PANEL: ARTIFICIAL INTELLIGENCE AND LEGAL RESPONSIBILITY

Margaret A. Boden - University of Sussex

Panellists: Margaret A. Boden (Chairman); Yorick Wilks;
Marshal S. Willick; Jay BloomBecker, Susan Nycum, Robert Kowalski

President Truman's famous remark — "The buck stops here" — was clearly correct. It's much less clear where the buck stops when one of the elements (I almost wrote "people") in the chain of responsibility is an AI-program.

There are two broad questions to be asked. First, to what extent (if any) can the making of — and the responsibility for — a given judgement or decision, or mistake,he attributed to a computer-program? And second, supposing that responsibility cannot be attributed to a program, with whom does it lie? Who is legally responsible for what a program does? Its user (person or institution, professional or client), its programmers (alive or dead), the domain-experts who provided the knowledge-base ... who?

In the past, the law has derided the notion that one might apply psychological predicates to a machine. Significantly, perhaps, this derision has sometimes resulted in a person's escaping responsibility for a clearly dishonest, and prima facie illegal, action. For example, on January 28th 1972, the London Times reported a case in which the defendant was found not guilty on the grounds that "machines cannot be deceived". A motorist had avoided paying a car-park fee, by manually lifting the exit-barrier instead of putting money in the exit-machine. His defense counsel said "The plaintiff has to be aware that deception has taken place for this case to be proved. It is impossible to deceive a machine as it has no mind and consequently cannot be aware of the deception as a car park attendant might". This argument was accepted by the Bench. In dismissing the case (and seven other similar cases), the Chairman said "Someone has got to be deceived in a case like this, but here this was not so." The penny-pinching motorist got off, because the magistrates ruled that it was in principle impossible to apply psychological categories to machines.

But car-park machines are different from powerful computers, and lifting a sixpenny barrier is different from giving a medical diagnosis, or advizing where to drill for oil. Is this "no-nonsense" judgment of January 1972 a useful precedent for the sorts of legal complications that are likely to arise with the increasing public use of complex AI-systems?

Hackers and laymen alike constantly refer to programs — and a fortiori to AI-programs — in psychological terms. We speak of their reasoning,judgments, evidence , knowledge, ignorance,and mistakes. We speak of what they are trying to do, and what priorities are guiding their decisionss Is this simply sentimentality, a sloppy way of speaking which can and should be avoided — above all, in the law courts? If it is not, if people as a matter of fact do not or cannot avoid using such terms in conceptualizing AI-systems, then what implications follow? If we are allowed to use some psychological words when describing AI-programs, why not all? If we use the language of knowledge and inference, and even of choice, then why not the language of purpose, effort — and even blame?

These questions are the focus of the first two speakers on the Panel, Yorick Wilks and Marshal Willick. The ascription of legal responsibility already varies depending on the "personal" category of the putative offender: states, companies, individuals, the sane, the insane, children, pets, wild animals, servants, and agents. What about computer programs?

Current intuitions about this question may seem absurd in a few years' time, when people are more used to AI-applications. Some of us may already feel uneasy with the judgment that "Machines cannot be deceived". If one wishes to prevent people from wilfully feeding false

information to a computerised system should a person or institution be found, or a legal fiction invented, to suffer (sic) the deception? Or should we be willintj to grant that machines can be <u>deceived</u>, though maybe not <u>disappointed</u>? Sherry Turkle, in her recent book <u>The Second Self</u>, reports that young children growing up in today's computer-culture spontaneously ascribe cognitive concepts (such as <u>knowledge</u>, <u>intelligence</u>, <u>deciding</u>, and <u>mistake</u>) to computers. They also use some conative concepts (like <u>purpose</u>, <u>goal</u> <u>wanting</u>, <u>trying</u>, and <u>failing</u>), at least in the context of problem-solving" on the computer's part. But they adamantly refuse to use affective concepts (such as <u>feeling</u> and <u>emotion</u>), and they also jib at such motivational concepts as <u>caring</u>, and the like. Indeed, the child's concept of what it is to be "alive" is apparently changing, so that affective and conative concepts are stressed at the expense of "mere" cognition. Does this imply that the litigants of tomorrow will allow that computers can make <u>mistakes</u>, but cannot truly have intentions?

Among the intentions which human beings harbour — and not only   in car-parks — are some which are criminal. The third panelist, Jay BloomBecker, discusses a range of examples taken from the current case-law on computer crime. He relates these to some relatively novel issues that may arise, once "Fifth Generation" systems are available. When dealing with programs capable of some degree of "autonomous" reasoning, both crime-detection and the ascription of responsibility are likely to be even more difficult than they are today.

An enormous amount of litigation, at least in the USA, concerns medical issues. Clearly, legal problems will arise in connection with the use, and misuse, and even non-use, of medical expert systems. Various loci of responsibility seem <u>prima</u> <u>facie</u> to be possible: the doctor who uses the system; the patient who knows this is happening (<u>caveat emptor</u>?); the hospital administration; the programmer/s; the specialist physician who supplied the relevant diagnostic or prescriptive rules in the first place; the author of the textbooks used. Many of these individuals may already be dead. But, as Norbert Wiener pointed out, "old programs never die"; could a doctor or hospital be sued for relying on an old out-of-date program? Could they be sued for not using any program at all? The fourth Panellist, Susan Nycum, considers some of the legal problems likely to dog applications of AI in the medical domain.

Finally, Bob Kowalski contributes some thoughts on how "legal" expert systems might be used. His own work includes the building of a system which incorporates the British nationality laws (a prime late-twentieth-century example of Baroque art). What implica-tions, if any, does this project have for the individual and society? Arguably, it would be an improvement on current practices to have nationality-decisions computerized. For a program cannot be affected by (Turkle's subjects would say, it does not care about) anyone's skin colour or physiognomy, or their manner of dress or speech. And arguably, the clarity of the programmed rules might help make clear any basic injustices in the programmed laws themselves: to change the world one has first to understand it. But where would responsibility lie if misclassification occurred? Should Kowalski start saving his pennies, in anticipation of his defense costs in the legal suits of the 'nineties?

And what about the legal implications of other legal or quasi-legal programs? If a program searching for precedents in case-law does not have analogical reasoning powerful enough to find the right one, to whom could the defendant complain?   If governmental and other institutions formulate policies based on legal "decisions" made by in-house programs, who is to know, who is to care, and what can be done?

The panel promises many questions. As for answers, those are more elusive. But since the Panellists include both specialist AI-practitioners and profess-ional attorneys who have already concerned themselves with these questions,we can expect a lively and informed discussion.