Responsibility - Are We Asking Machines to Take Responsibility for Complex Decisions and Delicate Judgments? Is This Wise? Is this Inevitable?" in Proceedings, 27, 28.

12. Dawson, op cit, 20.

13. ibid.

14. R. Kowalski and M. Sergot, "Computer Representation of the Law", Proceedings, 4.

15. J.A. Thom, "Social Implications of the Fifth Generation", Proc ACS (Vic), Feb 1986, mimeo.

16. H. Rosenbrock, "Deskilling or Job Enrichment" in Proceedings, 38.

17. Mumford, op cit, n 4.

18. Shackel, op cit, 30.